

A sokszínűség jegyében

szabo@hit.bme.hu

Jelen számunk rendhagyó abban az értelemben, hogy ezúttal nincs egyetlen központi téma. Válogatásunkkal a hírközlés és informatika több érdekes, részben határterületinek mondható témakörét mutatjuk be. Egyben földrajzilag is sokszínű a kép: szerzőink a győri Széchenyi István Egyetemről, a Veszprémi Egyetemről, Zürichből, Budapesten a Pázmány Péter Katolikus Egyetemről és a BME két tanszékéről jelentkeztek munkáikkal.

Első cikkünk az új Internet-protokollt, az IPv6-ot mutatja be. Ebből a cikkből képet alkothat az Olvasó az új képességekről és arról is, mi várható az általános bevezetését illetően.

A következő írás a távközlési szoftverek fejlesztéséhez létrehozott formális nyelvekkel, mindenekelőtt az SDL-lel foglalkozik.

Az előző számunk központi témája a digitális műsor-szórás volt, ehhez kapcsolódik a mostani DVB-T mérés technikával foglalkozó cikk, amelyben a szerző két magyarországi digitális földfelszíni televíziós állomás mérési eredményeit ismerteti.

Ezt követően az IP feletti beszédátvitellel foglalkozunk, ahol ma egyre inkább előtérbe kerül a szolgáltatásminőség biztosítása. Ehhez megfelelő felügyeleti rendszerre és a VoIP-szolgáltatásokra specializálódott hibamenedzsmentre van szükség. Írásunk átfogó hibamenedzsment-rendszert mutat be minőségi VoIP-szolgáltatást nyújtó hálózatokban való megvalósításra.

Érdekes és társadalmilag fontos alkalmazással foglalkozik két cikkünk: hallássérült emberek számára készülő szájmozgás-animációs program fejlesztésével. Az egyik munkában az audiovizuális beszéd-adatbázis összeállításáról számolnak be a szerzők, míg a második a fejlesztés mobiltelefon-specifikus részleteire és megoldásaira koncentrálnak.

Folyóiratunk e havi számában két cikk is adatbányászati témákkal foglalkozik. Mivel ez sok olvasónk számára valóban határterület, hadd mutassuk be egy kicsit részletesebben a két cikket környezetét.

Az adatbányászat azt a gépesített dokumentumfeldolgozó szakterületet jelenti, amely az internethez kapcsolódóan a nagymennyiségű weboldalon található képi, szöveges és egyéb alakú adatok feldolgozhatóságával, átalakításával foglalkozik. Általánosabb értelemben adatbányászatnak tekinthető minden olyan számítógépes adatfeldolgozó eljárás, melynek célja, hogy nagymennyiségű, az ember által nehezen vagy egyáltalán nem átlátható adatból információt nyerjen ki.

Az adatbányászat nagyon szerteágazó terület, az adatbányászati eljárások ezért számos szakterület ismeretéből merítenek, kezdve az algoritmuselmélettől, a gráfelmélettől a matematikai statisztikán át a mesterséges intelligencia legkülönbözőbb területeig.

Az első cikk a gépi tanulás és a neurális hálók egy, a közelmúltban kidolgozott ígéretes új irányához, a szupport vektor gépekhez (SVM) kapcsolódik. Az úgynevezett LS-SVM hálózatok olyan módosítását és alkalmazási lehetőségét mutatja be, mely a szupport vektor gépeknél jelentkező számítási komplexitás jelentős redukcióját teszi lehetővé. A második a Bayeshálók és azok adat-, illetve szövegelemzési feladatokra való alkalmazásával foglalkozik. Bár a Bayeshálók matematikai gyökerei több száz évre nyúlnak vissza, elméletük és gyakorlati alkalmazásai is az elmúlt 20-25 évben születtek meg.

E cikkekkal két új „adatbányászati” eszközre szeretnék felhívni a Híradástechnika olvasóinak figyelmét, esetleges segítséget adva ahhoz, hogy a jövőben ezen eszközöknek minél több híradástechnikai alkalmazása is megszülessen.

*Szabó Csaba Attila,
főszerkesztő*

IPv6 – A jövő Internet-protokollja?

FARKAS KÁROLY

Computer Engineering and Networks Laboratory (TIK)
Swiss Federal Institute of Technology Zurich (ETH Zurich)
farkas@tik.ee.ethz.ch

Kulcsszavak: Internet, IP, IPv4, IPv6

A kapocs, amely a globális Internetet összefűzi az Internet Protokoll vagy röviden IP. Ellentétben számos más hálózati rétegbeli protokollal, az IP-t a kezdetektől fogva arra tervezték, hogy alkalmas legyen különböző számítógép hálózatok közötti kommunikáció megvalósítására. Az IP elsődleges feladata datagramok¹ továbbítása a forrás- és célállomás között a lehetőségekhez képest legoptimálisabb (best effort) módon. A jelenleg használatban lévő az IPv4-es, amely egy rég² és világszerte elterjedt protokoll az Internet hálózatban, azonban úgy tűnik, hogy nem képes lépést tartani az óriási arányban növekvő, az Internethez csatlakozó eszközök számával és az egyre bonyolultabb útvonalválasztási megoldásokkal, még a folyamatosan kifejlesztett kiterjesztések ellenére sem. Mi lesz hát az Internettel, ha majd egy szép napon az IPv4 már nem lesz képes többé betölteni eredeti funkcióját? A javasolt megoldás egy új és sokkal hatékonyabb, az úgynevezett 6-os verzió (IPv6) bevezetése. De vajon valóban reális ez az elképzelés? Ha igen, mikor fog ez az átállás bekövetkezni?

1. Bevezetés

Az Internet napjainkban összeköttetést biztosít világszerte számos hoszt és számítógép hálózat között. Ez a világméretű kommunikációs hálózat az ARPANET nevezetű, az amerikai védelmi minisztérium által az 1960-as évek elején létrehozott hálózatból fejlődött ki. Miután 1983. január 1-től az ARPANET hivatalos protokollmodellje a TCP/IP (Transmission Control Protocol/Internet Protocol – átvitelt vezérlő protokoll/Internet protokoll)³ lett, az ARPANET-hez kapcsolódó hálózatok, gépek és felhasználók száma ugrásszerűen megnőtt. Ez az időpont tekinthető a mai Internet 'születésnapjának'.

Az Internet sikere a nagyszámú, különféle alkalmazás – mint például az e-mail, fájl transzfer, World Wide Web vagy többfelhasználós játékok –, illetve számos, különböző számítógép hálózat IP támogatásának az eredménye. Ezen alkalmazásoknak köszönhetően látványos multimédia tartalom válik elérhetővé akár egy egérkattintás hatására, mely nagymértékben hozzájárul az Internet gyors növekedéséhez. Az elmúlt évtized folyamán ez a növekedés exponenciális méretet öltött, és az iram várhatóan az elkövetkezendő években sem fog alábbhagyni, ami azt jelenti, hogy az Internethez kapcsolódó gépek száma évente mintegy megduplázódik.

Ennek az óriási sikernek a következménye ugyanakkor az Internet jelenlegi legnagyobb dilemmája: meddig lesz még alkalmas az aktuális IP verzió lépést tartani az Internet növekedési ütemével és mi fog történni, amikor már az IPv4 nem lesz tovább képes elvégezni feladatát? A probléma az Internet természetéből és működési módjából fakad. Minden gépnek, amely az Inter-

nethez csatlakozik, egy IP cím nevezetű, egyedi azonosítóval kell rendelkeznie. Ezen cím segítségével lehet az egyes gépeket megkülönböztetni egymástól és a logikai kommunikációs kapcsolatok végpontjait azonosítani. Ez azért fontos, mert a továbbítandó adat a küldő és fogadó hoszt között, ahogy az általában adathálózatokban lenni szokott, kis csomagokra van bontva és minden egyes csomagot a közbülső állomások a forrás- és célállomás közötti logikai útvonal mentén egyedileg kezelnek. Így minden egyes adatcsomagnak tartalmaznia kell információt a csomag feladójáról és a célállomásról, ahová a csomagnak el kell jutnia, ezáltal lehetővé téve a közbülső állomások számára a csomagok helyes irányba történő továbbítását.

Az eljárást lépésről lépésre történő útvonalválasztásnak (hop-by-hop routing) nevezzük. Minden közbülső állomás – amelyeket gyakran útvonalválasztóknak (router) hívunk – a beérkező adatcsomagban található, a célállomásra vonatkozó IP címet összehasonlítja a saját útvonalválasztási táblájában (routing table) szereplő bejegyzésekkel. Egy ilyen bejegyzéshez egy IP címtartomány és egy szomszédos csomópont azonosítója tartozik. Amennyiben a cél IP cím beleesik valamelyik bejegyzéshez tartozó IP címtartományba, az állomás továbbítja a csomagot a bejegyzéshez tartozó szomszédos csomópontjának. Ha nincs egyezés, a csomag az alapértelmezett útvonal (default route) mentén továbbítódik, amely útirány tartalmaz olyan útvonalválasztót, amelyik várhatóan ismeri a célállomáshoz vezető utat.

Az IPv4 32 bit hosszúságú bináris számokat használ IP címként, amely 'csak' $2^{32} \approx 4,3$ milliárd ($4,3 \times 10^9$) számítógép azonosítását és címzését teszi lehetővé⁴.

¹ Különálló adatcsomagok, melyeket a hálózat egyedileg kezel

² Az első IPv4 implementáció 1981-ben született

³ A TCP/IP tulajdonképpen az Internet esetén használt protokoll rétegmodell elnevezése, amely számos különböző Internet protokoll gyűjtőneveként használatos

⁴ A gyakorlatban ez a limit egy kicsivel alacsonyabb, mert ebből a tartományból nem minden szám használható fel egyedi hoszt IP címként

Jelenleg e címtartomány közel 70%-a van használatban vagy lefoglalva, míg a maradék, hozzávetőleg 1,3 milliárd cím még felhasználható. Mondhatnánk, hogy ez még mindig egy óriási szám! Igen ám, de annak fényében, hogy Kína egymaga fel tudná használni a maradék IPv4 címtartományt egy év leforgása alatt, a helyzet sokkal válságosabbnak tűnik és könnyen vezethet az IPv4 címtartomány kimerüléséhez (address space depletion).

Ezen felül további problémái az IPv4-nek a komplex protokoll fejrész, az új opciók és kiterjesztések hozzáadásának bonyolult folyamata, a különböző szolgáltatások korlátozott száma, az útvonalválasztási táblák folyamatosan növekvő mérete és komplexitása, valamint a korlátozott biztonság. Ezek áthidalása érdekében 1990-ben az IETF (Internet Engineering Task Force)⁵ [1] elkezdte a kidolgozását egy új IP verzióknak, amely manapság IPv6 néven ismert. Noha az új protokoll lényegi részeinek a szabványosítása már befejeződött, a világméretű áttérés az IPv6-ra még várat magára.

A továbbiakban áttekintjük az IPv6 fontosabb tulajdonságait és ismertetjük a főbb érveket, melyek az IPv4 teljes, IPv6 általi leváltása mellett és ellen szólnak, valamint kitérünk arra is, hogy ez az átállás várhatóan mikor fog bekövetkezni, ha egyáltalán bekövetkezik...

2. Az IPv6 protokoll

Előre látva az IPv4 problémáit, 1990-ben az IETF elkezdett munkálkodni egy új IP verzióknak, amely soha nem fogyna ki a címekből, megoldást nyújtana az említett problémákra és sokkal hatékonyabb és rugalmasabb lenne elődjénél [2]. A fő célkitűzések a következők voltak:

- Hosztok milliárdjainak támogatása, még tetszőleges címtartomány-hozzárendelés esetén is;
- Az útvonalválasztási táblák méretének csökkentése;
- A protokoll egyszerűsítése, lehetővé téve ezzel az útvonalválasztóknak a csomagok gyorsabb feldolgozását;
- Az IPv4-nél jobb biztonság (hitelesítés és titkosság) biztosítása;
- Lehetővé tenni, hogy egy számítógép a címének megváltoztatása nélkül barangolhasson a különböző hálózatok között;
- A protokoll fejlődésének biztosítása;
- Az új és a régi protokoll még évekig való egymás mellett létezésének lehetővé tétele.

Hosszú megbeszélések és viták után az IETF kiválasztott egyet a számos, következő IP verzióra vonatkozó javaslat közül, és az IPv6 elnevezést adta neki (az IPv5 már használatban volt egy kísérleti kutatási folyamathoz kapcsolódóan).

Az IPv6 egészen jól megfelel a kitűzött céloknak. Megtartja az IPv4 jó tulajdonságait, elveti vagy kevésbé hangsúlyossá teszi a rosszakait, és újakat is hozzá-

ad, ahol szükség van rá. Azonban fontos megemlítenünk, hogy az IPv6 nem kompatibilis az IPv4-gyel.

A legfontosabb előrelépés, hogy az IPv6 hosszabb címeket használ, mint az IPv4, és ezáltal a címtartomány kiterjedése is lényegesen nagyobb. A címek mérete 128 bit, amely megoldást nyújt az IPv4 címtartomány kimerülésének problémájára egy gyakorlatilag végtelen méretű Internet címtartománnyal. 128 bittel 2^{128} hosztot címezhetünk meg a hálózatban, ami hozzávetőleg 3×10^{38} . Ez egy igen nagy szám! Ha az egész Föld felülete, beleértve a szárazföldet és a vízfelszínt is, számítógépekkel lenne befedve, az IPv6 7×10^{23} IP címet tenne lehetővé négyzetméterenként.

Azonban a hosszú IP címek használatának van egy hátulütője: Hogyan tudják a rendszeradminisztrátorok ezeket a címeket hatékonyan ábrázolni és kezelni? Ezen probléma megoldására a következő jelölésrendszer lett bevezetve: a 16 bájtos címeket nyolc, egyenként négy hexadecimális számjegyből álló csoportként írjuk le a csoportok között kettősponttal, mint például 0000:0000:0000:0000:1234:5678:9ABC:DEFF.

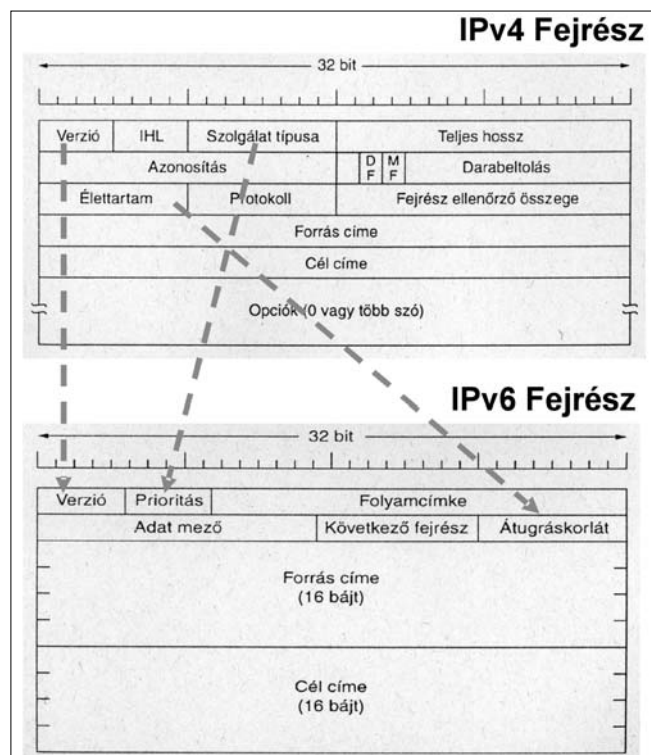
Mivel számos cím sok nullát fog tartalmazni, ezért három egyszerűsítés alkalmazható. Először is, egy csoporton belül a bevezető nullák elhagyhatók, így a 0345 345-nek írható. Másodsor, egy vagy több 16 nullából álló csoport két kettősponttal helyettesíthető. Így az előző példa a következő rövidebb formában írható:

::1234:5678:9ABC:DEFF.

És végül, az IPv4 címek két kettőspont, és a régi, pontokkal elválasztott decimális szám formájában írhatók le, például:

::129.132.66.157.

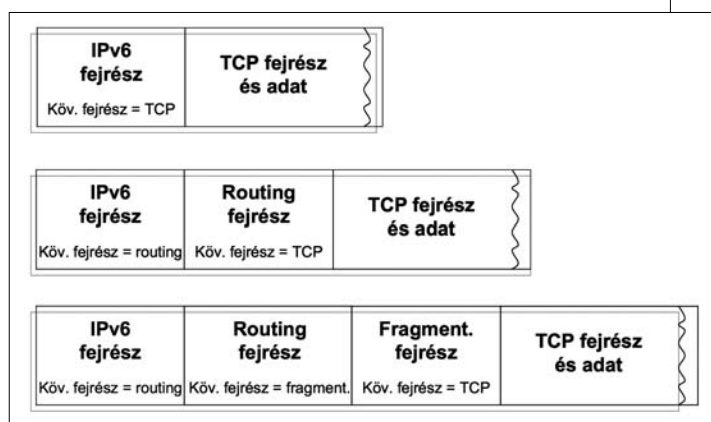
1. ábra Az IPv4 és IPv6 fejrész összehasonlítása



⁵ Az IETF egy nemzetközi szervezet, amely az Internettel kapcsolatos technológiák szabványosításáért felelős

Egy másik határozott előrelépés IPv6 használata esetén az IP fejrész jelentős egyszerűsödése. Ahogy azt már korábban említettük, minden egyes adatcsomagnak tartalmaznia kell bizonyos információt, hogy az útvonalválasztó csomópontok képesek legyenek a csomagok helyes irányba történő továbbítására. Ez az információ az IP fejrészben található. IPv6 esetén az IP fejrész mindössze 8 mezőből áll, ellentétben az IPv4 14 mezőjével. Ez az egyszerűsítés lehetővé teszi az útvonalválasztó csomópontok számára a beérkező adatcsomagok gyorsabb feldolgozását, így több adat átbocsátását, ezáltal növelve az útvonalválasztó kapacitását és hatékonyságát.

Az előző oldali 1. ábrán látható a két különböző IP verzió által használt protokoll fejrész, és a kapcsolat bizonyos mezők között, amelyeket többé-kevésbé átvett az IPv6 az IPv4-től. Egy érdekes dolog az úgynevezett *következő fejrész* (*next header*) mező használata. Ez az oka annak, amiért az IPv6 fejrész egyszerűsödött, hiszen ennek segítségével az IP fejrész szükség esetén kiterjeszthető további (opcionális) mezőkkel. A 'következő fejrész' mező határozza meg, hogy melyik fejrész következik a (jelenleg) 6 kiegészítő fejrész közül. Az utolsó fejrész esetén pedig ez a mező azt mondja meg, melyik transzport protokoll kezelőjének, például TCP, kell a csomagot továbbítani. A 2. ábra szemlélteti eme mechanizmus működését egy példával illusztrálva a fejrész kiterjesztések láncolt használatát.



2. ábra Láncolt fejrész kiterjesztések

Az IPv6 előrelépést jelent biztonság tekintetében is. A biztonságos adatátvitel szerves része az IPv6-nak, ellentétben az IPv4 esetén használt utólagos kiterjesztésekkel. A biztonságos adatátvitel magában foglalja az IP szinten történő adathitelesítést és adattitkosítást. A hitelesítés segítségével az adatcsomag címezhető biztos lehet abban, hogy a csomagot valóban a feladója küldte, és a csomag nem változott meg a továbbítás során. A titkosítás biztosítja, hogy csak az a vevő tudja elolvasni a csomagot, akinek azt valójában szántuk. Azonban még az IPv6 sem kínál megoldást hozzáférési jogosultságok kezelésére, azaz annak eldöntésére, hogy a felhasználónak a sikeres azonosítás után vajon van-e jogosultsága az erőforrások használatához. Ez továbbra is alkalmazás-szintű törődést igényel. Az em-

lített biztonsági megoldások implementálása egy titkos kulcsrendszer segítségével történik az IPv6-ban, azaz a küldő és fogadó állomásnak meg kell állapodnia egy vagy több titkos kulcsban – ami tulajdonképpen egy kód –, mely(ke)t csak ők ismernek, és ezen kulcs(ok) használatával történik később az adatcsomagok hitelesítése és titkosítása. Továbbá, a küldő és fogadó állomásnak egy biztonsági viszonyt kell kialakítania egymással. Ez a viszony tartalmazza a kulcsot, a hitelesítés és/vagy titkosítás algoritmusát, illetve egyéb paramétereket, mint például a kulcs érvényességi idejét.

A gyakorlatban a hitelesítés a *hitelesítő fejrész* (*authentication header*) – amely egy IPv6 kiegészítő fejrész – segítségével történik. A hitelesített adat – az adatcsomag feladójának címét és néhány más IP fejrész mezőt, mint bemeneti információt felhasználva – egy matematikai algoritmus alkalmazásával kerül előállításra, és az eredeti adatcsomaghoz csatolva kerül továbbításra. Az alkalmazott algoritmus az úgynevezett *kulcsolt MD5* (*keyed MD5 – Message Digest 5*), amelyet rendkívül nehéz visszafejteni és amely ezáltal biztosítja, hogy az adatcsomag sértetlen és a valódi feladójától ered. A hitelesítő fejrész azonban nem titkosítja magát az adatcsomagot. Erre a *titkosított biztonsági adatmező fejrész* (*encrypted security payload header*) használandó. Titkosítás használata esetén ennek a kiegészítő fejrésznek kell következnie rögtön az alapvető IPv6 fejrész után, hiszen csak így lehet biztosítani mind a hitelesítési információt, mind pedig az adatmező titkosítását. A titkosításhoz használt alapértelmezés szerinti eljárás a *CBC* (*Cipher Block Chaining – titkosított blokkok láncolata*), amely a DES (Data Encryption Standard) titkosítási szabványra épül.

A szolgáltatás minőségére (Quality of Service) vonatkozó kérdések tekintetében is előrelépést jelent az IPv6. Az IPv6 fejrész *prioritás* (*priority*) mezője (1. ábra) használatos azon csomagok közötti különbségtételre, amelyeknek a feladója képes forgalomszabályozásra, és amelyeknek nem. A 0-7-ig terjedő értékek azon adatkapcsolatokat jelölik – mint például honlapletöltés –, amelyek képesek torlódás esetén az átviteli sebesség csökkentésére. A 8-15-ig terjedő értékek az olyan valós idejű kapcsolatokhoz tartoznak, amelyeknek a küldési sebessége állandó és kicsi a késleltetés és késleltetés ingadozás az egyes csomagok között. Például hang és mozgókép átvitel tartozik ebbe a kategóriába. Ez a megkülönböztetés lehetővé teszi az útvonalválasztók számára, hogy a csomagokat hatékonyabban tudják kezelni torlódás esetén.

A *folyamcímke* (*flow label*) mező (1. ábra) még mindig kísérleti stádiumban van, de arra lehet majd használni, hogy a forrás- és célállomás létrehozasson egy virtuális összeköttetést bizonyos igényekkel és tulajdonságokkal. Az így létrehozott adatfolyamot előre fel lehet állítani a hálózatban. Minden folyamat egyedileg lehet azonosítani a forrás-, célcím és a folyamcímke által, így több folyamat is lehet aktív egyidejűleg két csomópont között.

Amikor egy nem nulla folyamcímekjű csomag jelenik meg egy útvonalválasztónál, az útvonalválasztó ki tudja keresni a belső táblázataiból, hogy a csomag milyen különleges bánásmódot igényel. Azonban a folyam szintű szolgáltatás minőségre vonatkozó eljárásoknak közismert skálázhatósági problémái vannak nagy hálózatokban, amely gondot okozhat az IPv6 folyamkezelés esetében is.

És végül, az IPv6 alapvető mobilitás támogatása – azaz annak lehetővé tétele, hogy a mozgó hosztok helyváltoztatásuk ellenére is az Internethez kapcsolódva maradhassanak – is egy újabb, jelentős előrelépés. Az IPv6 támogatja az IP címek automatikus konfigurációját mind állapot-követéses, mind állapotmentes módon.

Az állapot-követéses automatikus konfiguráció a DHCP (Dynamic Host Configuration Protocol – dinamikus hoszt konfigurációs protokoll) használatán alapul, akárcsak az IPv4 esetén, amely megköveteli DHCP szerverek telepítését és menedzselését a hálózatban. Az állapotmentes automatikus konfiguráció esetén nincs szükség ilyen szerverekre. Ebben az esetben a hoszt saját maga állítja elő az IPv6 címét a MAC (Medium Access Control – közeg hozzáférési kontroll)⁶ címnek és az úgynevezett *alhálózati prefixnek* (*subnet prefix*)⁷ a felhasználásával. A *szomszéd felderítési* (*neighbor discovery*) eljárások is tovább lettek fejlesztve az IPv6-ban.

Az összes olyan mechanizmus, amely az azonos hálózati szegmenshez csatlakozó hosztok és útvonalválasztók közötti interakcióval kapcsolatos, az IPv6 esetén egyetlen protokollba, mégpedig az ND (Neighbor Discovery – szomszéd felderítés) protokollba lett összevonva. Az ND felváltja az IPv4 által használt ARP (Address Resolution Protocol – cím feloldási protokoll), ICMP (Internet Control Message Protocol – Internet kontroll üzenet protokoll), RD (Router Discovery – útvonalválasztó felderítés) protokollokat, valamint további fejlesztéseket is tartalmaz. Ezen tulajdonságok elősegítik a mobil IP⁸ használatát IPv6 esetén is.

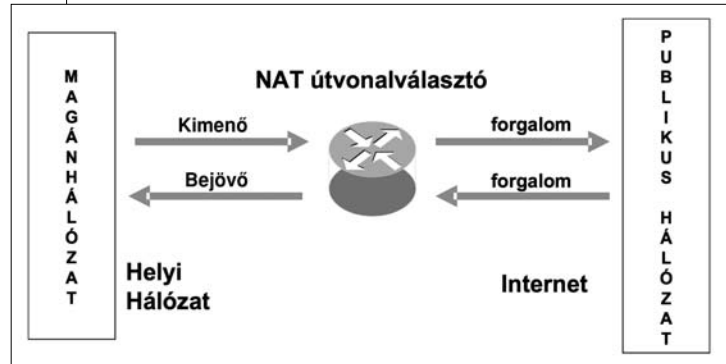
3. Az IPv6 a megoldás?

Most, hogy áttekintettük az IPv6 által bevezetett újdonságokat, feltehetjük a kérdést: Miért nem terjedt már el világszerte az IPv6, mikor számos előnnyel rendelkezik az IPv4-gyel szemben? Ráadásul az IPv6 lényegi részeinek a szabványosítása már befejeződött és az IPv6 Fórum⁹ [3] már évek óta támogatja és propagálja az IPv6 rendszeresített használatát.

Nem egyszerű válaszolni erre a kérdésre. Ehhez közelebbről meg kell vizsgálnunk, vajon az IPv6 által be-

vezetett újdonságok [4,5] valóban lényegi előrelépést jelentenek-e?

A gyakorlatilag kimeríthetetlen IPv6 címtartomány egy lényeges előrelépés. De valóban szükség van erre már manapság illetve a közeljövőben? Úgy tűnik, nem, mert a jelenleg használatban lévő, az IPv4 esetén alkalmazott IP cím-kímélő kiterjesztésekkel, mint például a NAT (Network Address Translation – hálózati címfordítás), és előrelátó címfogalási stratégiákkal az IPv4 címtartomány még biztosan kitarthat az elkövetkezendő pár év folyamán.



3. ábra NAT

Például, a nyilvános Internet és egy magánhálózat határán elhelyezkedő NAT útvonalválasztó használata esetén (3. ábra) mindössze pár nyilvános IP címre van szükség (gyakran csak egyre) ahelyett, hogy minden egyes hoszt a magánhálózatban egy külön nyilvános IP címmel rendelkezne. Ebben az esetben a NAT útvonalválasztó lefordítja, azaz kicseréli, a forrás IP címét minden olyan csomagnak, amely elhagyja a magánhálózatot egy, az útvonalválasztó által használható nyilvános IP címtartományból kiválasztott nyilvános IP címre. Ezen felül a NAT útvonalválasztó visszacseréli az Internet felől érkező csomagok cél IP címét a magánhálózatban lévő célállomás által használt magán IP címre. Ezzel a megoldással számos különböző magánhálózat használhatja ugyanazt a magán IP címtartományt a hálózaton belül – mert ezek a címek nem érhetők el a magánhálózaton kívülről, így nem okoznak címütközést –, ami növeli a kiosztható IP címek számát.

Azonban a NAT használatának is vannak korlátai. Például az olyan alkalmazásoknak, melyeknek egy NAT útvonalválasztón kell áthaladniuk, a magánhálózatból kell kezdeményezniük a logikai adatkapcsolat létrehozását, mivel az Internet felől csak bonyolult eljárásokkal – úgymint alkalmazás szintű DNS (Domain Name System – tartománynév rendszer) átjáróval – lehetséges ez, hiszen az Internet felől a NAT útvonalválasztó túloldalán lévő hálózat nem látható. Továbbá, összefüggő végpont-végpont (end-to-end) szintű csomagtovábbítás és biztonságos átvitel biztosítása is csak egyéb el-

⁶ A MAC cím általában az egyedi azonosítója az adott gép hálózati interfészének

⁷ Az alhálózati prefix annak a hálózati szegmensnek az azonosítója, amelyhez a hoszt csatlakozik és amelyet az úgynevezett szomszéd felderítési (neighbor discovery) eljárás segítségével kérdez le az adott szegmenshez csatlakozó útvonalválasztótól

⁸ A mobil IP egy protokoll szabvány mobilitás kezelésére az Internet hálózatban

⁹ Az IPv6 Fórum egy világméretű, az Internettel kapcsolatban lévő cégek és kutatási/oktatási hálózatok non-profit konzorciuma, amelynek célja az IPv6 népszerűsítése

járások segítségével lehetséges – mint például úgynevezett *alagút (tunnel)* használatával –, mivel alapesetben a logikai kapcsolatok végpontjai a NAT útvonalválasztóval, és nem közvetlenül egymással kommunikálnak.

Az IPv4 címtartomány kimerülésének tempóját vizsgálva a következőket figyelhetjük meg: 1994 előtt megközelítőleg 36%-a volt a rendelkezésre álló IPv4 címtartománynak kiosztva. Azóta további 20% lett felhasználva [4], és ez az időszak magába foglalja a teljes Internet robbanás időtartamát is! Amennyiben ez a tendencia folytatódik, akkor úgy tűnik, hogy még évekig ki fog tartani az IPv4 címtartomány.

A biztonságos kommunikáció is egyre nagyobb figyelmet igényel manapság. Míg az IPv6 esetén a biztonságos kommunikáció lehetőségét már a protokoll tervezésétől kezdve figyelembe vették, addig ez az IPv4 esetén csak különböző kiegészítésekkel oldható meg, mint például az IPsec használatával, amely egy protokoll gyűjtemény a biztonságos adatcsere lehetővé tételére. De valószínűleg mennyire fontos az, hogy a biztonságos kommunikációt a hálózat implementálja? Egy komoly alkalmazás, amely biztonságos adatátvitelt igényel, saját maga implementálja azt, és nem hagyatkozik a hálózatra, legalábbis manapság még nem.

A szolgáltatás minőség vagy QoS támogatás tekintetében még mindig az úgynevezett *túlkínálat (overprovisioning)* a legegyszerűbb és legolcsóbb megoldás az Internet hálózatban. Ez leegyszerűsítve azt jelenti, hogyha valahol a hálózatban több kapacitásra lenne szükségünk, akkor egy újabb kábel hozzáadása vagy az optikai szálak számának növelése egy összeköttetés mentén gyors és egyszerű megoldást kínál ellentétben bonyolult QoS irányelvek és technikák alkalmazásával. Továbbá, a korábban említett, IPv6 által támogatott folyamszintű QoS skálázhatósági korlátai még gondot jelenthetnek a jövőben.

És végül, a rohamosan növekvő, az Internethez csatlakozó mobil eszközök számával a mobilitást támogató megoldások egyre jobban előtérbe kerülnek. Az állapotmentes automatikus konfiguráció és a továbbfejlesztett szomszéd felderítési eljárások jelentős segítséget nyújthatnak IPv6 használata esetén. Azonban az IPv6 önmagában nem szolgál teljes megoldással erre a problémára, így néhány további protokoll használatára is szükség van (mint például a mobil IP-re, ellenben a mobil IP képes IPv4-gyel is együttműködni).

4. Áttérés az IPv6-ra?

Ezután a rövid elemzés után úgy tűnik, hogy a fő motívációt az IPv6 világméretű elterjedésére és használatára az a szituáció fogja jelenteni, amikor majdnem minden eszköz a hűtőgéptől a mikrohullámú sütőig az összes mobil eszközünkkel egyetemben az Internethez fog kapcsolódni, és egyedi IP címet fog igényelni. Azonban most még nehéz megjósolni, hogy ez mikor fog bekövetkezni. Természetesen ez nem egyik napról

a másikra fog végbemenni az átállás hatalmas terhei (idő, munka, pénz stb.), a szükséges működtetési tapasztalat hiánya és azon rizikó miatt, amelyet egy évtizedek óta működő rendszernek egy olyan rendszerrel való felváltása jelent, amellyel kapcsolatban vajmi kevés tapasztalattal rendelkezünk.

Mindamellet egy idő óta jó néhány kísérleti, sőt már kereskedelmi IPv6 hálózat is működik világszerte. Ebben a tekintetben Ázsia, különösen Japán és Dél-Korea viszi a vezető szerepet, de Európában is jelentős számú, IPv6-tal kapcsolatos kutatási projekt fut.

Svájcban például a SWITCH, a svájci oktatási és kutatási hálózat [6] nyújt IPv6 használata által való kapcsolódási lehetőséget az Internethez, már 1996 novembere óta. 2004 júniusától a SWITCH „SWITCHlambda” nevezetű gerinchálózata IPv4 és IPv6 támogatást nyújt egyidejűleg ugyanazon hálózati elemek használatával. A Swisscom Innovations, amely a központi innovációs részlege a Swisscom nevezetű svájci telekommunikációs cégnek, már jó ideje futtat IPv6 teszt hálózatot, és jelenleg belső Intranet hálózata is alkalmas IPv6 használatára [7].

Az USA is kezdi komolyan fontolóra venni az IPv6-ot. Az amerikai védelmi minisztérium 2003-ban írásba is fektette azon célját, mely szerint a teljes IP hálózatán IPv6-ra kell átállnia 2008-ra [8]. És végül, Magyarországon is lehet IPv6 címtartományt regisztráltatni az NIIF-nél (Nemzeti Információs Infrastruktúra Fejlesztési Program) [9], illetve működő teszthálózatok is találhatók, például a Budapesti Műszaki és Gazdaságtudományi Egyetemen [10] vagy az NIIF gondozásában [9].

Figyelembe véve a jelenlegi tendenciákat úgy tűnik, hogy az IPv6 világméretű elterjedésére 2008-2010-re lehet számolni. Meglátjuk...

Irodalom

- [1] IETF – www.ietf.org
- [2] A. S. Tanenbaum: Computer Networks, Prentice-Hall Inc., ISBN: 0-13-394248-1, 1996.
- [3] IPv6 Forum – www.ipv6forum.com
- [4] G. Huston: Waiting for IPv6, in the ISP Column, January 2003, <http://ispcolumn.isoc.org/2003-01/Waiting.html>
- [5] L. Ladid, J. Bound: Response by IPv6 Forum to ISP Column article entitled 'Waiting for IP version 6', <http://www.isoc.org/pubs/isp/ipv6response.shtml>
- [6] SWITCH IPv6 Pilot – www.switch.ch/network/ipv6/
- [7] Swisscom Innovations – IPv6 Labs and Services, <http://www.swisscom.com/Innovations/content/Labs/IPv6/>
- [8] DoD IPv6 General Information – <http://ipv6.disa.mil/>
- [9] NIIF/HUNGARNET IPv6 Projekt – <http://6net.iif.hu/>
- [10] BME – <http://portal.bme.hu/>

SDL:

A kommunikációs folyamatmodellek egyik szabványosított implementációs eszköze

MEDVE ANNA

Veszprémi Egyetem, Műszaki Informatikai Kar, Információs Rendszerek Tanszék
medve@almos.vein.hu

Kulcsszavak: SDL, UML2.0, FMT, formális nyelvek, ITU-T szabványok, ITU- nyelvek

Célunk a *Specification and Description Language (SDL)* formális módszertan modellező képességeinek ismertetése. A cikk motivációjának alapja, hogy az *Unified Modeling Language (UML)* modellező nyelvnek a '99 óta tartó fokozott közelítése a távközlés világában kifejlesztett SDL nyelvhez oly módon jelenik meg a fejlesztők mindennapjaiban, hogy az UML2.0 szabvány magába emelte a dinamikus viselkedések leírására az SDL nyelv teljes implementációs eszköztárát. Jóllehet a fejlesztőkörnyezetek gyártói közreadják az egyes terület-specifikus UML2.0 nyelvi implementációkba beépített lehetőségek segédletét, az előzőekkel szakító szemléletű UML2.0 verzió megismerésében jól hasznosítható az SDL viselkedés-modellező jellemzése.

1. Bevezetés

A távközlés digitalizálása és a számítástechnikával való konvergenciája, valamint az Internet hatásaként létrejött újszerű üzleti modellek visszahatnak a távközlési rendszerek fejlesztésére: az emelt szintű szolgáltatások (value added services) egyre dinamikusabbak és egyre gyorsabb fejlesztési folyamatot igényelnek, fokozottabb jósággal.

A kommunikációs szolgáltatások sajátossága napjainkban, hogy azok különböző infokommunikációs technológiákon alapulnak: a hagyományos nyilvános kapcsolt és mobil távközlő technológiák, vagy az Internet technológiák, vagy ezek keverékei által. Ugyanakkor nem mindig lehetséges élesen elhatárolni a kommunikációs szolgáltatásokat a kommunikáló protokolloktól, és miközben e téren a fejlesztés több évtizedes tapasztalatai kifejezetten a protokollfejlesztésben jegyeztek, aközben napjainkban az üzleti folyamatok szabályainak tervezésében és az információs rendszerek strukturálásában is a kommunikációs modell kerül előtérbe. Magyarázata lehet a fejlődés törvényszerűsége, oka a szoftverkrízis, amely kihat a globális piacon lévő termékek együttműködésére és átmeneti minőségcsökkenéssel jár. A megoldást várhatóan a szabványosítás és a szoftverfejlesztés kínálja [3,4]. A kommunikáló rendszerek működtetésének eszközei a *kommunikáló protokollok és kommunikációs szolgáltatások* [1,2], melyek modellezésére az International Telecommunication Union – Telecommunications Standardization Sector (ITU-T) a Z. formális nyelvek szabványcsaládját fejlesztette ki.

Annak ellenére, hogy a formális módszereket az iparban még kevésbé használják, a kutatók a jövő elkerülhetetlen szoftverfejlesztő eszközeinek tartják. A tapasztalat szerint az ipar olyan fejlesztési módszertant fogad el, amely a szükségletei szerint alkalmazható, kereskedelemben forgalmazott, fejlesztői környezettel jól támogatott, valamint fontos szempont annak szabványosítása is [9]. Az *UML (Unified Modeling Language)* félfor-

mális modellező nyelv az *Object Management Group (OMG)* gondozásában ezt az igényt táplálja és kínál megoldásokat egyben.

A távközlési korlátokat lebontó informatikai megoldások szoftvertechnológiája oly módon hatott az UML fejlődésére, hogy a konkrét fejlesztőkörnyezetű UML profilok terület-specifikus fogalomkategóriáira a fejlesztőkörnyezet-forgalmazók tudásbázis-alapú fejlesztéstámogató módszereket építettek be a távközlés területén bevált eszköztárból, kezdetben az integrációs tesztek támogatására. A felhasználóközpontú értéknövelt szolgáltatásokat támogató gyors szoftverfejlesztés technológiáiként előtérbe kerültek a távközlés területén korábban kifejlesztett komponens-alapú és modell-vezérelt modellező technológiák, az automatikus kódgenerálással együtt. A hazai könyvkiadásban égető hiány pótlásaként jelent meg az UML 2 modellező nyelvi kézikönyv [32], amely a szerzőtől már megszokott módon tárgyalja és segíti hozzá az UML-el ismerkedőket és UML-t már alkalmazókat is a nyelv lényegének megértéséhez és hatékony alkalmazásához.

Mivel az OMG kifejezetten a szoftverfejlesztés folyamatára optimalizálja az objektumtechnológiák fejlesztését, az UML fejlődésében és alkalmazás-támogató technológiáiban egyre hangsúlyosabbak és látványosak a távközlésben kidolgozott gyakorlatban jól bevált módszertanok. A cikkben ezek egyikének jellemzését, a dinamikus viselkedést leíró *Specification and Description Language (SDL)* formális nyelv tulajdonságait ismertetjük, mint a távközlés világában kifejlesztett és automatikus kódgenerálásra alkalmas szabványosított implementációs formális módszertant.

2. Az SDL specifikáló és leíró nyelv

Az SDL nyelvet az ITU-T a Z.100 -as ajánlásaiban dolgozta ki komplex rendszerek specifikálására, oly módon, hogy az eseményvezérelt, valósidejű és interaktív

folyamatok konkurens módon diszkrét jelekkel kommunikálnak [10-13].

Jóllehet manapság az SDL nyelv használható bármely valós idejű rendszer specifikálására és implementálására, gyökerei a távközlésben vannak.

Az SDL fejlesztése egy 1968-as ITU tanulmányból indult ki a programvezérlésű kapcsolórendszerek kezelésére vonatkozóan, aminek eredményeként 1972-ben egyetértés született arról, hogy nyelvek szükségesek a gépek és berendezések interakcióinak leírására és programozására. 1976-ban közzétették a leíró nyelv szabványát, egy alap grafikus leírónyelv fejlesztésével, majd négyévenként további fejlesztéseket jelentettek meg.

1980-ban a processz szemantika meghatározásával, 1984-ben struktúrák és adatok hozzáadásával 1988-ban konkrét eszközként jelentették meg az SDL-88-at, egy jól meghatározott szintaktikájú Meta IV formális jelöléssel támogatva a beszélt nyelvi leírást. Az 1992-ben bevezetett típuskonstrukció az SDL-t objektum-orientáltá tette és az 1996-ban az Addendum szabványkiegészítés pedig a hatékonyabb fejlesztőeszközöket eredményezte.

Az SDL eszközök piaca 1996-2000 között jelentősen növekedett. A felhasználók nyomására az *SDL formális* jellegét fokozták, ami a tervezésben a nagyobb pontosságot, megfelelést, és átláthatóságot biztosítja. A számítógépes grafika és a beépített tudástárak fejlődésével lehetővé vált mára az SDL eszközök nagyfokú funkcionalitása és a bizonyítottan jó, eredményes nyomkövetése. Az ipari alkalmazásának elterjedését legfőképpen a megvalósításnak a specifikációból való generálhatósága fokozta: a fejlesztőkörnyezetek generálnak programozási nyelvekre forráskódokat (általában C/C++), amelyeket be lehet szerkeszteni a valós idejű rendszerek termékgyártásába. A generált C++ kód kezelése közbülső nyelvű elemként történik, mint ahogy a kompájlerek kezelik a gépi kódot. Erőssége az eszközöknek, hogy absztrakt módon is lehet használni az SDL nyelvet a megvalósításhoz szükséges követelmények informális leírására [5,6,29].

Az SDL nyelvet 1996 óta használják a távközlési iparon kívül is, elsősorban az orvosi felszerelések iparágában, az autó- és repülőgépiparban.

A hazai kutatások az SDL alkalmasságát célozzák meg a hardver-szoftver együttes tervezésére, és jelentős eredmények születtek a kódbázisú (újabb specifikáció-bázisú) tesztelés és validálás terén az automatikus tesztgenerálásra és szelektálásra adott számos algoritmussal [7,8], a konformancia teszteléssel összefüggésben, és a formális módszertani összefüggésben [1,9].

2.1. Az SDL alkalmazásának jellemzői

Az SDL elsősorban a reaktív és diszkrét rendszerek esetében alkalmazható:

- *specifikálásra*: rendszer szolgáltatásainak specifikálása, a rendszer belső szerkezete nélkül, főleg az interfészek és a kommunikáció jellemzése fontos,

- *tervezésre*: hardver- és szoftverfejlesztők által (bővítőmennyiségű folyamat, a rendszerspecifikáció transzformálásával, valamint a belső szerkezetre és működésre vonatkozó információkkal)

- *megvalósításra*: rendszerintegráló mérnökök által, cél: a végső tervezésből futtatható rendszert készíteni (automatikus: a tervezéshez futás-idejű információkat kell hozzáadni),

- *dokumentálásra*: felhasználók és rendszerfenntartók számára (az aktuális működésről).

Az SDL alkalmazását megkönnyíti az egymásnak kölcsönösen megfeleltethető grafikus és szöveges nyelvi implementációja, a jól meghatározott fogalomköre, a tiszta, egyértelmű és pontos specifikálás következetessége. Az SDL egyik legfontosabb jellemzője a formálissága, amelyre alapozható a megvalósítás konformanciája és az automatikus kódgenerálás. Az SDL nyelv hierarchiája és formálissága kínálja a spirális fejlesztés menetét, ezzel áthidalja a hézagot a specifikálás és megvalósítás között, biztosítva az absztrakt szintű modellezést a megvalósításhoz szükséges részletezett-séggel.

Az SDL specifikációk jóságához szükséges az *MSC (Message Sequence Charts)* nyelv használata, amely kezdetben az SDL nyelv alkalmazását megadó szabvány része volt, 1992-ben választották külön szabványba, amit az 1996-os négyévenkénti soronkövetkező fejlesztés az in-line kifejezések és adattípus definíciók hozzáadásával a követelménytervezésben sokoldalúan alkalmazhatóvá tett [14,17,18].

Az MSC nyelv szerkezeteivel az eseményekre forgatókönyvet szerkeszthetünk, amelyek sorozatával informálisan leírható a megfigyelt viselkedés üzenetcsere sorozataként. A megfigyelt viselkedések egy halmaza valahány rendszerfolyamatot ábrázol, valamint általuk megjeleníthetők a rendszerentitások közötti kapcsolatok (interakciók), ezáltal a forgatókönyvek sorozatai segítik a rendszerentitások hierarchiába szervezését, valamint a dinamikus viselkedés formális analízisét, ezáltal a két nyelv komplett megoldást szolgáltat a specifikáláshoz és megvalósításhoz.

Az adatok ábrázolására SDL-ben az *ASN.1 (Abstract Syntax Notation One)* adatleíró nyelvet érdemes alkalmazni [15]. Ennek nagy előnye a nyílt rendszerek ábrázolásakor érvényesítődik az együttműködőképesség fokozásával, a zárt kommunikációs rendszerekben az SDL beépített típusgeneráló eszközeivel érdemes élni, főleg ha gyakori a rendszerváltozatok továbbfejlesztése, mivel az absztrakt típusok művelet-definíciós algebrai eszköztára jól hasznosítható az algoritmikus leírások és a kommunikációs folyamatok elkülönítésére az implementáció során, ezáltal a fejlesztői változtatások egyértelműsítésére is.

2.2. Az SDL szabványai

Az SDL nemzetközileg szabványosított nyelv. A nyelv fejlesztése az IUT-T ajánlások Z sorozatában közreadott szabványokban követhető, revíziójuk folyamatos. A fő szabvány a Z.100, amely tartalmazza az SDL de-

finícióját precíz és tömör módon. Nyelvi referencia, annak ellenére, hogy nem javasolt a nyelv tanulására. Nem tartalmaz példákat, de jól mutatja a szintaktikai szabályokat [19].

A Z.100-as szabványban a konkrét szintaktikai szabályok BNF dialektusban (MetaIV absztrakt nyelvtanban) definiáltak. Mindegyik szöveges nyelvtani elemnek (SDL/PR) adott a grafikus megfelelője (SDL/GR), formális és absztrakt szintaxisként, amelyhez kapcsolódik egy végrehajtásmodell. A formális szemantikát az F melléklet tartalmazza, amely egyenértékű a Z.100-as szabványban természetes nyelven megadott „Szemantikus leírás és Modellek” fejezettel. További szabványok terjesztik ki a Z.100-as szabványt a nyelv továbbfejlesztéseivel.

A Z.105-ös szabvány megadja az ASN.1 modulok kapcsolását az SDL adatleírásokba direkt módon, azaz a Z.105 megadja az ASN.1 használatát SDL-ben.

A Z.106-os szabványt 1996-ban az egyes fejlesztői környezetek közötti átjárhatóságra kiemelt, SDL/PR CIF néven találjuk meg a hétköznapiakban.

A Z.107-es szabvány kiegészíti a Z.100 és Z.105-ös szabványokat, az ASN.1 adattípusok használhatók az SDL adatdefiníciókban, azaz külön szabványban rögzítik az ASN.1 beágyazását az SDL nyelvbe.

A Z.109-es szabvány definiálja az SDL-nek megfelelő UML profilt, így hozzáadja az UML elv szerinti ábrázolást. (A Z.120 hasonló a Z.100-as szabványhoz az MSC definiálására.

A nyelv jellemzése

A nyelv fő szerkezeteit az objektum-orientált *absztrakt típusok* jelentik, oly módon, hogy maga az SDL típusok definíciója tartalmazza a szoftverfejlesztés mérnöki megközelítését [20].

Az *OO-elv az SDL-ben* a típusdeklarációk által érvényesül úgy az öröklődés, mint a specializációk vonatkozásában.

Az öröklődéshez a típusdeklarációk elhelyezhetők a rendszerstruktúra bármely szintjén, ha a fastruktúrában érvényesítettük az öröklődést. Abban az esetben, ha más rendszerekkel is megosztjuk a deklarációkat, akkor *package* egységekben a rendszeren kívül helyezzük el a deklarációkat. A specializációk szerkesztéséhez altípusokat szerkeszthetünk a főtípushoz adott új tulajdonságokkal (például újabb állapotátmenetek a *process* típushoz, újabb processzek a *block* típushoz), vagy a főtípusban megadott virtuális típus és virtuális állapotátmenet újradefiniálásával.

Az ábrázolandó rendszer szerkezetének és viselkedésének együttes leírására az SDL szerkezeti (strukturális) típusai adják a nyelvi elemeket. A *szerkezetet* hierarchikusan a *rendszer*, *blokk*, *processz* és *eljárás* egységek ilyen sorrendű egymásba ágyazása jelenti a *system type*, *block type*, *process type*, *procedure type* szerkezeti típusok megadásával, amint az 1. ábrán látható.

Az ábrázolandó rendszer felosztásának célja az *átláthatóság*, a részletek alsóbb szinteken taglalhatóságával; a *funkcióorientáltság*, a természetes funkciók követhetőségével; a *modellezhetőség*, az interfészek és kapcsolatok ábrázolhatóságával; az *újratervezhetőség*, az alrendszerek megvalósításainak különböző hardver és szoftver környezetekbe ágyazhatóságával; az *újrafelhasználhatóság*, a már létező modulok felhasználásával; a *tesztelhetőség*, a szoftverelemek belső kommunikációjának statikus ábrázolásával; a *karbantarthatóság*, a „fentről-le” felbontással valamint az interfészek és kapcsolatok ábrázolhatóságával, amely egyben a fejlesztőcsoport tagjainak együttműködését biztosítja, a fejlesztés alatti karbantarthatóság mai irányzatainak megfelelően; az *üzembiztonság*, a szerkezeti típusok és a többi típusok közötti – főleg a kommunikáció-, kapcsolat- és viselkedéstípusok közötti – formális szemantika támogatásával. SDL nyelven folyamatorientált, időben diszkrét rendszermodellezést végezhetünk [21].

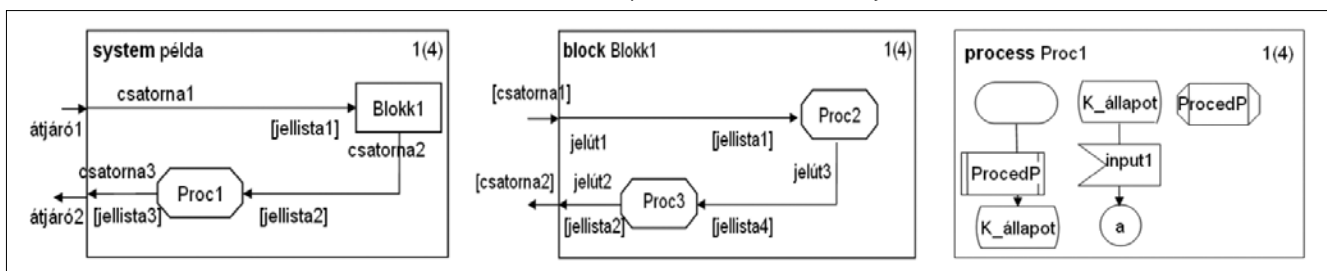
A stabil rendszerarchitektúra megvalósításához az SDL támogatja a távoli eljárshívást, a rekurzitást, valamint a típusok lokális és globális hatókörét az egyes állapotgépek külön-külön memóriaterületével is.

Az SDL-ben az elemi működési funkciók láncolatai a kommunikáló automaták állapot-átmenetei, a rész-funkciók közötti kölcsönhatások pedig a funkció szerint hierarchiába szervezett automaták kapcsolatai, rendszer-, blokk- és processz szinten. A kommunikáló automatákat a *process* szerkezeti és viselkedési típussal ábrázoljuk. A *process* szerkezeti típus az automatáknak a rendszer egészéhez való kapcsolódásainak jellemzőit rögzíti, míg a *process* viselkedési típus, mint látni fogjuk, az elemi funkciókat ellátó automaták működését ábrázolja.

A rendszer dinamikus *viselkedését* a processzek kommunikációja határozza meg, a *processzek kommunikációját* az üzenetküldések és üzenetfogadások adott módú leírásával adjuk meg.

A jelszerét az üzenetküldések és üzenetfogadások valósítják meg a processz (az automata) erre alkalmas

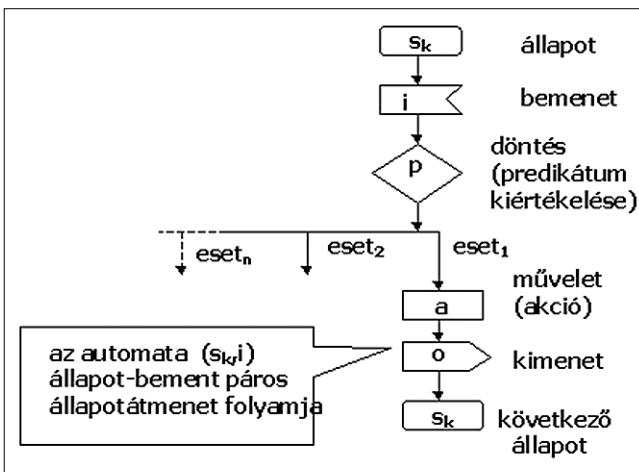
1. ábra Az SDL specifikáció hierarchiája



állapotaiban, ezáltal a *dinamikus jelleg leképezhető az állapotok változásával*.

A kommunikáló automatáknál az állapotok változása gerjesztés-válasz elven történik, és állapotátmenet a neve. Az *állapothoz (state)* kapcsolhatók az adott állapotban fogadott *bemenőjelek (input)* és az adott állapotban a bemenőjelhez társított állapotátmenet leírása. Az *állapotátmenet (transition)* megadja egy { állapot, bemenőjel } párhoz tartozó műveletsorozatok halmozát és az esetleges kimenőjelet valamint az új állapotot. Az állapotátmenet megadása többféle szerkezetben lehetséges, ez adja egyben a nyelv erejét is. A műveletsorozat neve *akció (action)* és *feltétel (decision)* alapján is megválasztható. Az állapotátmenetben az akciót követheti üzenetküldés is és maradhat az automata ugyanabban az állapotában is (2. ábra):

2. ábra Állapotátmenet



A **kommunikáció eszközeinek leírására** szolgál a *csatorna és jelút*, az *aszinkron jelek* és a *távoli szinkron eljárás* típusai (1. ábra).

A FIFO elven működő csatornákon és jelutakon, a kommunikációs mechanizmus egy vagy kétirányú lehet, és a hordozott paraméterekkel információcsere és szinkronizáció valósul meg az SDL processzek között, valamint az SDL rendszer és környezete között (nem SDL-szerűen viselkedő alkalmazás, vagy másik SDL rendszer).

Az SDL a jelutak és csatornák kombinációjának szerkezetével jól meghatározott interfészeket definiál a blokkok és processzek között. A kommunikációs típusokhoz nem rendelhető prioritás, ezt a megvalósítás szintjén a fejlesztő teheti meg.

Az SDL gyengesége, hogy a kommunikációs csatornák és jelutak állapotának kezelésére nem ad definíciós eszközt, a megvalósítás szintjén pedig a bejövő jelek feldolgozásánál megengedett jelsorrend csere a SAVE mechanizmussal jelentősen rontja a tesztelhetőséget [22].

Egy jel egyszerre csak egy meghatározott processzpéldánynak küldhető, a többesküldésre nincs szintaxis, a többesküldést mint funkciót a fejlesztőnek kell megterveznie.

Az **adatok leírására az SDL nyelvben** az *absztrakt adattípus definíciót (ADT)* és/vagy az *ASN.1 adatleíró nyelvet* kell használni, azaz az SDL nem tartalmaz adattípusokat, hanem eszközöket ad a fejlesztőnek bármely adatszerkezet megszerkesztésére.

Az *absztrakt adattípusok (ADT)* olyan típusok, amelyeknek nincs adatszerkezetük, hanem az adat tulajdonságainak a leírására való szerkezetekkel és a fejlesztő kényelmére előredefiniált, a programozási nyelvekből ismert egyszerű és összetett alap-típuskészlettel bármely kommunikált adat leírható. Rendkívül előnyös, mert a fejlesztő specifikálhatja ily módon, az adattípus jelkészletét, műveleteit, függvényeit is, bármely terület-specifikus jellemvonás ábrázolható, mi több, az algoritmikus, számításos műveletek típusműveletekbe szerkesztve lerövidülnek, jól áttekinthetővé válnak az állapotátmenet leírások.

A másik lehetőség az adattípusok definiálására, az *ASN.1 absztrakt jelölő nyelv*, amely a magas szintű nyelvek adatdefiníciójához ad átjárást és lehetőséget a létező adatszerkezetek újrafelhasználásához. SDL verziótól és fejlesztőkörnyezettől függ, hogy deklaráción szinten vagy importálással alkalmazhatók az ASN.1 definíciók [12,15].

2.3. Az SDL jelene és jövője

Manapság két fő irányban végzik az SDL-2000 fejlesztést, és pedig az objektummodellezés és a termékfejlesztés irányában.

Az objektummodellezés jobbító eszközeiként az SDL-2000-be az interfészeket és az ágens elvet vezették be egységesen mindenik SDL architektúrási elemre, közelítve ezzel az UML sztereotípusokhoz. A termékfejlesztés jobbításához az adatmodell revíziója, a kivételkezelés és a diagramokon belüli szöveges algoritmusok bevezetése jelentős. Továbbra is fennmaradó gyengesége a nyelvnek az időkezelés és a szinkron műveletek kezelése.

Az SDL nyelv fejlődése a nyelv széleskörű alkalmazásához vezetett az ipari automatizálás területein. A beépített alpműveletek mellé, adatdefiníciós eszközökkel saját típust és műveleteit szerkeszthetjük meg az állapotátmenetek műveletsorozatainak programozásához.

Az SDL nyelv kifejezőereje nagy és rendkívül kényelmes a kommunikáció-vezérelt rendszerek modellezésére, úgy véljük, ez lassítja a matematikai módszerek alkalmazását az újabb verzió formális szemantikájának automatikus helyességbizonyítására és szintézisére (hasonló a jelenség ahhoz, mint amit a VHDL nyelv kialakulásakor tapasztalhattunk) [9].

Az SDL objektum-orientáltsága elérte a napjainkban elvárt objektum-orientált szintet az SDL2000 verzióval, mondhatnánk azt is, hogy a modellfejlesztés általános paradigmái ontológikusan is utolérték az SDL modellezésben de facto létező technikákat, most újabb kihívás az ontológiák alapján kimutatni az SDL-ben eredendően jelenlevő MDA paradigma – megfelelést [23].

2.4. SDL fejlesztőkészletek

Az SDL fejlesztőkörnyezetek legtöbbjének közös funkciói a grafikus szerkesztő, a szöveges és grafikus konverzió, a statikus elemző, a kódgeneráló, a szimuláló és validáló a dinamikus elemzés, az MSC-vel kombinált támogatás.

A Telelogic TAU család egy ipari érvényességűnek (industrial-strength) tartott fejlesztőkészlet, amely a távközlési nyelvek családját támogatja és fejlesztéskövető eszközöket is kínál. Erőssége az MSC, SDL, TTCN modellek környezettámogatásos összeláncolhatósága, ugyanakkor az UML kiindulású folyamatra két fejlesztőeszközt forgalmaz – egyik a rendszerelemzésre, másik az UML2.0 alapon a teljes életciklusra használatos [24].

A Tau SDL Suite 4.6-ös verzióján tart, szolgáltatása a szerkesztés közbeni, tudásmenedzsment alapú szintaktikai ellenőrzés, a szimulációs modellvégrehajtás, a holtpontok és specifikálatlan bemenetek jelzése az állapot-robbanás módszerével, oly módon, hogy a hibakövetéshez MSC nyomvonalat (trace) generál. A TAU család tartalmazza a TTCN fejlesztőkészletet, amellyel SDL modellből is nyerhetünk teszt sorozatokat. A 4.0 verziótól lehetséges az UML modelleknek SDL modellbe transzformálása, majd a validálás, kódgenerálás esetleg teszt sor-generálás SDL alapon. A TAU támogatja a csapatmunkát, akár egyazon időben egyazon modellen is. A TAU család támogatja végrehajtható kód generálását számos valós idejű operációs rendszerben [25].

További SDL szabványt támogató termék a jelenleg Telelogic termék, az ObjectGeode hasonló a Tau SDL-hez, jóval több adattárral és konvertálóval támogatja a valós idejű rendszerek fejlesztését az ipari automatizálásban. A SOLINET cég SDL fejlesztőkészlete valós idejű rendszerek fejlesztéséhez ad támogatást, akadémiai körökben és ipari automatizálásra alkalmazott a Cinderella, demo változata letölthető [26].

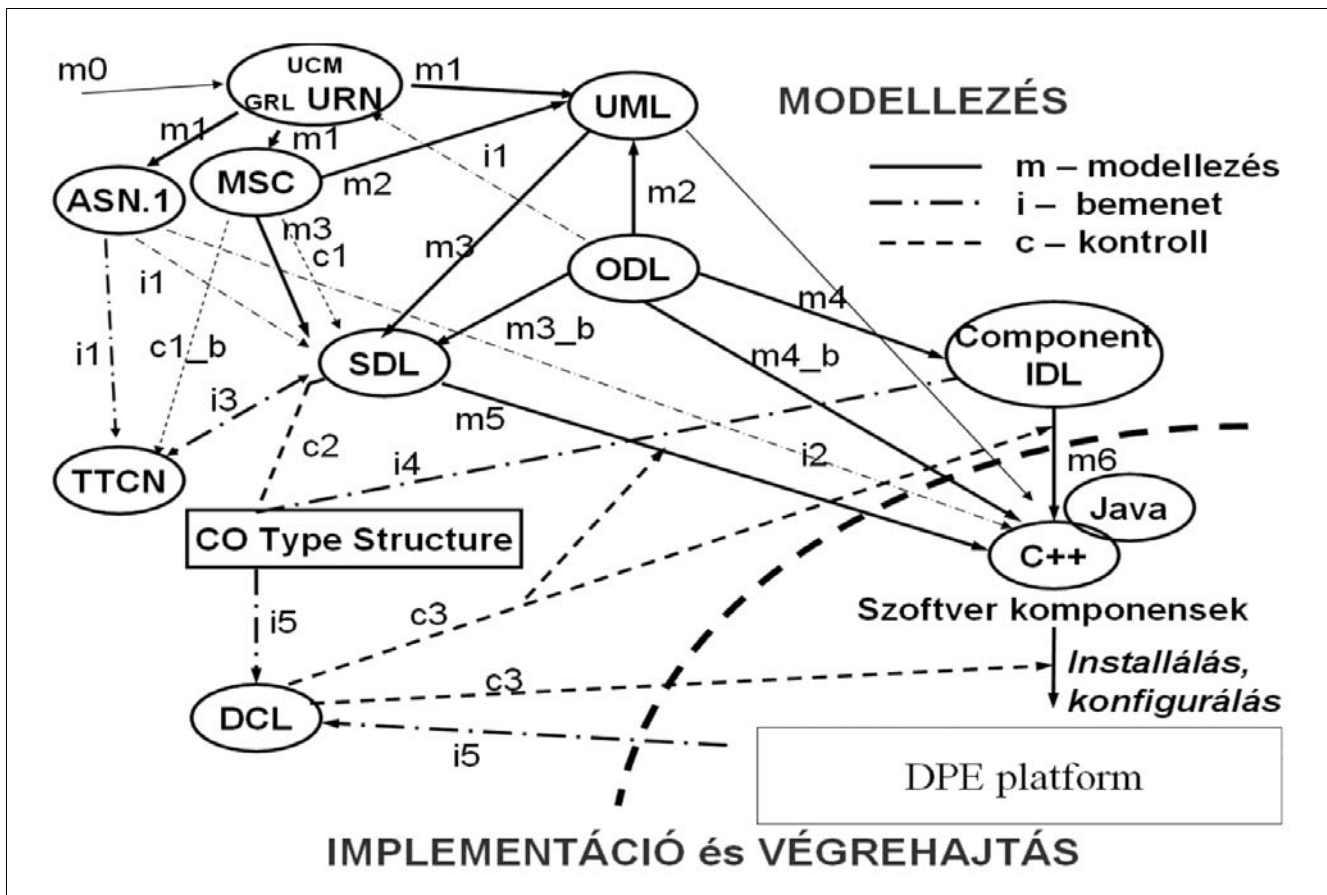
3. A távközlés szabványosított formális nyelvei a fejlesztés fázisaiban

Teret nyertek a fejlesztésben a több paradigma mentén és több nyelvi eszközt használó módszertanok [31]. Az ITU-T és az ETSI munkacsoportjai javasolnak követésre módszertant a távközlési szoftverek fejlesztésére a szabványosított eszközök alapján, de jellemzően a legtöbb módszertan és fejlesztőkészlet ipari környezetben jött létre.

Az ITU-T nyelvek fejlesztőcsoportja meghatározta a System Design Language néven a Z. szabványcsalád azon elemeit, amelyek a teljes fejlesztési folyamatot támogatják.

Az ITU-T szabványkészlete és módszertana az URN-MSC-UML-SDL-TTCN szabványokra épülő inkrementális módszertant ajánlja a munkacsoportjainak eredményeit ismertető konferencianapokon [19].

3. ábra Formális nyelvek egymásra épülése a fejlesztés fázisaiban [20]



A *User Requirements Notation (URN)* szabványa két formális nyelvet, a *Use Case Map (UCM)* és *Goal Requirements Language (GRL)* formális nyelveket tartalmazza, a funkcionális illetve a nem-funkcionális követelmények ábrázolására.

Az UCM leképezhető MSC diagramokra, amelyek többféle módon építhetők be az UML modellbe szekvencia diagramként, az SDL modellbe a statikus és dinamikus leírás során. Az ASN.1 deklarációkat a fordító lefordítja SDL típusra, gyorsítja és egyszerűsíti az átjárást a *Test and Test Control Notation (TTCN-3)* modulokkal a validálás és tesztelés generálás folyamataiban.

A rendszer életciklusaiban alkalmazható formális nyelvek kapcsolatát szemlélteti 3. ábra (az előző oldalon). Tükrözi azt az egységesítési folyamatot, amely jelentősen növeli a fejlesztés hatékonyságát és egyben kiterjeszti a távközlés formális nyelveinek alkalmazását az ipari automatizálás több területére [28].

A rendszerfeltárást URN (UCM és GRL) nyelven ábrázoljuk, az elemzést UML nyelven ábrázoljuk és/vagy elemezzük, fejlesztőeszköztől függően, a koncepciók tervben UCM és MSC diagramokkal ábrázoljuk a rendszer eseményeit, ez konvertálható UML-be és SDL-be is, ASN.1-ben adjuk meg az elemzés során feltárt adatokat, amelyek így beépíthetők az SDL és TTCN specifikációkba.

Az SDL specifikáció helyességét ellenőrizhetjük MSC idősor diagramokkal. Az SDL specifikációt felhasználjuk a TTCN tesztgenerálás specifikálásához és a forráskód készítéséhez C++ vagy JavaScript nyelven. A *Deployment and Configuration language (DCL)* telepítő és konfigurálás leíró nyelvet az ITU munkacsoportja kidolgozás alatt tartja. Case eszközök automatikusan generálják a kész terméket az SDL specifikációból, és elvégeztethetjük a konformancia tesztelést.

4. Összefoglalás

A távközlés eszközeinek széleskörű használatával az elmúlt években a fejlesztői szemlélet is bevonult a különféle heterogén és egymásraépülő rendszerek fejlesztésébe. A több nyelven alapuló technikák nagyon eredményesek lehetnek, mert az összetett rendszerek különféle aspektusainak fejlesztésekor a célnak megfelelőbb eszközt alkalmazhatjuk. Modellelemek újrafelhasználása válik célszerűvé, amely technikáit és munkafolyamatszervezését az automatikus programfejlesztő környezetek és az alkalmazásterület CASE eszközei támogatják.

A távközlési nyelvek néven ismert ITU-T formális módszerek csoportjának és az UML nyelv konvergenciája figyelhető meg az UML2.0 szabvány megjelenésével [30], amely magába integrálta az MSC teljes eszköztárát az interakciók ábrázolására, valamint az SDL nyelv dinamikus viselkedést leíró eszköztárát az implementációs folyamatok állapotgépeinek ábrázolására.

Az SDL mint a formális specifikáció implementációjára jól kidolgozott nyelv, összefogja az ITU-T System De-

sign Language gyűjtőnévű szabványosított formális módszereit. A módszerekben rejlő lehetőségek konkrét hasznosításához továbbra is feladatunk a modellfejlesztés folyamatainak kutatása az UML és a formális modellező nyelvek konvergens elemeire építve az egyes alkalmazás-területek mentén [33,34].

Köszönetnyilvánítás

Köszönettel tartozom Dr. Tarnay Katalinnak, aki a Veszprémi Egyetem Műszaki Informatikai Karán iskolát teremtett a távközlési szoftverek formális módszertanainak kutatására és áldozatos munkájával segített.

A cikkben megjelenített munkát az OTKA 29556 számú kutatási szerződés támogatta.

Irodalom

- [1] Tarnay K.: Kommunikációs protokollok modellezése és konformancia vizsgálata. Doktori értekezés, 1990.
- [2] Bögel Gy.: Az infokommunikációs hullám sajátosságai, Híradástechnika 2003/5.
- [3] Czinkóczy A.: A távközlési hálózatok fejlődési iránya a következő 5-10 évben, Híradástechnika 2003/10.
- [4] Inside the personal communication portal www.intel.com, Intel Glenayre 2002/06.
- [5] Z. Mammeri: SDL modelisation de protocoles et systemes réactifs Hermes 2000.
- [6] J. Ellsberger, D. Hogrefe, A. Sarma: SDL Formal Object-oriented Language for Communicating Systems, Prentice Hall, 1997.
- [7] G. Kovács, Z. Pap, D. Le Viet, A. Wu-Hen-Chang, Gy. Csopaki: Applying Mutation Analysis to SDL Specifications, SDL-Forum, Springer-Verlag. 2003.
- [8] Gy. Csopaki, K. J. Turner: Modelling Digital Logic in SDL. In Proc. Formal Description Techniques X/Protocol Specification, Testing and Verification XVII, pp.367–382., Chapman and Hall, London, Nov.1997.
- [9] Pataricza A.: Formális módszerek az informatikában, Typotex, Budapest, 2004.
- [10] ITU-T Recommendation Z.100 (1995) "Specification and Description Language (SDL)"
- [11] ITU-T Recommendation Z.100 (1999) "Specification and Description Language (SDL)"
- [12] ITU-T Recommendation Z.105 (1995) "SDL Combined with ASN.1 (SDL/ASN.1)"
- [13] ITU-T Recommendation Z.109 (1999) "SDL Combined with UML (SDL/UML)"
- [14] ITU-T Recommendation Z.120 (1999) "Message Sequence Chart (MSC)"
- [15] ITU-T Recommendation X.680 (1994) "Data Networks and open System communications"

- OSI networking and system aspect – Abstract syntax Notation One (ASN.1)”
- [16] ITU-T Recommendation X.292 (1998), Z.140-Z.149 (2003):
“OSI conformance testing methodology and framework for protocol Recommendation for ITU-T applications – The Tree and Tabular Combined Notation (TTCN)”
- [17] A. Medve:
MSC and the Aspect-Oriented paradigm in protocol engineering, Int. Conf. MicroCAD’2003 Miskolc, pp.71–79.
- [18] T. Dulai, A. Medve:
The common use of SDL, CSP and MSC in protocol design, Int. Conf. MicroCAD’2003 Miskolc, pp.29–34.
- [19] www.itu.int.org/recommendation/zseries/languages
- [20] Medve A.:
A formális módszerek szerepe a távközlési szoftverek fejlesztésében.
Networkshop 2001, www.niif.hu/networkshop
- [21] Medve A.:
Az SDL nyelv jellemzése. Elektronikus segédlet, VE Információs Rendszerek Tanszék, 2003.
- [22] A. Medve:
Relations of testability and quality parameters of SDL implementation at the early stage of protocol development life cycle, CSCS’2002, Szeged.
- [23] A. Medve:
Standardized formal languages for reliable model engineering, Int. Conf. MicroCAD’2005 Miskolc
- [24] www.telelogic.com
- [25] www.telelogic.com/SDL/default.asp
- [26] www.sdl-forum.org/tools
- [27] ITU-T Recommendation Z.150-152 (2003)
User Requirements Notation – GoalRequirements Language Use Case Map Notation.
- [28] Medve A.:
A formális nyelvek egymásra épülése a fejlesztés folyamatában,
XXIV Neumann Kollokvium, VEAB Veszprém, 2003.
- [29] H. Farman:
Tau UML Suite Telelogic Szakmai Napok, Inventix Kft.2000.
- [30] www.omg.org
- [31] V. Vittorini, M. Iacono, N. Mazzocca, G. Franceschinis:
TheOsMoSys approach to multi-formalism modeling of systems,
Software System Modeling 2004/ 3, pp.68–81.
- [32] Raffai M:
UML 2 Modellező nyelvi kézikönyv,
Objektumtechnológia sorozat 4. kötet,
Palatia Nyomda és Kiadó, 2005.
- [33] A. Medve:
Enterprise Modeling with the joint use of User Requirements Notation and UML,
Idea Group Publishing, 2005. (fejezet átdolg. alatt).
- [34] A. Medve:
Process Driven Combination of Scenario-based and State-based Modeling Languages, TSE Special Issue on Interaction and State-Based Modelling, 2005. (beküldve)

Hírek

Az Oracle bejelentette, hogy az **Oracle Database 10g2** új rekordot állított fel a TPC-H két 300 GB-os adattár-házas sebességpróbáján is, ami újra igazolja a rendszer kimagasló teljesítményét és a vállalati szintű adatbáziskezelők között kivívott vezető helyét. A négy 3,33 GHz-es Intel Xeon processzorral és 8 MB harmadik szintű gyorsítótárral (EM64T) ellátott két csomópontos Dell PowerEdge 6800 szerverfürtön az Oracle Database 10g2 – Real Application fürtözéssel és Red Hat Enterprise Linux Advanced Server3 operációs rendszeren – világrekordnak számító 22 USD/QphH@300 GB ár-teljesítmény arányt ért el 11.742,8/QphH@300 GB teljesítmény mellett. A második sebességpróbán az Oracle Database 10g2 és az RAC együttesét négy 3,0 GHz-es Intel Xeon processzorral felszerelt, két csomópontos Dell PowerEdge 6600 szerverfürtön, Red Hat Enterprise Linux Advanced Server 3 operációs rendszeren futtatva a 10g első változatához képest 27%-os teljesítménynövekedést értek el. Ennél a benchmarknál 30 USD/QphH@300 GB ár-teljesítmény arány mellett 8604 QphH@300 GB-os teljesítményértéket mértek a 10g előző változatának 6795 QphH@300 GB-os eredményéhez képest.

A Java-fejlesztők már ingyenesen használhatják az **Oracle JDeveloper 10g** fejlesztőrendszert. Az Oracle ezzel a gesztussal a Java-fejlesztők iránti töretlen elkötelezettségét kívánja hangsúlyozni. A cég emellett felajánlotta, hogy az Eclipse Foundation nyílt forráskódú fejlesztői közösségben elvállalja az egységes fejlesztőeszközök kialakítását célzó JavaServer Faces (JSF) projekt vezetését, továbbá bejelentette, hogy fő támogatóként csatlakozik az Apache MyFaces kezdeményezéshez. A J2EE platform népszerűvé válásának egyik feltétele, hogy rendelkezzen olyan eszközökkel, amelyekkel a nagyvállalati Java-alkalmazások egyszerűen fejleszthetők. A Java-alkalmazások felhasználói felületének gyors kialakításához használható JavaServer Faces keretrendszer, valamint az ingyenes Oracle JDeveloper szoftver megkönnyíti az összetett alkalmazások fejlesztését, és így végső soron elősegíti a J2EE platform elterjedését. A webes alkalmazásokat kiszolgáló JSF keretrendszer és az Oracle JDeveloper 10g integrált fejlesztői környezet együttműködésével az összetett alkalmazások gyorsan létrehozhatók és egyszerűen rendszerbe állíthatók.

DVB-T mérési tapasztalatok Győrben és Budapesten

HONFY JÓZSEF

Széchenyi István Egyetem, Győr
honfy@sze.hu

Kulcsszavak: digitális televízió, DVB-T, OFDM, mérési eredmények

A televízióműsorok földfelszíni digitális sugárzása Magyarországon is elkezdődött, egyelőre a budapesti és a kabhegyi televízió állomásokról. Az alábbi cikk bemutatja a győri Széchenyi István Egyetemen és a budapesti Puskás Tivadar Távközlési Technikumban folytatott DVB-T vételi méréseket, illetve a vételi tapasztalatokat.

1. Bevezetés

Az Antenna Hungária Rt. a magyarországi digitális televízió műsorszórás adáskísérleteit 1999 júliusában kezdte Budapesten a 43. tévécsatornán egy kölcsön berendezéssel, amit 2001 októberében kiváltott egy saját berendezéssel. A következő lépés 2002 májusában a Kabhegyen üzembe helyezett adó volt, amelyet a meglévő UHF sávú antennarendszerre kapcsoltak. Bár a publikált adatok szerint a kisugárzott teljesítmény Kabhegyen 10 kW ERP volt, Győrben a Széchenyi István Egyetemen – ahol a vidéki kísérleti sugárzás beindítását óriási várakozás és alapos előkészületek előzték meg – nem sikerül feldolgozható jelet kapni.

Tovább lépésre csak 2004 őszén volt lehetőségünk, amikor Budapesten az 51. csatornán belépett a kísérletbe egy további adóberendezés. Ugyanekkor Kabhegyen is figyelemre méltó változtatásokat hajtott végre az AH Rt. Külön antennát szereltek fel a 64. csatornán sugárzó DVB-T adó számára, ezzel pedig Győrben is megnyílt a lehetőség a vételre és a vizsgálatokra.

A győri vételi próbálkozásokkal párhuzamosan egyetemünk kutatói megteremtették a bekapcsolódási lehetőséget a DVB-T adáskísérletekbe egy budapesti megfigyelő állomás létrehozásával. A témára nyerhető pályázati támogatásokkal és a Puskás Tivadar Távközlési Technikum hathatós támogatásával a budapesti megfigyelő állomás a Gyáli úton kezdte meg működését még 2003 elején.

Ezzel a lépéssel abba a szerencsés helyzetbe kerültünk, hogy a magyarországi DVB-T adáskísérleteket a 43., az 51., és a 64. tévécsatornákon is teljes egészében figyelemmel kísérhetjük és mindhárom adó jelét mérhetjük. Szerencsésnek tartom, hogy mindkét laboratóriumunk egy-egy oktatási intézményben kapott helyet, mégpedig úgy hogy a laboratóriumi kutatómunkát nem választottuk el az oktatástól, így a fiatalságot aktívan be tudjuk vonni bizonyos részfeladatokba, mérési gyakorlatokat vezettünk be a DVB-T adások elemzésére, diplomatervezési feladatokat adunk ki az adáskísérletek elemzésére, tudományos diákköri dolgozatok készülnek, és így tovább.

2. A kísérleti adók főbb műszaki adatai

Az Antenna Hungária Rt. jelenleg Budapesten és Kabhegyen folytat DVB-T sugárzást, tájékoztatásuk szerint ezek az adók már a tervezett hazai digitális televízió műsorszóró adóhálózat végleges elemei. Az egyes telephelyeken működő adók műszaki adatait a következőkben foglalom össze:

Budapest, Országos Mikrohullámú Központ:

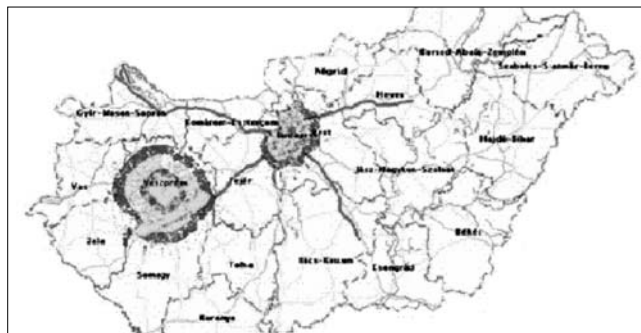
Frekvencia: UHF 43. (650 MHz) és 51. (714 MHz),
Kisugárzott teljesítmény: 1 kW ERP,
Polarizáció: horizontális,
Moduláció: OFDM/64QAM,
Mód: FFT/8k (6817 alvivő),
Védőintervallum: 1/32,
Hibajavítás: alpha 1, (NH), CR: 2/3

Kabhegy, televízió állomás:

Frekvencia: UHF 64. (814 MHz),
Kisugárzott teljesítmény: 2,5 kW ERP,
Polarizáció: horizontális,
Moduláció: OFDM/64QAM,
Mód: FFT/8k (6817 alvivő),
Védőintervallum: 1/32,
Hibajavítás: alpha 1, (NH), CR: 2/3

Az adóberendezések által besugárzott számítással meghatározott területet az AH Rt. honlapján találhatjuk.

1. ábra A DVB-T adók számított vételi területe

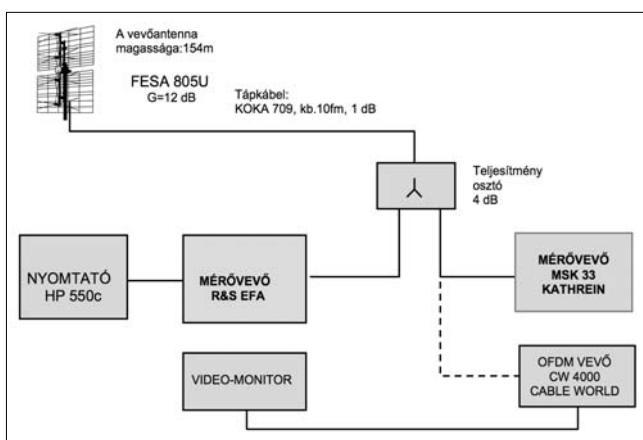


3. A DVB-T mérések

A bevezetőben elmondottak szerint a DVB-T sugárzás vizsgálatára Budapesten és Győrben hoztunk létre megfigyelő állomásokat. Ezeket az alábbiakban részletebben is bemutatjuk, továbbá néhány gondolatban elemezzük, hogy mit és miért célszerű mérni a vételmegfigyelés során, majd az igényeket összevetjük a rendelkezésünkre álló mérőeszközök adta lehetőségekkel.

3.1. Mérő- és megfigyelő-állomások

Budapesten a Puskás Tivadar Távközlési Technikum műholdas oktató laboratóriumában rendeztük be a digitális műsorsugárzások megfigyelő és mérőállomását. A DVB-T mérő munkahely rendszertechnikai kialakítása a 2. ábrán látható.



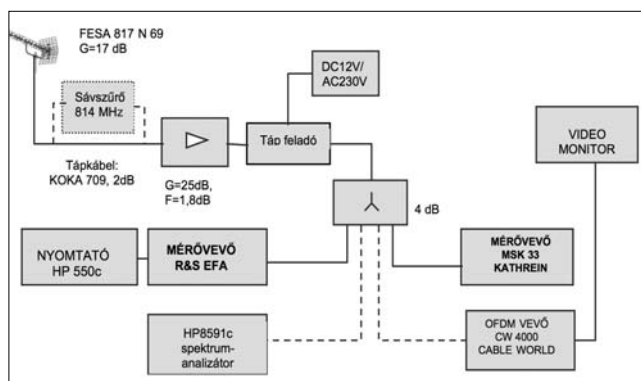
2. ábra A budapesti mérőállomás blokkvázlata

A felszereltségét tekintve a legfontosabb műszere a Rohde&Schwarz DVB-T mérővevő (EFA 40/43). Kiszolgálására építettük köré a többi egységet (nyomtató, videó monitor, OFDM vevő és MPEG dekóder stb.), amelyek az ábrán láthatók. Kontroll műszerként a Kathrein gyártmányú digitális mérővevőt (MSK 33) használtuk. Az állomást egy Hirschmann gyártmányú, szélessávú UHF panelantenna szolgálja ki jellel. A rendszer kialakításánál kényelmes helyzetben voltunk, mert az OMK és a laboratórium távolsága mintegy 8 km, a közvetlen átlátást semmilyen akadály sem zavarja.

Győrött a mérőállomás a Széchenyi István Egyetemen a Távközlési Tanszék Műholdas laboratóriumában kapott helyet, de végeztünk méréseket a győri mikrohullámú állomáson is. A vevőrendszer kialakítása Győrben több problémát is felvetett, mivel a kabhegyi adóállomás mintegy 50 km távolságban van, továbbá a domborzati viszonyok sem kedvezőek a vétel szempontjából. Az 1. ábra szerint az ellátottsági számítások nem is adnak vételi lehetőséget.

Győrben a mérések tervezésekor éppen ezért körültekintő figyelemmel kellett eljárni. A vételre – miután Kabhegyről a 64. tévécsatornát kell vennünk – a Hirschmann cég Fesa 817 N 69 típusú antennáját használjuk, amely a 63-69 tévécsatornákon 17dB nyereségű. A várható kis jelszintre való tekintettel az antenná-

ra még egy 25dB erősítésű, 1,8dB zajtényezőjű fejelesztőt is felszereltünk, továbbá az analóg TV-adók jeleinek leválasztásához a 60-69-es csatornákra terveztünk egy sávszűrőt is. A mérőállomás rendszertechnikájának kialakítása a továbbiakban hasonló a budapestihez (R&S EFA 40/43, nyomtató, videó monitorok, OFDM vevő és MPEG dekóder, digitális mérővevő (MSK 33), stb.) de a mérőállomás rendelkezik még egy HP 8591 C típusú kábeltelevíziós spektrum analízátorral is. Az összeállítás rendszertechnikáját a 3. ábra mutatja.



3. ábra A győri mérőállomás blokkvázlata

3.2. Mit és miért mérünk?

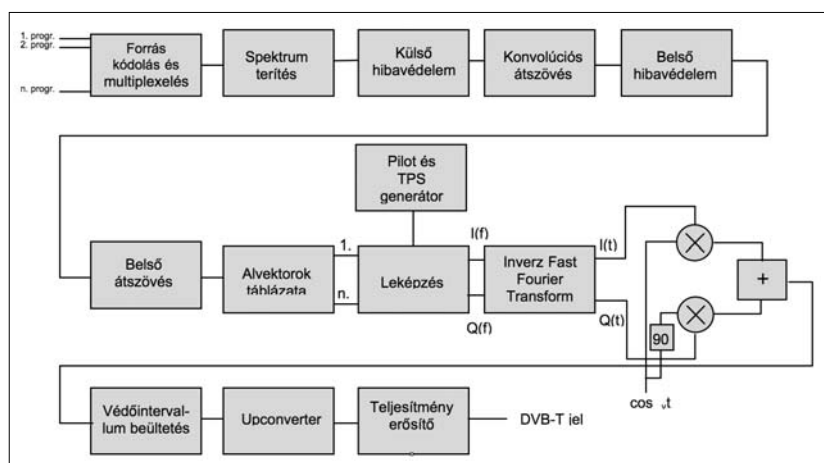
Ahhoz hogy a kérdésre választ adjunk, és a mérési eredményeinket értékelhessük, tekintsük át röviden az OFDM jel leggyakrabban alkalmazott előállítás és vételi folyamatát.

A modulációs eljárás a 4. ábra nyomán röviden a következő:

- A videó, hang és egyéb adatfolyamokat a megfelelő forráskódolás után egyetlen bitfolyamba nyálabolják össze (multiplexelés). A kimeneti spektrum szétterítése érdekében ehhez a bitfolyamhoz egy álvéletlen generátor jelét keverik hozzá. Ezt követően hibavédő kóddal látják el a jelfolyamot, ezek a lépések sorrendben a Reed-Solomon kódolás, a külső átszövés, majd a konvolúciós kódolás (belső hibavédelem) és a belső átszövés.

- Az így kapott jelfolyamból a következő lépésben egy sokvivős modulált jelet állítanak elő. Ehhez a jelfolyam először áthalad egy leképezőn, amely a folytonos jelfolyam egymást követő elemeit egy-egy komplex számnak megfelelő alvektor valós és képzetes koordinátájává teszi. Minden egyes alvektor egy modulált állapotú vivőnek felel meg, ahol az alvektor valós és képzetes része a modulált vivő koszinuszos és szinuszos összetevőnek felel meg. Mindez a frekvenciatartomány egy adott időintervallumbeli leírását adja eredményül.

- Az eljárás következő lépése az IFFT (gyors inverz Fourier-transzformáció) lesz, amely a jelfolyamot átranzformálja a frekvenciatartományból az időtartományba. A transzformáció bemenő adata a kettő hatványának megfelelő számú jelminta. 8k üzemmódban a DVB-T jel 6817 vivőt tartalmaz, amely 8192 (2^{13}) frekvenciaminta egy részhalmaza (8k mód). Kidolgozták még a 2k módot is, ez 1705 vivőt használ, amely pedig 2048 (2^{11})



4. ábra A DVB-T jel előállításának sémája

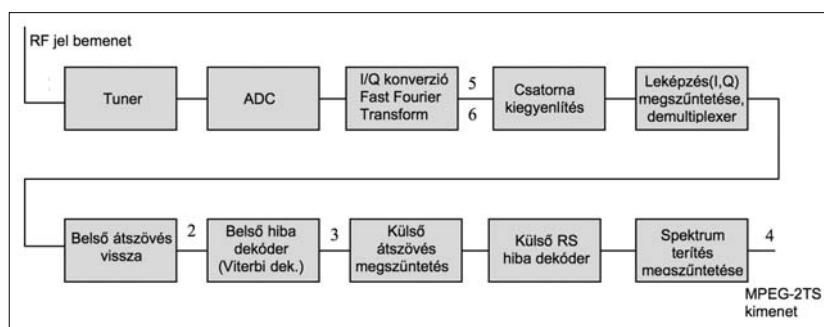
frekvenciaminta egy adott részhalmaza. A DVB-T jelet specifikáló EN 300 744 szabvány minden egyes OFDM szimbólumot az adott módnak megfelelő vivőkészlet összegeként definiál. 8k módban a 6817 tényleges vivőből csak 6048 darab vivő hordoz adatot, a pilotvivők száma 769, 2k módban pedig 1512 hasznos adathordozó vivő és 193 pilotvivő van. A pilotvivők a frekvenciatengelyen lehetnek fix vagy változó pozícióban, a változó pilotok helyzetét egy álvéletlen függvény határozza meg. A pilotjelek feladata a frekvencia-, idő- és keretszinkronizálás.

- Az IFFT blokkot követi a komplex szorzó (I/Q modulátor), melynek kimenetén az időtartományban megjelenik az OFDM jel. A komplex szorzó utáni jel a fix pilotvivőket kivéve véletlenszerűnek tekinthető annak köszönhetően, hogy több ezer, egymástól függetlenül modulált alvivőből áll.

- Ezután a védőintervallum beültetés következik, ennek eredményeképpen a földfelszíni adás kimenő jelében modulált vivős hasznos adatátviteli időszakok és kikapcsolt vivős védőintervallumok váltogatják egymást. A hasznos adatátviteli idő mindig a vivőtávolság reciproka, a védelmi idő pedig ennek 1/4, 1/8, 1/16 vagy 1/32 része. A szimbólumidő a hasznos adatátviteli idő és a védelmi idő összege.

Az 5. ábrán felrajzoltuk a DVB-T vevő funkcionális blokkvázlatát. A működés magyarázatát mellőzzük, hiszen gyakorlatilag az adó oldali kódolási- és modulációs eljárások inverzét kell elvégezni a jel dekódolásához.

5. ábra A DVB-T jel vétele és dekódolása



A rendelkezésünkre álló lehetőségek és eszközök a vételi oldal vizsgálatára adnak lehetőséget. Törekedtünk arra, hogy a vizsgálatunk a DVB-T ellátottság ellenőrzésére nemzetközileg ajánlott méréseket is tartalmazzák, illetve annak megfelelőek legyenek. Az általunk vizsgált jellemzők az alábbiak voltak:

- a vett jel szintje,
- a hibaarány, BER a Viterbi dekódolás (konvolúciós kód dekódolása) előtt,
- a hibaarány az RS (Reed-Solomon) dekódoló előtt,
- a hibaarány a RS dekódoló után,

- MER (modulációs hibaarány),
- konstellációs ábra,
- impulzus válasz (a reflektált jel vizsgálata),
- spektrumanalízis, például a szomszédos csatornák vizsgálatára,
- a vett jel képi megjelenítése.

Az 5. ábrán bejelöltük az egyes paraméterek vizsgálati pontjait. A mért jellemzők sajátosságait, megjelenítési formáját, illetve értelmezését célszerű a mérési adatok tárgyalása előtt röviden áttekinteni.

A digitális átvitel minőségének legfontosabb paramétere a BER (Bit Error Ratio), a hibás bitek számának és valamennyi bit számának hányadosa. Az átvitt adatok visszaállításának folyamatában több helyen is mérhetünk, az egyes mérési pontokban a kapott értékek különböző átviteli jellemzőkre adnak minősítést.

A Viterbi dekódolás előtti BER mérése arról ad információt, hogy az adóhálózatnak, a csatornának és a vevőnek milyen a hibavédelem nélküli minősége. A mérés az 5. ábra 2-es mérőpontján történik. A vevő Viterbi dekódolás utáni adatfolyamát ugyanannak a konvolúciós kódolásnak vetik alá, mint amivel azt az adóoldalon kódolták. Ez egy olyan adatfolyamot eredményez, amely az adó kimenetén lévő adatfolyam becslése lesz. Ezt az adatfolyamot bitszinten hasonlítják össze azzal az adatfolyammal, amely a Viterbi dekódolás előtt rendelkezésre áll.

A BER mérésére a vételi oldalon az úgynevezett üzem közbeni mérési eljárást használják. Ez azon az alapfeltételezésen alapul, hogy az RS (Reed-Solomon) bájtokat az átviteli úton, annak minden összeköttetési egységében korrigálják. Normális esetben az RS dekódoló minden hibás bájtot képes kijavítani és hibamentes TS packetet biztosít. Nagy hibacso-

magok esetében azonban az RS dekódor már nem képes a hibák kijavítására. Ebben az esetben a TS hibajelző bitet be kell állítani és egyetlen további adatbit értékét sem szabad megváltoztatni, miközben a következő átviteli szakaszra való küldés előtt az RS redundancia bájtokat újra meg kell határozni. A mérés az RS dekódor által szolgáltatott bitfolyam és az RS dekódolás előtti bitfolyam ösz-

szezhasonlításán alapul. Ennek megfelelően az átviteli úton bármely RS BER mérés csak az adott szakaszon keletkezett hibákat tartalmazza. A hibák száma megkapható úgy, hogy az RS dekódolás utáni biteket összezhasonlítják az RS dekódolás előttiékel.

A BER értékek küszöbszintje az $1,0E-4$ értéknél van. Ha ennél rosszabb, akkor már nem beszélhetünk üzemképes átvitelről.

A RS utáni BER méréseknél kijelzett $0,0E-9$ -es érték azt jelzi, hogy tulajdonképpen nem volt hiba, de nem történt olyan időtartamú vizsgálat, hogy 10^{-9} -nél jobb értéket lehessen kiírni. Mivel a BER mérése meglehetősen hosszadalmas, azt is célszerű kiírni, hogy a kijelzett érték hány mérés átlagát mutatja.

Az IQ analízis az OFDM jel egyetlen vivőjére is alkalmazható, de elvégezhető a vivők tetszőleges csoportjára is. Ha a vivőcsoportok konstellációs diagram analízisét végezzük, akkor a kijelölt vivőhöz tartozó valamennyi szimbólumot egyetlen IQ diagramban összegezten kell felrajzolni. A folytonos és a szórt pilotokon, illetve a TPS vivőkön eltérő a modulációs leképzés, ezért azok az IQ síkú ábrázolásban nem szerepelnek. A TPS vivő (Transmission Parameter Signaling) az OFDM jelre vonatkozó legfontosabb adatokat tartalmazza; ezek az FFT mód, a védőintervallum, az alvivők modulációs módja, a hierarchiatényező, a kódarány. A TPS adatok – amelyek a hasznos adatok vételéhez és dekódoláshoz szükségesek – nagyon magas hibavédelemmel rendelkeznek, a vivők modulációja DBPSK.

Az átvitel minősítésére használt legfontosabb mérési paraméter a modulációs hibaarány. A MER alkalmas arra, hogy a vett jel megbízhatóságát jellemezze. Tartalmaz minden olyan hibát, amely a vevő döntési áramkörének bemenetén megjelenik. A mérés kiindulási feltétele, hogy a szimbólumidőzítés és a vivőfrekvencia hibátlanul detektált legyen. A MER meghatározásának menete:

- N darab vett szimbólum (I_j , Q_j) koordináta értékpárjait tárolják;
- Minden egyes értékpár esetében dönteni kell arról, hogy vajon milyen szimbólumot küldtek;
- A döntés után meg kell határozni azt a vektort (I_j , Q_j), amely a feltételezett adás oldali szimbólumhoz tartozik;
- A hibavektor (δI_j , δQ_j) az ideális pont (I_j , Q_j) és a ténylegesen vett pont (I_j , Q_j) távolságvektora;
- Az ideális szimbólumvektorok koordinátáinak négyzetösszegét valamennyi vett szimbólumra (N) összegzik, majd elosztják a szimbólum hibavektorok koordinátáinak négyzetösszegeiből képzett összeggel.

Az eredményt, mint teljesítmény-hányadost például logaritmikus egységekben kifejezve kapjuk a MER-t:

$$MER_{[dB]} = 10 \log \left[\frac{\sum_{j=1}^N (I_j^2 + Q_j^2)}{\sum_{j=1}^N (\delta I_j^2 + \delta Q_j^2)} \right]$$

A MER úgy tekinthető, mint egyfajta jel/zaj viszony, amely pontosan megmutatja a vevő minőségét. Nem csak a zajt, de minden egyéb, a minőséget rontó paraméter hatását is tartalmazza. Gyakran emlegetik, mint a digitális moduláció jel/zaj viszony ekvivalensét.

A konstellációs ábrából további, a vett jel különböző torzításaira jellemző paraméterek határozhatók meg:

- amplitúdó kiegyenlítetlenség (Amplitude Imbalance),
- kvadratúra hiba (Quadrature Error),
- a vivő elnyomás (Carrier Suppression),
- a fázis dzsitter (Phase Jitter).

Természetesen további paraméterek is generálhatók a konstellációs ábrából az átvitel minősítésére az itt felsoroltakon kívül, ez a rövid áttekintés azonban nem ad lehetőséget valamennyi jellemző bemutatására.

3.3. A mérőműszerek

A méréseink megtervezésekor az alapvető kérdés az volt, hogy mit mérünk. A korrekt válaszhoz azonban mérlegelnünk kell azt is, hogy mit tudunk mérni, milyen műszereink vannak. Rendszerünk alapvetően a Rohde & Schwarz EFA mérővevőre épül. A 6. ábrán áttekinthetők a műszerrel mérhető jellemzők a különböző felhasználási területeknek megfelelően csoportosítva. Segítségül bejelöltük a vételi megfigyeléseinkhez használható választéket.

DVB-T OFDM application	Level	BER	MER	SNR	Carrier suppression	Quadrature error	Amplitude imbalance	Phase jitter	Constellation diagram	MER(1)	VOH	Spectrum-shoulder measurement	Amplitude imbalance (1)	Amplitude imbalance (2)	Amplitude imbalance (3)	Impulse response	History
Production of modulators and transmitters	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Transmitter installation and SFN adjustments	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Coverage measurement of terrestrial signals	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Monitoring of TV transmitters and transposers	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Research & Development	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Service	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

6. ábra Az OFDM jel Rohde&Schwarz cég EFA típusú mérővevőjével mérhető jellemzői

Ezek tulajdonképpen azok a vételi jellemzők, amelyeket a műszerrel a vizsgálataink során mértünk, illetve felhasználtunk. Bemutatásukhoz a Győrben, az egyetemen végzett méréseink adatait használjuk fel segítségül.

3.3.1. R&S EFA DVB-T mérővevő

A fő mérési menü

Minden képernyő jellemzője, hogy a legfontosabb adatokat a keret feletti fejlécben mindig megadja, ezek: a csatorna középponti frekvencia, az üzemmód, a védőintervallum, az OFDM jel szintje, az alvivők modulációja, a moduláció hierarchia tényezője, a kódarány, és a BER értékek a Viterbi dekóder előtt, a Reed-Solomon dekóder előtt és után.

A bekeretezett fő mérési adatok a demodulált DVBT csatorna fő paramétereit adják meg, könnyen, gyorsan áttekinthetően.

DATE 22.03.05 TIME 10:32:42
 RF 818.000000 MHz LEVEL 54.0 dBuV
 FFT 8K GI 1/32 64 QAM ALPHA 1 NH CR 2/3
 BER BEFORE VIT 2.1E-3 BER BEFORE RS 9.4E-8
 BER AFTER RS 0.0E-9

DVB-T MEASURE			
SET RF (8MHz)	ATTEN : 0 dB		
818.00 MHz	54.0 dBuV		
FREQUENCY/MER/BER:		CONSTELL. DIAGRAM...	
FREQUENCY OFFSET	-0.052 kHz	FREQUENCY DOMAIN...	
BITRATE OFFSET	7.0 ppm		
MER (RMS)	24.7 dB	SPECTRUM/TIME DOMAIN.	
BER BEFORE VIT	2.1E-3 (10/10)		
BER BEFORE RS	9.4E-8 (100/100)	OFDM PARAMETERS...	
BER AFTER RS	0.0E-9 (248/10K0)		
OFDM/CODE RATE:		RESET BER	
FFT MODE	8K (TPS: 8K)		
GUARD INTERVAL	1/32 (TPS: 1/32)	ADD. NOISE OFF	
ORDER OF QAM	64 (TPS: 64)		
ALPHA	1 NH (TPS: 1 NH)		
CODE RATE	2/3 (TPS: 2/3)		
CELL ID	0000		
TPS RES (F1-F4)	00,00,00,00		
TS BIT RATE 24.12834 MBit/s			

7. ábra

A R&S EFA típusú mérővevőjének fő mérési jellemzőket összegző képernyője a kabhegyi adó mérési adataival

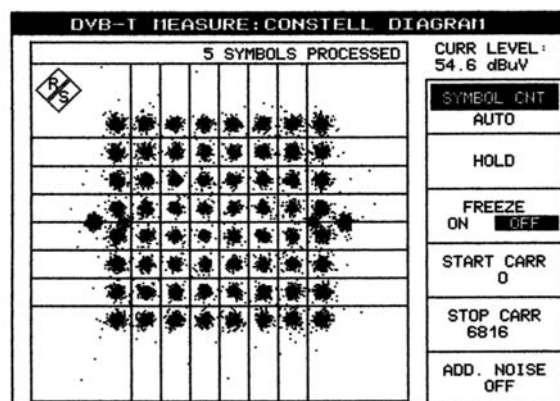
Itt olvashatók ki (7. ábra) a vételi frekvenciára vonatkozó adatok, a vett OFDM jel szintje, a BER értékek, az átlagolt MER és, az OFDM adási paraméterek, az átvitt TPS pilotok (a belső demodulációs beállításokkal összehasonlítva). A jobb oldalon láthatók a lehetőségek a belépésre a további mérési almenükbe. Ugyanitt aktiválható a műszer a belső zajgenerátora is.

Konstellációs ábra

A konstellációs ábra a legjobb módszer a digitális moduláció megjelenítésére, a mérési eredmények értelmezésére, mint például a vivőelnyomás vagy az I/Q amplitúdó-kiegyensúlyozatlanság mérésére. Mélyanalízis elvégzése céljából lehetőség van a képernyőn megjelenített szimbólumok számának beállítására (a példa öt szimbólumot jelenít meg). Ha szükséges, az EFA automatikusan is beállíthatja a szimbólum számot úgy, hogy optimális legyen a képrfrissítési arány (8. ábra).

8. ábra A kabhegyi adó jelének konstellációs ábrája

DATE 22.03.05 TIME 10:25:31
 RF 818.000000 MHz LEVEL 54.6 dBuV
 FFT 8K GI 1/32 64 QAM ALPHA 1 NH CR 2/3
 BER BEFORE VIT 1.9E-3 BER BEFORE RS 9.0E-8
 BER AFTER RS 4.0E-8



DATE 22.03.05 TIME 10:44:43
 RF 818.000000 MHz LEVEL 54.5 dBuV
 FFT 8K GI 1/32 64 QAM ALPHA 1 NH CR 2/3
 BER BEFORE VIT 1.7E-3 BER BEFORE RS 6.1E-8
 BER AFTER RS 0.0E-9

DVB-T MEASURE: OFDM PARAMETERS			
SET RF (8MHz) 818.00 MHz		ATTEN : 0 dB 54.5 dBuV	
PARAMETERS: CENTR CARR EXCLUD		CONSTELL DIAGRAM...	
CODER:		FREQUENCY DOMAIN...	
I/Q AMPL IMBALANCE		-0.01 %	
I/Q QUADRATURE ERROR		+0.01 °	
CARRIER SUPPRESSION		---- dB	
PHASE		---- °	
TRANSMISSION:		SPECTRUM/ TIME DOMAIN.	
PHASE JITTER (RMS)		0.00 °	
SIGNAL/NOISE RATIO		27.9 dB	
SUMMARY:		START CARR 0	
MOD ERR RATIO (RMS)		25.5 dB	
MOD ERR RATIO (MIN)		-0.7 dB	
MOD ERR RATIO (RMS)		5.3 %	
MOD ERR RATIO (MAX)		108.3 %	
AVERAGE: 100 %		STOP CARR 6816	
		ADD. NOISE OFF	

9. ábra

A kabhegyi adó jelének átlagolt OFDM paramétereit mutató munkaablak

Hasonlóképpen beállítható, hogy hány alvivő modulációját jelenítsük meg az ábrán, lehetőség van például akár csak egy, tetszőlegesen kiválasztott vivő beállítására, vagy akár az összes vivő modulációjának egyidejű megjelenítésére is. Vizsgálhatók a szinkronjelek, vagy a TPS vivők, illetve modulációjuk is.

OFDM paraméterek

Minden OFDM paramétert a kiválasztott vivők konstellációs diagrammából számít ki a rendszer (9. ábra). A vizsgálni kívánt vivők, vagy vivőcsoportok tetszőlegesen beállíthatók.

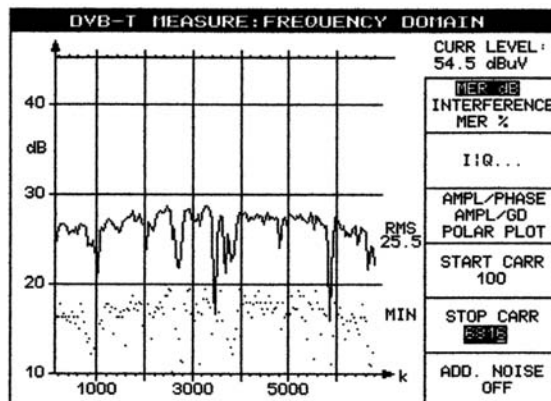
MER a frekvencia függvényében

A MER (Modulációs hibaarány) mérése és ábrázolása az alvivők függvényében az egyik leghatásosabb műszeres vizsgálati mód, mivel lehetővé teszi a rádiócsatorna általános állapotának megfigyelését (10. ábra).

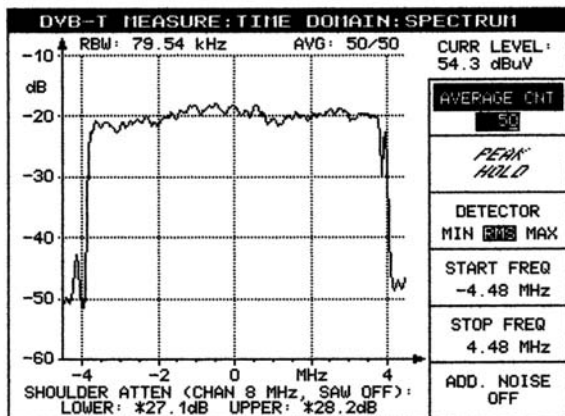
10. ábra

A MER vizsgálata és ábrázolása az alvivők függvényében

DATE 22.03.05 TIME 10:49:58
 RF 818.000000 MHz LEVEL 54.6 dBuV
 FFT 8K GI 1/32 64 QAM ALPHA 1 NH CR 2/3
 BER BEFORE VIT 1.6E-3 BER BEFORE RS 8.4E-8
 BER AFTER RS 0.0E-9



DATE 22.03.05 TIME 10:52:18
 RF 818.000000 MHz LEVEL 54.6 dBuV
 FFT 8K GI 1/32 64 QAM ALPHA 1 NH CR 2/3
 BER BEFORE VIT 1.5E-3 BER BEFORE RS 7.5E-8
 BER AFTER RS 0.0E-9



11. ábra A kabhegyi DVB-T adó vett jelének spektruma

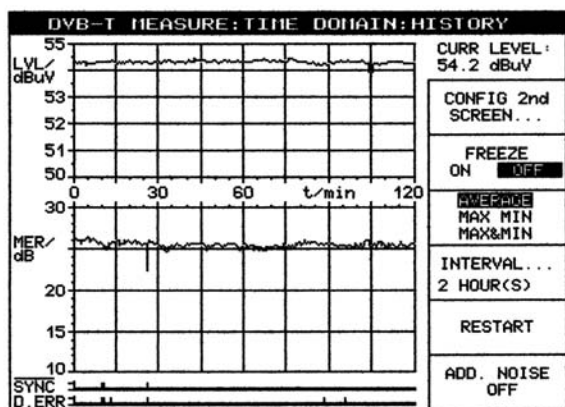
Természetesen lehetőség van az OFDM jel bármely károsodott alvivőjének gyors megkeresésére és elemzésére is. A szomszéd csatornás interferencia is mérhető és a képernyőre vihető. Nagyon fontos azonban megjegyeznünk, hogy az egyvivős esetekre kidolgozott MER érték fogalma az OFDM modulációban, mint egyetlen fő paraméter, nem létezik. Bár a műszer elvégzi az egyes alvivők MER értékeinek átlagolását és a fő mérési menüben ki is jelzi, ezt az értéket csak egy olyan tájékoztatásként célszerű kezelni, amely a csatorna átlagos állapotáról ad információt.

Spektrum-analízis

A beépített funkciónak köszönhetően tulajdonképpen nincs szükség külön spektrum-analizátorra. A műszer rendelkezik minden alapszintű spektrum-analizátor funkcióval, például a start/stop frekvenciamód (vagy a szakasz középpontozás), és ugyanígy elvégezhető számos detektálási és átlagolási módszer is (11. ábra).

12. ábra A kabhegyi adó naplózott mérési adatai: a jelszint és az összes adatvivő átlagolt MER értékei 120 perc időtartományban felvéve

DATE 22.03.05 TIME 13:47:25
 RF 818.000000 MHz LEVEL 54.0 dBuV
 FFT 8K GI 1/32 64 QAM ALPHA 1 NH CR 2/3
 BER BEFORE VIT 1.4E-3 BER BEFORE RS 4.4E-8
 BER AFTER RS 0.0E-8



Naplózás funkció

A hosszú távú DVB-T átvitel megfigyelését segíti. A legtöbb lényeges paraméter (jelszint, MER/dB, MER/%, BER és szinkronizációs adatok stb.) naplózása kiválasztható és grafikus formában, beállítható időszakassal, időfüggvényként rögzíthető. Ez a mód képes minden értéket numerikusan is megjeleníteni (átlag, max., min., aktuális érték). A BER és a szintmérések naplózási funkciója folyamatosan fut, mialatt más mérések ettől függetlenül végezhetők (12. ábra). Az ábrán egy kétórás időtartamú felvétel látható Győrben az Egyetemen rögzítve.

Impulzus válasz

Az impulzus válasz mérése (a védő-intervallumon belül) nagyon hasznos információt adhat különösen az egyfrekvenciás hálózatok (SFN) beállításánál. Ez a mérési funkció képszerűen és numerikus értékekkel is ábrázolja a fő DVB-T jelet (0dB, referenciával), a reflexiókat és az elő-reflexiókat. A zoom funkció a gyors reflexiók megjelenítésére szolgál, melyek a beépített környezetben léphetnek fel (az épületekről való visszaverődés következményeképp).

Hogy igényeink szerint állíthassuk be az alkalmazást, az x tengely egységeit és léptékét is megváltoztathatjuk, például μ s-ról kilométerre vagy éppen mérföldre (13. ábra).

3.3.2. MSK 33 mérővevő

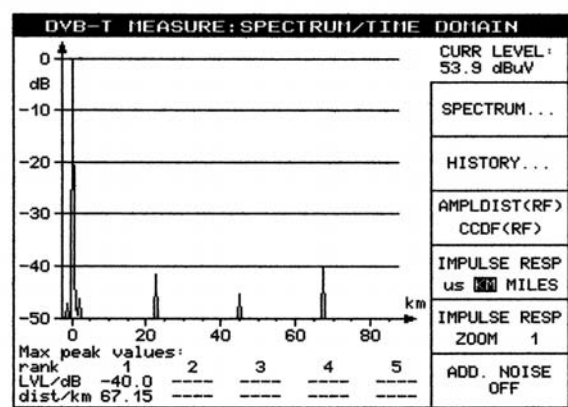
A műszer ellentétben az EFA mérővevővel szerviz műszer. Akkumulátoros táplálása folytán kitűnően alkalmazható terepi- és mobil mérésekre, kielégítő pontosságú és részletességű mérési adatokat szolgáltat. A műszerrel mérhető, illetve megjeleníthető jellemzők:

- konstellációs ábra megjelenítése, az összes vivőre, illetve a tetszőlegesen kiválasztott vivőre egyaránt,
- bit-hibaaarány mérése (BER),
- jel-zaj viszony mérése (SNR),
- carrier offset mérése,
- moduláció mérése,

13. ábra

A R&S mérővevő impulzus válasz elemző munkablaka

DATE 21.03.05 TIME 17:42:54
 RF 818.000000 MHz LEVEL 54.0 dBuV
 FFT 8K GI 1/32 64 QAM ALPHA 1 NH CR 2/3
 BER BEFORE VIT 1.7E-3 BER BEFORE RS 4.3E-8
 BER AFTER RS 0.0E-8





14. ábra Az MSK 33 mérővevő munkaablakai az adásmód és a konstellációs ábra megjelenítésére



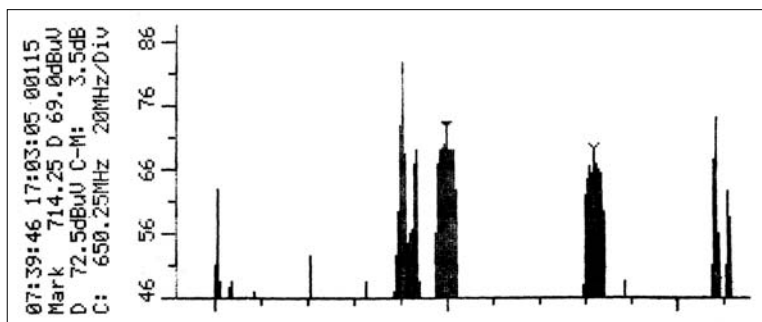
15. ábra Spektrumanalízis az MSK 33 mérővevővel.

A jobb oldalon a 41. és az 58. csatornás analóg és a 43., illetve az 51. csatornákon sugárzott digitális adások spektrumai.

- hierarchia moduláció mérése,
- kódarány mérése (C/R),
- védőintervallum kijelzése,
- inverz üzemmód,
- vételi csatorna adatainak megjelenítése,
- impulzus válasz megjelenítése (terjedési reflexiók),
- spektrum-analizátor mód,
- OFDM jelszint mérés.

16. ábra

Az MSK 33 mérővevő beépített nyomtatójával megjelenített spektrum. A két marker a 43. és 51. csatornák spektrumait mutatja.



A műszer jellemző szolgáltatásait a Budapesten végzett mérések eredményeinek megjelenítésével mutatjuk be. A 14. és a 15. ábrákon a képernyőn megjelenített adatok, a 16. ábrán a saját nyomtatójával készült spektrumfelvétel látható. Ez a spektrum azért figyelemre méltó, mert a két digitális adó spektruma (a 43. és 51. csatornákon) alatt, illetve felett látható a két analóg, őket közrefogó adó spektruma is (41. és 57. csatornák).

4. Mérési eredmények és értékelésük

4.1. Győrben a Széchenyi Egyetemen végzett mérések

A 7-13. ábrák mutatták be az egyetemen végzett mérések adatait. Elemzésükkel és figyelmes tanulmányozásukkal rendkívül sok információt kaphatunk a DVB-T vételről. A mért jellemzők mindegyikének a külön-külön bemutatására terjedelmi okokból nincs módunk, a mérések legfontosabb eredményeit az alábbiakban az 1. táblázatban foglaltuk össze. A mért jelszint felhasználásával (a nap-

lózott jelszint-mérés átlagával), ismerve a vevőrendszerünk elemeinek jellemzőit, visszaszámoltuk a vételi pontban a térerősség értékét.

4.2. Mérések a győri mikrohullámú állomáson

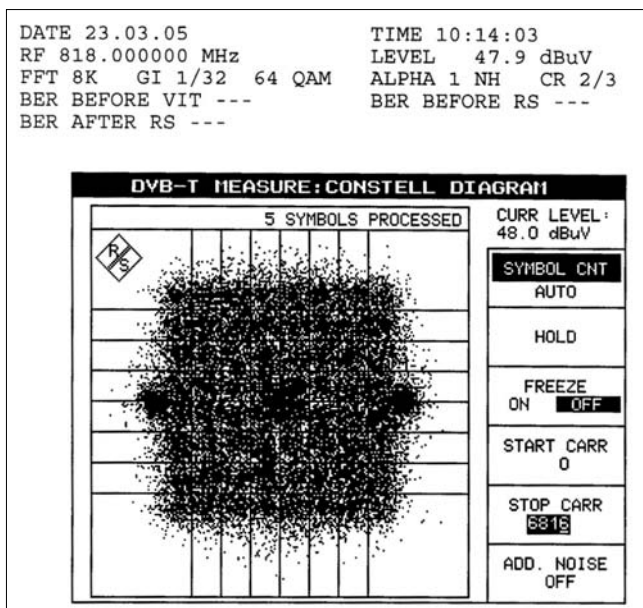
A kabhegyi adó vételének lehetőségét megvizsgálandó, méréseket végeztünk Győr Szabadhegyen, az Antenna Hungária Rt. mikrohullámú, és televízió állomásán is. Ez a helyszín – lévén az állomáson a mérések alatt is működő URH-FM rádió adóberendezések és analóg televízió adóberendezések is vannak – rendkívül zajos környezet a DVB-T vételére. A mérési eredmények, amelyeket a 2. táblázatban foglaltunk össze, ezt tükrözik is.

Érdekes és a DVB-T rendszerben alkalmazott hibajavító kódolási eljárás rendkívüli hatékonyságára utal, hogy a mikrohullámú állomáson felvett konstellációs ábráról (17. ábra) látható rendkívül rossz jelminőség ellenére is a megjelenített kép hibátlan, zavartalan.

A „maszatos” konstellációs ábra nyilvánvalóan annak következménye, hogy a vételi frekvenciasávban nagyon magas az állomás működéséből eredő intermodulációs produktumok szintje. A 18. ábrán az impulzusválasz felvételén jól látszik a „szokásos” reflektált szimbólumok helyetti folyamatos időtartamú zavaró interferencia.

17. ábra

A kabhegyi adó jelének konstellációs ábrája a győri mikrohullámú állomáson



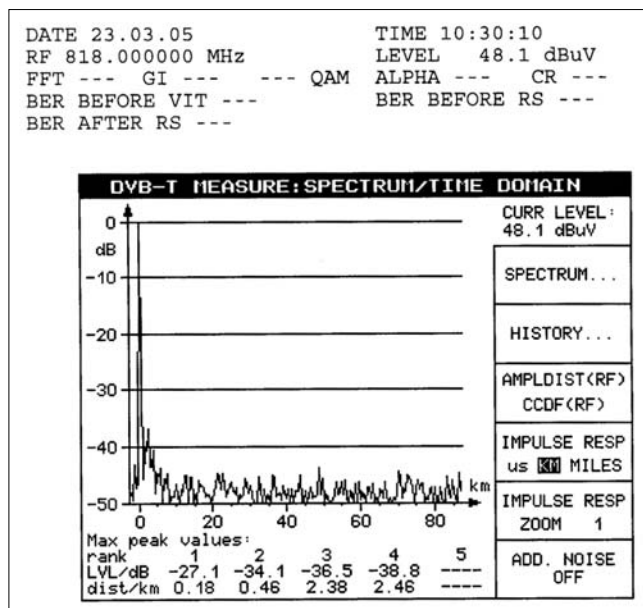
Mérések Győrben az egyetemen	Csatorna: 64 (818 MHz) Hierarchia nincs ($\alpha=1$) 64 QAM GI=1/32, C/R=2/3	
Jellemzők		
A mérés időpontja	03. 21	03. 22
Jelszint	53,7 dB μ V	54,5dB μ V
BER a Viterbi dekóder előtt	9,5E-4 (10/10)	2,1E-3 (10/10)
BER a Reed-Solomon dekóder előtt	1,1E-8 (42/100)	9,4E-8 (100/100)
BER a Reed-Solomon dekóder után	0,0E-7 (42/100)	0,0E-9(248/10K0)
MER átlag, k= 0..6816	22,5dB	24,7dB
A vételi pont térerőssége a vett jel szintjéből	45dB μ V/m	
Vett jelből kapott dekódolt kép minőség	Folytonosan hibátlan minőségű	

Mérések Győrben a mikrohullámú állomáson	Csatorna: 64 (818 MHz) Hierarchia nincs ($\alpha=1$) 64 QAM GI=1/32, C/R=2/3	
Jellemzők		
A mérés időpontja	03. 23	
Jelszint	48 dB μ V	
BER a Viterbi dekóder előtt	4,6E-2 (4/10)	
BER a Reed-Solomon dekóder előtt	6,8E-4 (3/10)	
BER a Reed-Solomon dekóder után	1,3E-5 (2/10)	
MER átlag, k= 0..6816	16,7dB	
A vételi pont térerőssége a vett jel szintjéből	40dB μ V/m	
Vett jelből kapott dekódolt kép minőség	Folytonosan hibátlan minőségű	

1. és 2. táblázat
 Mérések a győri Széchenyi Egyetemen (fent)
 és a győri mikrohullámú állomáson (lent)

A Győrben végzett mérések egyik legnagyobb tanulsága, hogy bár a város messze kívül esik a számított besugárzási határon, igényes, gondosan beállított eszközökkel, nagy nyereségű, az épületek fölé szerelt antennával, antennaerősítővel lehetséges olyan jel-szintű vétel, amely elegendő a DVB-T jel hibátlan feldolgozásához.

18. ábra
 Zavarjelek a győri mikrohullámú állomáson



Mérések Budapesten a Puskás Technikumban	Hierarchia nincs ($\alpha=1$) 64 QAM GI=1/32, C/R=2/3	
Jellemzők		
Csatorna	43. (650 MHz)	51. (714 MHz)
A mérés időpontja	03. 25	03. 25
Jelszint	68 dB μ V	62,0dB μ V
BER a Viterbi dekóder előtt	1,4E-2 (10/10)	1,2E-5 (10/10)
BER a Reed-Solomon dekóder előtt	2,0E-4 (10/100)	0,0E-9 (129/1000)
BER a Reed-Solomon dekóder után	1,1E-8 (28/1000)	0,0E-8(131/1000)
MER átlag, k= 0..6816	22,2dB	30,5dB
A vételi pont télerőssége a vett jel szintjéből	84dB μ V/m	80,5dB μ V/m
Vett jelből kapott dekódolt kép minőség	Folytonosan hibátlan minőségű	

3. táblázat

Mérések a budapesti Puskás Technikumban

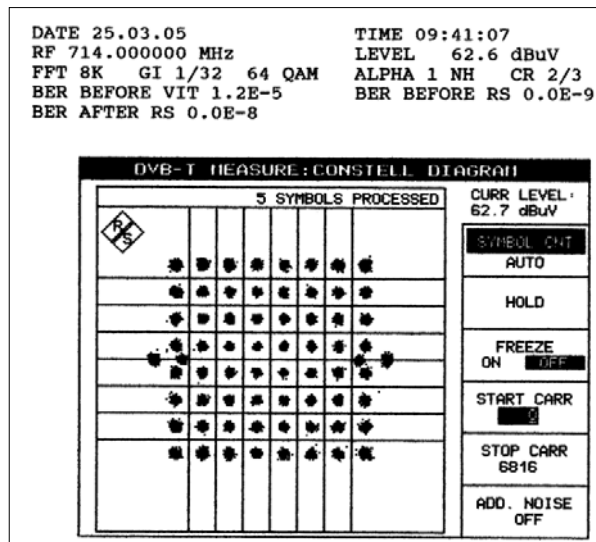
4.3. A budapesti mérések

A Puskás Tivadar Távközlési Technikum laboratóriumában végzett mérések az előzőekkel szemben lényegesen jobb eredményt adtak. Nem csoda, hiszen közvetlen átlátás van a labor és a mintegy 8 kilométerre lévő Széchenyi hegy (OMK) között. A vételi vizsgálatok itt nem az ellátottság ellenőrzése, a lehetséges vétel feltételeinek a meghatározása volt, hanem a két különböző csatornán folyó sugárzás összehasonlítása. A mérési eredményeket hasonlóan az előzőkhöz táblázatban mutatjuk be (3. táblázat).

A táblázat adataihoz nem fűzünk sok kommentárt, a kitűnő jelminőség önmagáért beszél, ilyen adatokkal élvezet tulajdonképpen a mérés, mert „tanítani” való eredményeket kaptunk. A 19. és 20. ábrákon az 51. csatornán kapott eredményeket mutatjuk meg. Jól látható hogy a teljes vivőcsoport ábrázolása mellett is tiszta a konstellációs ábra és a MER az alvivők függvényében a csatorna teljes keresztmetszetében azonosan kitűnő, 30 dB körüli értéket mutat.

19. ábra

A budapesti 51. csatornán dolgozó DVB-T adó összes alvivőt figyelembe vevő konstellációs ábrája



Érdekes eredményt hozott azonban a 43. csatornán végzett mérésünk. Itt a BER értékek rosszabbak, jelezvén, hogy az átvitelben valamilyen terjedési, vagy reflexiós probléma van. Ez megmutatkozik a csatornáról készített, átnézeti MER vizsgálat eredményeiből is (21. ábra).

Látható, hogy a csatornában a terjedés nem homogén, a 3000. vivő környékén jelentős interferencia lép fel, itt a MER értéke lezuhan a korábbi 32-33 dB-ről 10 dB környékére, majd ismét

nő, de a korábbi értéket már nem éri el, csak mintegy 30 dB-ig emelkedik, majd a 6000. vivő környékén újra meredeken lecsökken. A vett jel képi megjelenésén azonban a reflexiós probléma nem látható.

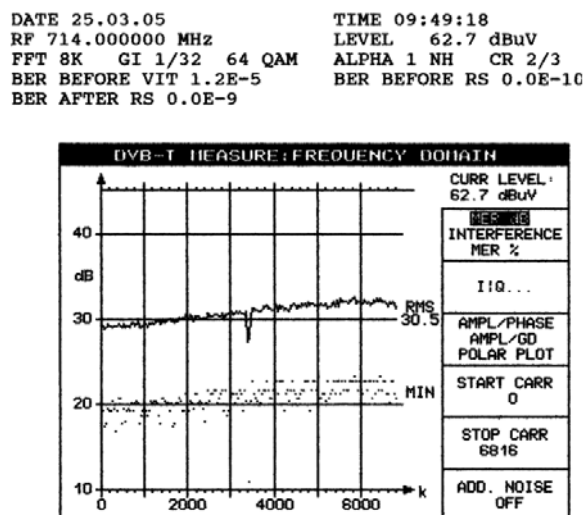
Az eset kapcsán érdekességgént bemutatjuk a Rohde&Schwarz EFA mérőművel végezhető mélyanalízist, amellyel megkereshetők és külön is vizsgálhatók a károsodott alvivők. A többször is szűkített keresés után kijelöljük a 3000-3010. alvivő csoportot, mert itt a legnagyobb a reflexió, ezzel a MER(k) függvényt kinyitjuk, majd a most már vivőnként meghatározott MER értékek közül kiemeljük a legrosszabbakat. Például a 3006. és a 3007 vivő ilyenek. A „legrosszabb” MER értékű, 3006. vivő konstellációs ábrája látható a 23. ábrán.

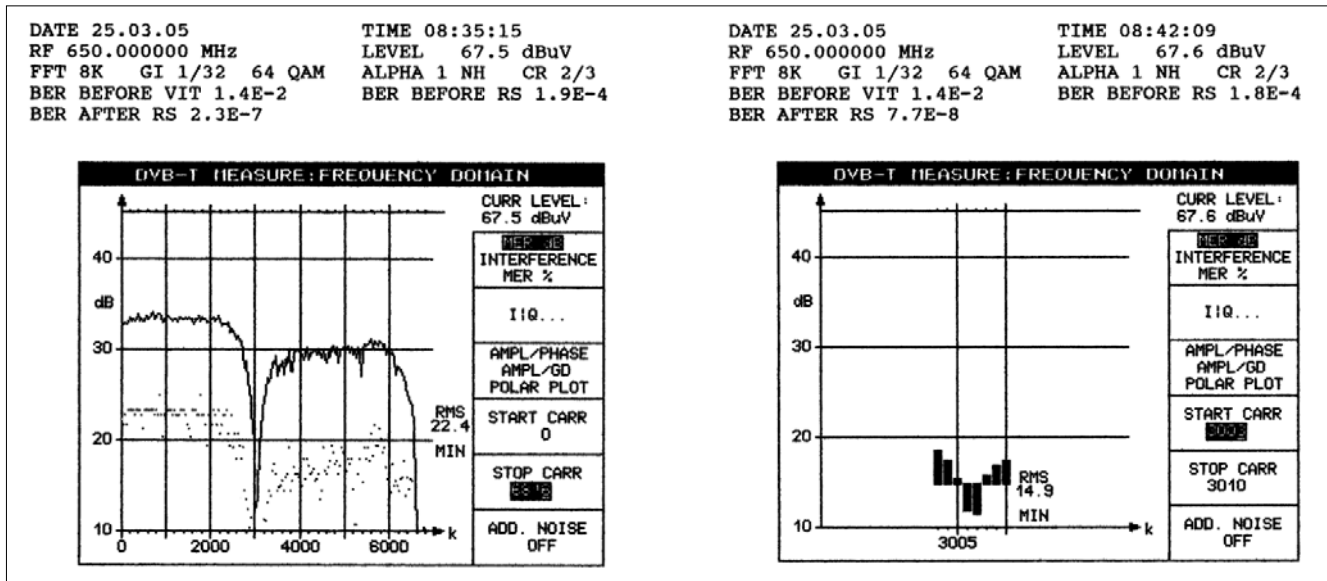
5. Összefoglalás

Jelen cikk, illetve az ezt megelőzően a témában elhangzott előadásunk rövid betekintést ad a témába végzett munkánkról. Tisztában vagyunk vele, hogy kevés a gyakorlati tapasztalatunk, hiszen mi is most tanuljuk a DVB technológiát, de mint felsőfokú oktatási in-

20. ábra

Az 51. csatornán a MER értéke az alvivők függvényében





21. ábra

A 43. csatorna vett jelének MER értéke az alvivők függvényében

22. ábra

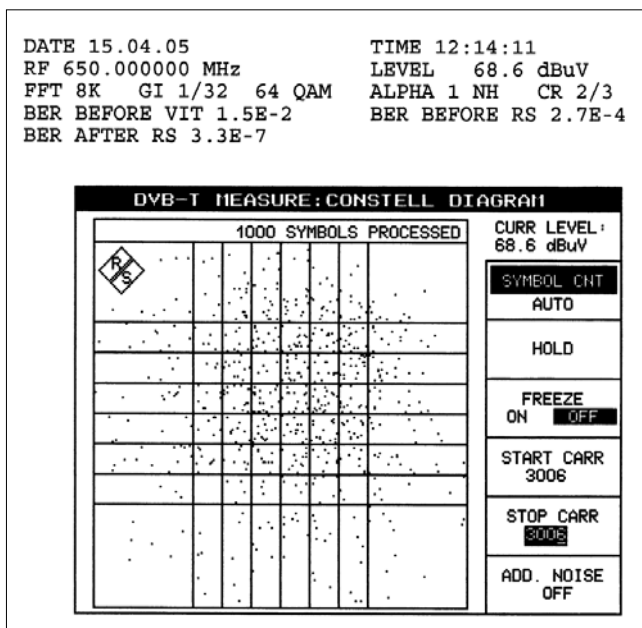
A segédvivők számának szűkítése lehetőséget ad részletesebb vizsgálatokra is

tézmény, amely kihatással van a szakmában élenjáró középfokú oktatásra is, jelezzük hogy részt kérünk és részt is vállalunk a feladatokból.

A Széchenyi István Egyetemen Győrben, illetve a Puskás Tivadar Távközlési Technikumban Budapesten már nem csak elvont kutatási munkák folynak, a DVB technológia bevonult a napi oktatásunkba is. Hallgatóink, diákjaink figyelemmel követik a hazai fejlesztéseket, megtalálhatók a szakmai fórumokon, konferenciákon is. Tudományos diákköri dolgozatokban és diplomatervekben jelzik, hogy ez a tématerület nem idegen számukra, sőt, ha lehet mondani már az ő technológiájuk, szeretnének minél aktívabban bekapcsolódni a megismerésébe, bevezetésébe.

23. ábra

Az analízissel kiemelt segédvivő konstellációs ábrája



Ehhez mára megteremtettük és folyamatos pályázatokkal szinten tartjuk intézményeink laboratóriumai-ban a műszaki, technikai feltételeket. A cikkben bemutatott mérésekben, azok előkészítésében, az egyes részfeladatok megoldásában diákjaink munkája is megtalálható.

A DVB-T kísérletek méréseiben, vizsgálataiban tehát potenciális tényezőként a budapesti és a kabhegyi sugárzást is felölölően készséggel közreműködünk, illetve részt kérünk a feladatokból. Ilyen feladat lehet például a hierarchia modulációs lehetőségek vizsgálata, továbbá a különböző reflexiós és terjedési jelenség tanulmányozása és elemzése is.

Irodalom

- [1] EN 300 744 – Digital Broadcasting Systems for Television, Sound and Data Services? Framing Structure, Channel Coding and Modulation for Digital Terrestrial Television.
- [2] W. Fischer: Digital Television Springer-Verlag Berlin, Heidelberg 2004.
- [3] EBU/CENELEC/ETSI, ETR-290, 1997. május, Digital Video Broadcasting? Measurement guidelines for DVB Systems
- [4] Rohde&Schwarz: Digital TV. Rigs and Recipes, part 4: DVB-T, Rohde&Schwarz, Broadcasting Division, 2001.
- [5] Rohde&Schwarz: TV Test Receiver EFA, Models 40/43 (DVB_T)
- [6] www.ahrt.hu/hu/szolgalatasok/tv_digitalis.php
- [7] Kathrein: SAT-TV-FM Test Receiver, MSK 33 Operating Manual

Audiovizuális beszéd-adatbázis és alkalmazásai

BÁRDI TAMÁS, FELDHOFFER GERGELY, HARCZOS TAMÁS,
SRANCSIK BÁLINT, SZABÓ GÁBOR DÁNIEL

Pázmány Péter Katolikus Egyetem, Információs Technológia Kar
{bardi, flugi, harczos, sraba, szasza}@digitus.itk.ppke.hu

Kulcsszavak: beszédszintézis, hallássérültek, beszélőfej-animáció

Cikkünkben egy audiovizuális beszéd-adatbázis összeállításáról számolunk be, melyet hallássérült emberek számára készülő szájmozgás-animációs programunk fejlesztésében szeretnénk hasznosítani. Részletesen kitérünk a tervezési szempontokra, és arra a siketeket érintő vizsgálatssorozatra, melyen alapulnak.

1. Bevezetés

Az utóbbi évtizedben a multimédia kommunikációs technológiai fejlődése következtében az emberi érzékelés kép és hang modalitásainak összefüggései, egymásra hatásai a tudományos érdeklődés homlokterébe kerültek, és megélénkültek a kutatások az audiovizuális beszédfeldolgozás területén.

A beszédhang, a kép és az elmondott szöveg hármasán belül lehetséges konverziók közül két irány élvez nagyobb népszerűséget a kutatók között. Az egyik az audiovizuális beszédszintézis, ahol egy animált fej szájmozgásait és esetleg az arc mimikai mozgásait, valamint az ezzel szinkronban lévő beszédhangot képezük a szöveg alapján. A másik az audiovizuális beszédfelismerés, ahol a hagyományos beszédfelismerőhöz képest a hang mellett a képi információt is felhasználjuk a biztosabb felismerés érdekében [1].

Mi az elmúlt évben azonban egy harmadik iránnyal kezdtünk foglalkozni, mely a felismerésnek és a szintézisnek egy sajátos keveréke, a beszédhangból beszélőfej animáció képzése –, angolul: speech to animation conversion.

A felmerülő első kérdés: mire használható egy ilyen? Egyrészt a szórakoztató iparban rajzfilmek, animációs filmek, valamint játékszoftverek készítését teszi gyorsabbá és olcsóbbá, mintha képkockaként „kézzel” kellene őket megrajzolni. Másrészt hallássérült emberek hátrányos társadalmi helyzetének kiküszöbölésében is alkalmazható. A siketek és nagyothallók többsége jól tud szájról olvasni, így egy megfelelő minőségű szájmozgás-animáció „láthatóvá” tenné számukra a beszédet.

Ilyen, „megfelelő” minőségű speech to animation conversion alkalmazás azonban működő rendszerben ma még nem létezik, de reméljük, hogy jelenleg is folyó munkánk eredményeképpen megoldást fogunk adni a problémára.

Jelen cikkünkben beszámolunk az ezirányú kutatásokról: röviden a fontosabb nemzetközi eredményekről, valamint saját erőfeszítéseinkről és a jövőbeni törekvéseinkről.

2. Helyzetkép

2.1. WISDOM

A WISDOM (Wireless Information Services for Deaf People) egy EU kutatás-fejlesztési projekt, amely 2001-ben indult és 2003 végén fejeződött be. A projektben az Egyesült Királyság, Svédország, Spanyolország és Németország vettek részt. Kutatólaboratóriumok, mobil szolgáltatók, rendszerszállítók, egyetemek fogtak össze. A célközösség természetesen a siketeket jelentette az adott országokban.

Szándékuk az volt, hogy a kutatási eredményeikre alapozva olyan szolgáltatásokat lehessen nyújtani, melyek megoldást adnak személyek közötti élő beszélgetésnél, és információ lekérdezésénél nyilvántartó rendszerekből. Mindezeket egy hordozható, kép átvitelére is alkalmas berendezésre és a siketek jelnyelvére alapozták.

A WISDOM projekt kifejezetten a harmadik generációs mobil rendszerekre épül, mert ezekkel nyílt meg a mobil mozgókép-továbbítás lehetősége. Bluetooth képességekkel rendelkező eszközöket kapcsoltak össze videojel-felvétel, tömörítés, szövegkezelés, és jelnyelvi felismerés céljaira.

2.2. SYNFACE

A SYNFACE a stockholmi Királyi Műszaki Egyetem (KTH) projektje (2001-2003) [2,3]. A projektben fix telefonkészülékekhez köthető, valós idejű, laptopon futó rendszert állítottak össze. (Alapelveiben jelen cikk szerzői is egy, a fentihez hasonló rendszer kialakításában gondolkodnak.)

A SYNFACE projektben olyan moduláris megközelítést választottak, amelyben a feladatot gyakorlatilag kettévágták egy felismerő és egy szintézis részre. Rendelkezésükre állt egy 3D-s animált beszélőfej, amit TalkingHead néven fejlesztettek egy korábbi projektjükben [4].

A TalkingHead bemenetként fonéma kódokból álló sztringet vár és ehhez képezi a szájmozgás animációt. (Csak a modell szája mozog, az arc többi része, a szemkörnyék és a homlok merev. Az élő emberi arcon itt tör-

ténnek a hangsúlyozáshoz és érzelem kifejezésekhez tartozó finom mimikai mozgások.) Ehhez készítették egy rejtett Markov-modelles (HMM) beszédfelismerőt, amely a beszédjelből előállítja a TalkingHead számára a fonéma sorozatot [5]. A felismerőjük nem szavakat ismer fel, hanem csak fonémákat. Így nagyobb a tévesztési arány, viszont a késleltetés kevesebb, mert nem szó hanem csak fonéma hosszúságú.

A beszéd fonémáknak megfelelő vizuális alapegységei a vízémák. Ezekből kevesebb van, mint a fonémákból, mert ha két fonéma képzése csak olyan jellemzőkben különbözik, ami kívülről nem látszik, akkor ugyanaz a vízéma tartozik hozzájuk. Ilyen jellemzők például, hogy a hangszalagok rezegnek vagy sem, vagy a nyelvcsap enged-e levegőt az orrüregén át vagy sem. Ennek megfelelően a 'papa', 'baba', 'mama' szavak szájáról olvasva elvileg megkülönböztethetetlenek. (Vannak igen gyakorlott siketek, akik esküsznek rá, hogy ilyenkor is látnak különbséget.)

A projektben próbálkoztak fonéma felismerő helyett vízéma felismerővel, ami elvileg egy könnyebb osztályozási feladat, hiszen kevesebb osztályba kell sorolni az akusztikus kulcsokat. De a tévesztési arány így nagyobb volt, vélhetően azért, mert vannak akusztikusan nagyon hasonló beszédhangok, amik különböző vízémákhoz tartoznak, és vannak amelyek akusztikusan nagyon különbözőek, de ugyanahhoz a vízémához tartoznak.

A program a sok munka és a részletek színvonalas kidolgozottsága ellenére a teljesen siketek esetében nem bizonyult használhatónak. Viszont a nagyothalló alanyok esetében sikeres volt: a telefonon érkező beszédet jobban értették, ha láthatták hozzá az animált szájmozgást a laptop képernyőjén.

3. A választott megközelítés

A kutatási projektünk eredeti célja segédeszközök fejlesztése siketek mobiltelefonos kommunikációjának elősegítésére. A segédeszközök között szerepelnek a telefonhoz illesztett kiegészítő eszközök, vagy a mobiltelefonon futtatott olyan alkalmazások, amelyek megkönnyítik, vagy lehetővé teszik siket emberek hallókkal folytatott mindennapi beszélgetését.

A segédeszközök fejlesztése érdekében számos előzetes vizsgálatot végeztünk siket emberek kommunikációs képességét és szokásait illetően. Azokban a helyzetekben, amikor a siketek nem láthatják a beszélő személyek száját, arcát (például telefonos kapcsolatnál), akkor a beszédjelből olyan mozgóképek ábrázolását választottuk, melyek a megértéshez elegendő képi információt hordoznak a beszédjelről.

A spektrogramok „olvasását” is meg lehet tanulni tartó gyakorlással. A szájáról olvasást viszont nem kell tanulniuk és gyakorolniuk a siketeknek, mert sokéves gyakorlással szinte művészetté tökéletesítették ezt a képességüket. Ezért gondoltunk olyan segédeszközben, ami ezt ki tudja használni. Úgy látjuk, hogy nem egy-

szerű, nem megoldott, de reálisan kitűzhető feladat a beszédjelből a szájmozgás képeinek gépi előállítás.

Ezzel szemben egy beszédhangból jelnyelvi animációt képző program készítése ma még reménytelen feladatnak látszik, mivel egy ilyen programnak meg kellene „érteni” a beszélő mondanivalóját. A jeltolmácsok emberi intelligenciája, amit beszédre jelre vagy jelről beszédre fordításkor aktívan használnak, kérdéses hogy gépi módszerekkel kiváltható-e egyáltalán.

3.1. Előzetes vizsgálatok

Nem szerettünk volna olyan eszköz fejlesztésébe kezdeni, amit a siket és nagyothalló emberek nem akarnak, vagy nem tudnak használni. Ezért előzetes „igényfelmérés” gyanánt interjúkat készítettünk, valamint a hallókétól eltérő képességeiket kísérletekben mértük. Ezekben a felmérésekben összesen 14 siket és erősen nagyothalló személy vett részt. Összehasonlításként mindegyiket elvégeztettük halló alanyokkal is.

Először interjúkat készítettünk velük, ahol konkrét történeteket meséltek el, melyekben a halló és siket emberek közötti párbeszéd valamiért elakadt, sikertelen volt, vagy félreértés történt. Ezután a hallókétól eltérő nyelvi képességeiket, szövegértésüket írásbeli feladatlapokkal vizsgáltuk.

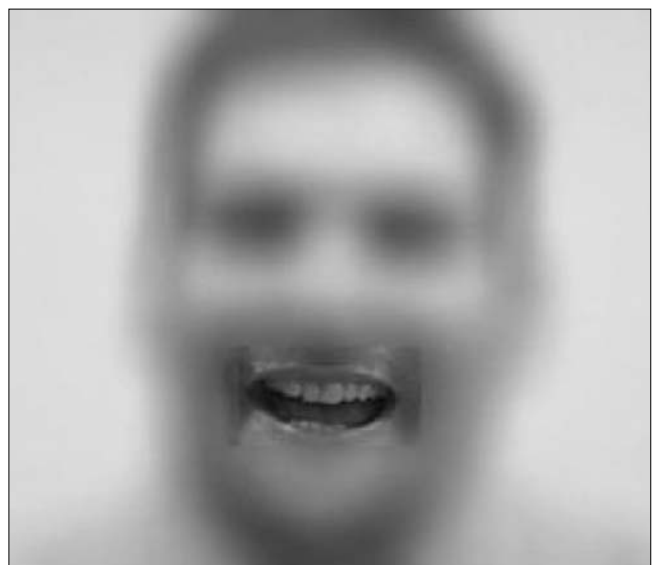
Szájról olvasási képességüket különböző videó tesztekkel vizsgáltuk. A vetített felvételeken egy-egy ember beszélt szemből, közelről a kamerába nézve, és a szájáról kellett leolvasni, hogy mit mond. A hangot minden esetben kikapcsoltuk.

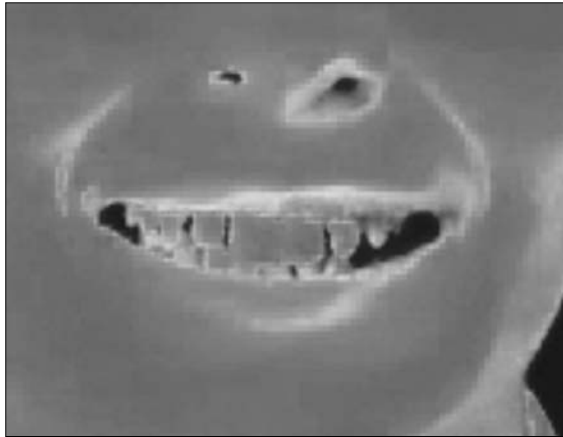
A felvételek különböző minőségűek és felbontásúak voltak. A frame-rate 15, 25 és 30 kép volt másodpercenként. Jó minőségű kivetítőként projektort használtunk. A gyengébb minőségű kivetítő eszközt a Sony Ericsson P910-es mobiltelefon 208x320 pixeles színes kijelzője jelentette (40x61 mm).

A videókat különböző módon manipuláltuk:

- „blur”: a száj és közvetlen környékén kívül a képet elhomályosítottuk (1. ábra),

1. ábra „Blur” fedi az arc nagy részét





2. ábra Kékített videó

- kékítés: az RGB színtelepeket eltoltuk a kék felé (2. ábra).
- binarizálás: a képet a pixelek világossága alapján kétszínűvé, fekete-fehérré konvertáltuk. Egy világosság-küszöböt beállítva fekete és fehér pixeleket kaptunk, a köztes szürke értékeket nem engedték meg (3. ábra).



3. ábra Binarizált videó

Összehasonlításképpen a kékített és a kétszínű videók mellett vetítettünk felvételeket az eredeti színekkel is. Ezeken a formák térbelisége jobban kivehető (4. ábra).

4. ábra Videó az eredeti színekkel



3.2. A vizsgálatok tanulságai

Az elvégzett vizsgálatok alapján a fejlesztési folyamatot is befolyásoló tanulságokat szűrtünk le.

Ne legyen feltűnő

A legtöbb hallássérült nem szívesen hívja fel a figyelmet erre a tulajdonságára, mindaddig, amíg ez nem feltétlenül szükséges. A fehér botot használó vakokkal, vagy a kerekesszékekkel közlekedő mozgássérültekkel szemben a siketség (nagyothallás) első ránézésre egyáltalán nem látszik, és ők törekednek is rá, hogy amíg csak lehet, olyanak tűnjenek, mint a halló többség. Emiatt sokuk tartózkodik olyan technikai eszközök használatától, ami már messziről „elárulja” hallássérülését. Vannak akik még a hallókészüléket is szégyellik hordani, bár ma már ez alig látszik.

Ebből a szempontból egy mobiltelefonon futó program mint segédeszköz úgyszólván tökéletes. Mobilja szinte mindenkinek van, használója általában nem kelt feltűnést. Siketek is előszeretettel használják, jellemzően az üzenetküldő szolgáltatásokat. Emellett egy rendkívül kompakt eszköz, vihető bárhová és hosszabb ideig bírja akkumulátor feltöltés nélkül, mint egy laptop.

A szájról olvasás vizuális feltételei

A szájról olvasás még a leggyakorlottabb siketek számára is nehéz és fárasztó feladat, ami koncentrált figyelmet igényel. A beszélőnek és a szájról olvasónak egyenesen egymásra kell néznie közben, a fejüket nem forgathatják, és ügyelniük kell a megfelelő távolságra.

A beszéd szájról olvashatóságának két legfontosabb kritériuma a beszéd sebessége és artikuláltsága. Lassan és tisztán, látványosan jól artikuláltan kell beszélni. A hallók számára megszokott beszédtempó a szájról olvasáshoz túl gyors, nem érthető, ilyenkor gyakran előfordul, hogy a siket nem érti amit mondanak neki, csak bólogat. A halló emberek beszéd artikulációja nagyon változó, szájról olvashatóság szempontjából igen erős szórást mutat. Vannak, akik „lapos szájjal”, nagyon alulartikuláltan beszélnek, és a siketek akkor sem értik őket, ha külön a kedvükért lelassítanak, sőt még ismétlés után sem.

A legérthetőbben artikulálók jellemzően azok a siketek közt élő vagy dolgozó emberek (halló rokonok, jel-tolmácsok), akiknek a kommunikációs és empátikus készségük egy folyamatos visszacsatolást biztosít, aminek köszönhetően javítani tudják a beszédjük szájról olvashatóságát. Ők egy idő után ösztönösen kihangsúlyozzák a szájmozgásukban azokat az apró elemeket, melyek alapján a hasonló vizémák megkülönböztethetők. Jól választják meg a tempót, és a nehezebben érthető részeket az arc kifejezésükkel hangsúlyozzák, és ott még lassabbra váltanak.

Azt tapasztaltuk, hogy ezek az artikulációs jellemzők sokkal többet számítanak, mint a kép felbontása vagy a frame-rate. Mindezek alapján arra a következtetésre jutottunk, hogy az audiovizuális beszéd-adatbázisunk felvételéhez jeltolmácsokat alkalmazunk. Így remélhetőleg olyan tanító adatokat kapunk, mely alapján az animált beszélőfejük szájmozgásában az említett artikulációs jellemzők érvényesülnek majd.

A szájról olvasás nyelvi feltételei

Szájról olvasáskor fontos szerepe van a nyelvi intelligenciának. Beszélgetéskor a halló emberek sem értenek tisztán minden hangot az elhangzott szavakból, hanem a hiányzó részeket a nyelvi tudásukra és a társalgás aktuális kontextusára alapozva kiegészítik agyuk. Az egyes beszédhangok azonosítása a szájmozgás alapján bizonytalanabb, mint hallás alapján. Emiatt szájról olvasáskor több a hiányzó vagy bizonytalan elem, amit a kontextus és a nyelv alapján kell kirakni fejben, mint a hallók beszédértésében.

A siketek nyelvhasználatukban aránylag kevés figyelmet fordítanak a ragozásra. Magyarban a ragozott szóalakok jellemzően csak egy-két nehezen leolvasható, könnyen összetéveszthető mássalhangzóban különböznek. Erre mindig tekintettel kell lenni a megfogalmazásban. Például annál a mondatnál, hogy *‘Haza megyek’* jobban értik azt, hogy *‘Én haza megyek’*.

Mindezek fényében teljesen irreális volna azt várni, hogy tetszőleges témájú, nyelvileg bonyolult mondatokat szájról olvasás után szó szerint visszaadjanak. Ehelyett a témát előzetesen megadtuk, és csak egy-két szóval kellett kérdésekre válaszolniuk, amivel ellenőrizhettük, hogy mit értettek meg. A vizsgálatok egy részében tudhatták, hogy kétjegyű számokat kell leolvasniuk, ezzel a „keresési tartományt” jelentősen leszűkítettük.

A mobilkészülék kijelzője elegendő

Úgy találtuk, hogy a Sony Ericsson P910-es készülék kijelzőjének elég jó a felbontása a szájról olvasáshoz. A jól artikulált felvételek esetén a mobiltelefon kijelzőjéről és a jó minőségű projektor képéről egyaránt 80-90%-os felismerési aránnyal olvasták le a kétjegyű számokat. A felismerés a kékitett és a binarizált felvételeken is ugyanilyen megbízható volt.

2D vagy 3D

Az emberi agy több módszert is bevet a látott kép térbeli rekonstrukciójában. Egy arc nézésekor leginkább a gömbölyded felületek folytonos színáttnései és a finom fényárnyék hatások játszanak szerepet. A kékitett és a binarizált felvételeken viszont ezek jelentősen torzulnak. Úgy gondoltuk, ha ezekről tudnak szájról olvasni, akkor elég 2 dimenziós fejmodellben gondolkodni. A felmérések eredményei megerősítették várakozásainkat. Ennek alapján 2D-s audiovizuális adatbázis felvétele mellett döntöttünk. Így a több kamera képéből számított 3D-s mozgás követés technikai problémáival nem kell törődnünk.

4. Adatbázis-összeállítás és -felvétel

Egy rendszer tanításához és ellenőrzéséhez olyan adatbázisra van szükség, ami megfelel a teljességi, a pontossági, és a statisztikai elvárásoknak. Az adatok felvétele előtt mondatokat és diádokat állítottunk össze úgy, hogy a lehetséges fonéma együttállásokat a felvétel lefedje, ehhez a lehetséges diádok felsorolását választottuk. A példamondatok a magyar fonémakészletet a nyelv átlagos eloszlása szerint használták.

Ezek a követelmények beszédfelismerési feladatokban már ismertek és jól dokumentáltak, az újdonságot a videó anyag rögzítése jelentette. Olyan videó anyagra van szükség, ami pontos és használható információt tartalmaz. Ha egyszerű, jól megvilágított arcot fényképezünk le, akkor abból nehéz absztrakt információt kivonni, ezért kitüntetett pontok követését tűztük ki célul.

Az arc teljes körű modellezését már elvégezték és szabványosították MPEG-4 néven. Ez a szabvány leír többek között 84 feature pointot (FP) az arcon. Úgy döntöttünk, hogy amennyire lehetséges, e szabvány szerint dolgozunk. A képeket úgy vettük fel, hogy az arcon megjelöltük néhány FP-t. Ez a jelölés bőrszemleges textílfestéssel történt, ami könnyen lemosható, nem irritálja a bőrt, és a kiszínezésének köszönhetően kényelmesen felvihető 2-3 mm vastag foltokban. Ezzel az eljárással az a probléma, hogy az ajkak belső oldalára nem lehet pontot tenni, mert az elkenődne beszéd közben, így csak bizonyos pontokat vettünk fel. Nem jelöltük meg a siketek által nem figyelt pontokat, csak a száj környékén szemből látható területen törekedtünk teljességre (5. és 6. ábra).

A megvilágítást úgy választottuk meg, hogy a későbbiekben automatikusan tudjuk a pontokat megkeresni a felvételen. Mivel a beszéd közben ezen pontok mozgását kell megállapítani, fontos, hogy a felvétel felbontása minél jobb legyen. A kameráink a piacon elterjedt digitális kamerákhoz hasonlóan 720x576 képpont felbontásúak. Két kamerával dolgoztunk, amik egyszerre indultak el. Az egyik kamera a teljes arcot fényképezte, 90 fokos elforgatással, hogy a vízszintes képet jobban kihasználjuk, a másik kamera csak a száj környékét, hogy a számunkra legfontosabb FP-k mozgását minél pontosabban meghatározhassuk. A két kamera és a hang szinkronitását ellenőriztük, az esetleges elcsúszások nem haladják meg az egy frame-t. Azzal, hogy az egyik kamera csak a száj körüli pontokat figyelte, elértük, hogy egy FP mozgását átlagosan vízszintesen 40-60 pixel, függőlegesen 80-140 pixel felbontással lehessen rögzíteni.

A kamerák képét közvetlenül videószalagra készítettük, majd a felvételek után DV AVI formátumban digitalizáltuk. A pontosság érdekében ennél jobban nem tömörítettük. Összesen négy jeltolmács beszélt a felvételeken, 76 perccel. A teljes kép- és hanganyag 36 GB-ot foglal.

5. Az MPEG-4 FAP pontok kinyerése a videó felvételekből

A videó felvételen előszűrést végeztünk. A megvilágítástól függően olyan intenzitásintéket állítottunk be, amik kiemelik a megfestett pontokat. Egy vágás után a pontok egy-egy összefüggő foltot jelentenek, amin eróziót hajtunk végre, hogy megkapjuk a foltok közepét pixel pontossággal. Ez az erózió akkor pontos, ha a folt konvex, ami sokféleképpen sérülhet. A legegyszerűbb

eset a gyors mozgásnál fellépő pontatlan interlaced kép, ami egy folt helyett vízszintes vonalakat ad. Ennek elkerülése érdekében elmostuk a képet, mielőtt a vágás megtörtént volna. Így már csak ritkán volt probléma a pontok követésével, ez pedig abból ered, hogy bizonyos szájtartásoknál az ajak lebiggyed, és az általunk festett pontot eltakarja, vagy árnyékot vet rá. A számításokat MATLAB programmal végeztük, mert a videó fájlokhoz való hozzáférést ez nagyon könnyűvé tette.

6. Továbbtekintés

A továbbiakban az adatbázis képanyagából rendelkezésünkre álló pont koordinátákat és a beszédhangból nyert sajátság vektorokat (MFCC) neurális hálók tanításához fogjuk felhasználni. Az így betanított neurális hálók fogják majd a beszédhez tartozó animációs paramétereket automatikusan előállítani. Az adatbázis feldolgozása és a tanítási folyamat erősen számítás igényes, ehhez a ma kapható csúcskategóriás PC-ket kell használnunk. De a létrejövő rendszer várhatóan egy jobb mobiltelefonon is futtatható lesz. [6] Reméljük, hogy olyan minőségű szájmozgás animációt sikerül ezzel elérni, amelyet siketek az élő arcot közelítő megbízhatósággal le tudnak olvasni. Ilyen tudomásunk szerint a világon ma még nincs.

Köszönetnyilvánítás

A szerzők hálás köszönetüket szeretnék kifejezni az NKTH-nak a projekt támogatásáért, a közreműködésért a SINOSZ-nak, valamint a siketeknek és jeltolmácsok-

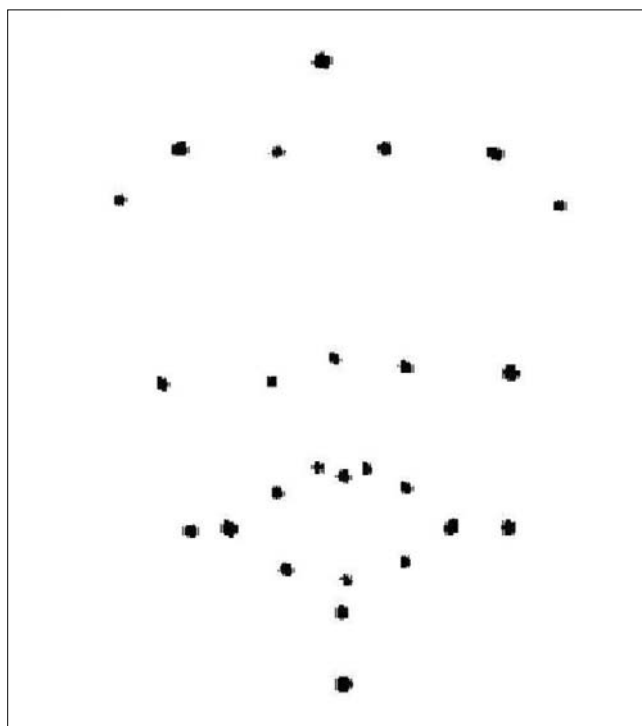
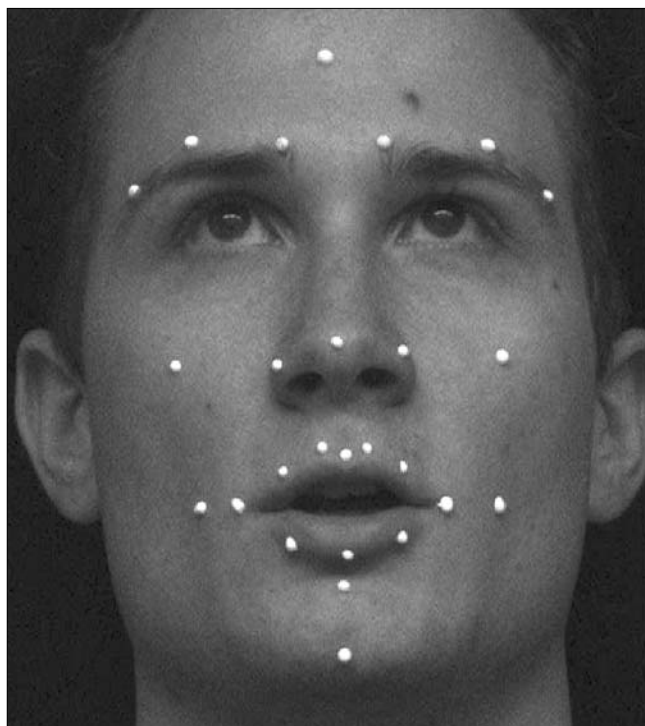
nak, akik részt vettek a programban. A T-Mobile-nak köszönjük, hogy a telefon készülékeket a rendelkezésünkre bocsátotta. Külön köszönjük Dr. Takács Györgynek az értékes segítségét és a témavezetést.

Irodalom

- [1] Czap L.: Audiovizuális beszédfelismerés és szintézis, PhD értekezés, 2005.
- [2] B. Granström, I. Karlsson, K-E Spens: „SYNFACE – a project presentation” Proc. of Fonetik 2002, TMH-QPSR, 44: 93–96.
- [3] M. Sheard, N. Thomas: „The SYNFACE project: development and evaluation of a talking face telephone” The HCI 2003 – Designing for Society Conference in Bath, England
- [4] J. Beskow: Talking Heads – Models and Applications for Multimodal Speech Synthesis. Doctoral Thesis. Department of Speech, Music and Hearing, KTH, Stockholm (2003)
- [5] G. Salvi: „Truncation error and dynamics in very low latency phonetic recognition” Proc. of ISCA workshop on Non-linear Speech Processing (2003)
- [6] Feldhoffer G., Bárdi T., Jung G., Hegedűs I. M.: „Mobiltelefon alkalmazások siket felhasználóknak”, Híradástechnika, 2005.

5. és 6. ábra

Megfestett referencia pontok az arcon, felvétel közben és a referencia pontok kinyerése a felvételből



Mobiltelefon alkalmazások siket felhasználóknak

FELDHOFER GERGELY, BÁRDI TAMÁS, JUNG GERGELY, HEGEDŰS IVÁN MIHÁLY

*Pázmány Péter Katolikus Egyetem, Információs Technológia Kar
{flugi, bardi, junge, hegivmi}@digitus.itk.ppke.hu*

Kulcsszavak: mobiltelefon, Symbian mobil operációs rendszer, hallássérültek

Cikkünkben egy olyan rendszert írunk le, ami lehetővé teszi siket vagy hallássérült emberek beszédérzékelését olyan esetekben, amikor nincs lehetőségük szájáról olvasni. A célunk hordozható készülék, lehetőleg egy mobiltelefon fejlesztése, ami beszédhangból közvetlenül egy modellezett száj vagy fej mozgását tudja előállítani. Ez a cikk a fejlesztés mobiltelefon-specifikus részleteire koncentrál.

1. Bevezető

A napjainkban kapható legfejlettebb mobiltelefon készülékek gyors processzoroknak, nagy tárhelynek és multimédiás képességeiknek köszönhetően igen sokrétűen használhatók. A már megszokott hang és üzenetküldő szolgáltatások mellett fényképezőgépként, szövegszerkesztőként, és már televízióként is működnek. A készülékekben rejlő lehetőségek a már meglévő hasznos és kevésbé hasznos alkalmazásokon túlmutatnak. A gyártók és a szolgáltatók üdvözlnek minden jó elképzelést, amely valóban hasznos és vonzó funkciókkal egészíti ki a telefonjaikat. Ilyen jellegű fejlesztésekre igyekeznek ösztönözni külső cégeket, szervezeteiket.

A PPKE ITK-n az elmúlt évben egy kutatás-fejlesztési projektbe kezdtünk, melynek fő célja, hogy hallássérült felhasználók mobiltelefonos kommunikációs lehetőségeit javítsuk, minél jobban kiaknázva e készülékek képességeit. A beszéd teremti talán a legközvetlenebb és hatékonyabb kommunikációs kapcsolatot ember és ember között, ezért nem meglepő, hogy a telefon eredendően és legnagyobb részben még ma is beszéd átvitelére szolgál. Azonban a hang információ a siketek és nagyothallók számára hozzáférhetetlen.

A projektünk célja, hogy beszédből egy okostelefon által más modalitásban megjeleníthető információt képezzünk, amely tényleges segítséget jelenthet a siketek hallókkal való kapcsolat tartásában. Konkrétan egy a telefon grafikus kijelzőjén megjelenő animált beszélő fejre gondoltunk, melynek szájmozgásairól a beszéd leolvasható [1].

A másik lehetőség egy a telefonhoz csatlakoztatható taktilis kijelző, mely a tapintási érzékünkre hat. Mindkét elképzeléshez szükség van valós időben működő beszédhangelemző modulra, ezért első lépésben ezt készítettük el.

Jelen cikkünk tárgya ez a Symbian operációs rendszerre fejlesztett program, mely a beérkező beszédjelből a legtöbb automatikus felismerő rendszerben használatos mel-kepsztrum sajátásvvektorokat képezi. Ér-

demesnek látjuk beszámolni a fejlesztési folyamat mobiltelefonos környezet adta nehézségeiről, sajátosságairól.

2. A mobiltelefon mint platform

Egy ma kapható okostelefonban 100-200 MHz-es órajelű processzor van, de a leggyorsabbak akár 250 MHz-nél magasabb órajelű frekvenciával is működnek és belső memóriájuk 64 MB is lehet. A 90-es évek elején az első hordozható személyi számítógépeknek ennél jóval kisebb volt a teljesítményük. Például az első kereskedelmi forgalomban megjelent laptop, a NEC Ultra-lite-nak csak 8 MHz-es órajele volt. Nem meglepő tehát, hogy napjaink telefonjai a legkülönbözőbb alkalmazások futtatására képesek.

A mobiltelefonokra írt alkalmazások gyors terjedésének a másik oka, hogy több rendszerhez is létezik nyilvános fejlesztőkörnyezet. Ilyen a Symbian OS is, de ilyennek tekinthető bármely Java MIDP futtatásra alkalmas készülék is.

Létezik ezen kívül a PocketPC család, ami Windows CE vagy Windows Mobile rendszert használ, ezek egyelőre mobiltelefon funkcióval nem terjedtek el, mint ahogy a PalmOS alapú kézi számítógépekre sem jellemző a GSM kommunikáció. A Symbian-os készülékek megjelenése előtt a telefonon futó programokat kizárólag a gyártók készítették, és a felhasználó utólag nem tudott azokra újabbakat telepíteni. A szabványosított nyílt rendszerek ezzel szemben lehetővé teszik, hogy a felhasználó harmadik fél által készített alkalmazásokat telepítsen, így létrejött a telefonon futó programok piaca [2,3].

2.1. A Symbian mobil operációs rendszer

A Symbian konzorciumot a Nokia, az Ericsson, a Psion és a Motorola alapította 1998-ban. A Symbian-t úgy tervezték, hogy kompakt hordozható eszközök operációs rendszereként működjön. A mobil piacon az első sikert a 2000-ben bemutatott 6.0 verzió aratta. A 7.0-as verzió már kifejezetten mobil telefonokra optimalizált

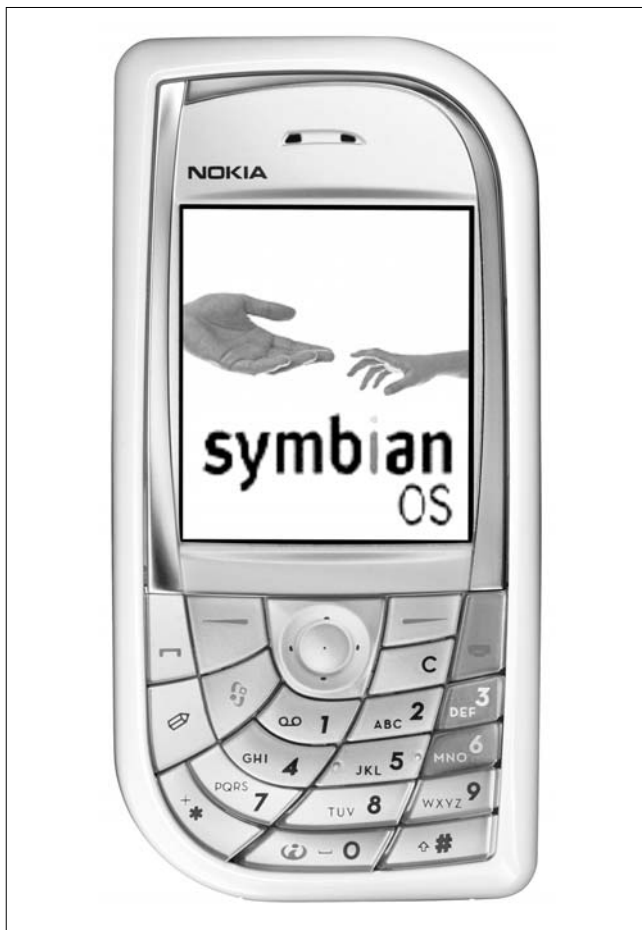
rendszernek számított, az idén megjelent 9.0-ás változat pedig a 3G mobil hálózatokon is működtethető.

Ahogy a telefonok piaca fejlődik, a keretrendszereknek alkalmazkodniuk kell a növekvő igényekhez. Ez olyan fejlesztési eljárás létrejöttéhez vezetett, ahol az operációs rendszerek szinte készülékről készülékre változnak. Miközben a jó operációs rendszer egyik fontos tulajdonsága a hordozhatóság, jóformán alig találni két egyformán programozható telefonmodellt, mert egymást érik a fejlesztések. Egy konkrét példa erre a Nokia telefonok Series 60-as sorozatának (S60) fejlődése. Az S60-as szabványnak a 2.0 verziója 2003-ban, a 3.0-ás 2005 közepén jelenik meg, és a kettő között háromszor bővült a rendszer olyan szolgáltatásokkal, melyek az előző szabványokban nem szerepelnek. Szerencsére ezek a sorozatok visszafelé kompatibilisek.

2.2. Az alkalmazott készülékek

Két típuson dolgoztunk: Nokia 7610-esen (1. ábra) és Sony Ericsson P910-esen. Mindkettőnek StrongARM processzora van. Ez egy RISC processzor, amit csekély tárigényűre és fogyasztásúra alakítottak ki. A gépi kód mérete ilyen processzorokon töredéke az x86-osokénak. A Nokia 176x208-as és az Ericsson 208x320-as képfelbontása alacsonynak tűnhet, de ha a kijelző méretét is figyelembe vesszük, DPI-ben ezek a folyadékkristályos kijelzők a modern monitorokéval egy kategóriába esnek.

1. ábra A Nokia 7610-es



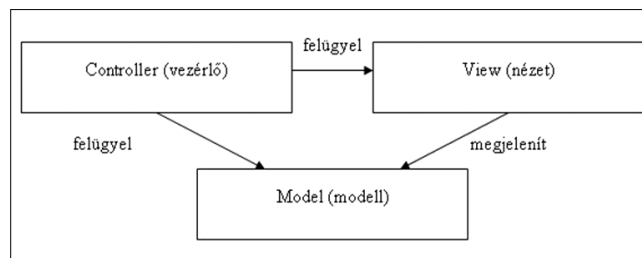
3. Fejlesztés Symbianra

A hagyományos gyakorlat azt várná el, hogy a platform hordozza a fejlesztőeszközöket, de nem nehéz belátni, hogy a mobiltelefon nem a legalkalmasabb eszköz programszöveg kezelésére. Ehelyett keresztfordító programokat használnak, amelyek lehetővé teszik, hogy a fejlesztő PC-n megírja a programot, majd lefordítsa a készülékre.

A folyamat kényelmesebbé tehető, ha van egy emulátor az eszközhöz, amivel az alkalmazás már a PC-n ellenőrizhető. Az említett gyors piaci fejlődés az operációs rendszerre való fejlesztésben viszont nehézségeket okoz. Az emulátorok általában csak a nagy verzióugrásoknál készülnek el, ezért gyakori, hogy egy-egy alkalmazás csak telefonon fut, az emulátoron viszont nem, mert kihasznál az emulátorból még hiányzó szolgáltatásokat. Az ellenkező eset is felléphet, az emulátorban működő program nem feltétlenül fordítható le az eszközre, mert PC-n jóval enyhébb megkötések vannak például a verem méretére. Összességében az emulátorokról azt lehet mondani, hogy segítenek az apró fejlesztési lépésekben, de nem garantálják a helyes működést.

Az operációs rendszer szolgáltatásait C++ osztály hierarchiába szervezték, támogatva az objektumorientált programozást. Nem ritka, hogy egy szolgáltatás igénybe vételéhez a hiányzó lényegi részeket absztrakt osztálytól kell örökölni, és megvalósítani.

Az operációs rendszer szolgáltatásait ugyanis nem egy egyszerű API, hanem egy teljes fejlesztői könyvtár formájában nyújtják, ahol a programot előre adott szerkezetben érdemes felépíteni. Ez a Model–View–Controller (MVC) architektúra, ahol a program tartalma, ki- és bemeneti lehetőségei és képe egymástól elkülönül (2. ábra). Ezt a szerkezetet ösztönzi, hogy a Symbian rendszer rendelkezik ősosztályokkal, amikből az MVC komponensek örökölhetők, és így a főbb kontrollfolyamatokat átvállalják a programozótól. Ez nagymértékben hasonló a mai ablakkezelést támogató PC-s fejlesztőkörnyezetekhez.



2. ábra Az MVC modell

Mindazonáltal a szabványos C++ nyelvnek nem teljes egésze használható a Symbian fejlesztőkörnyezetben. Az STL (Standard Template Library) nem használható, ezért az eredetileg a C++ szabványhoz tartozó könyvtár helyett a Symbian saját konténereit ajánlják. Az STL hiányát azóta egy a Symbiantól független csoport orvosolta, és elkészült az STL port, ami egyszerre

rúbb rendszereszközökkel valósítja meg az STL szolgáltatásait. Azonban a Symbian még a C++ nyelv szabályait is szigorította, melyek egy része nem feltétlenül látszik szükségesnek.

A legfőbb lehetőségek, amiktől a Symbian megfosztja a C++ programozót, a kivételkezelés, a többszörös öröklődés egy része, és több kényelmes lehetőség az objektumok kezelésére. A megszorításokat a Symbian alkotói minden esetben a hardver korlátaival magyarázzák, ami a gyorsan fejlődő eszközök mellett több esetben nehezen érthető.

3.1. A hangkezelés nehézségei

Egy olyan alkalmazásnak, amely hang elemzésével képi és taktilis információkat állít elő, bizonyos alapkövetelményeknek meg kell felelnie. Képesnek kell lennie a hang folytonos feldolgozására, lehetőség szerint minél kisebb átmeneti tárolóval, és ügyelni kell arra, hogy a rendszer késleltetése ne növekedjen. A taktilis információk közvetítéséhez a hardverre és a szoftverre egyaránt szükség van. A hang elemzéséhez pedig memória- és processzorkímélő algoritmusok olyan kialakításra van szükségünk, ami a kritikus követelményeknek megfelel.

A Symbian első verziói a hang kezelésében a felvételre és lejátszásra szorítkoztak, ilyenkor egy függvényhívással megkezdődött a rögzítés, egy másik hívással pedig befejeződött. Ez a működés kényelmetlen a folyamatos hangkezeléshez, mivel nincs kettős puffereles, így a szakasz felvételének befejezése és a következő kezdése közötti idő elvész.

Az újabb verziókban ezt a területet kibővítették és lehetőség nyílik audio stream használatára, ahol már csak a callback függvényt kell megírni, amivel a megtelt puffer tartalmát feldolgozhatjuk. Az általunk használt telefon rendszere nem tartalmazza ezt, ezért egy hiánypótló MdaAudioInputStream könyvtárat használtunk, ami kettős puffereles nélkül tud folyamatosan hangot kezelni.

A grafika az egyik legkényelmesebben használható terület a Symbian OS programozásában. Mivel a programunk beleágyazódik az űsosztályokba, a kirajzolásért felelős programrészek egyszerűen kezelhetők.

A mobiltelefonok egy része szabványos csatlófelülettel rendelkezik, ahol lehetőség van tetszőleges hardver csatlakoztatására. Ez adta az ötletet, hogy a siketek körében sikerrel alkalmazott taktilis kijelzők, mint például a rezgőmotorral ellátott gyűrű, részét képezze a rendszernek. Folyik a kutatás egy olyan vibrátor építésére, ami olyan hangelemző eljárás eredményeképpen rezeg, ami a siketeknek a legjobban feldolgozható információt adja.

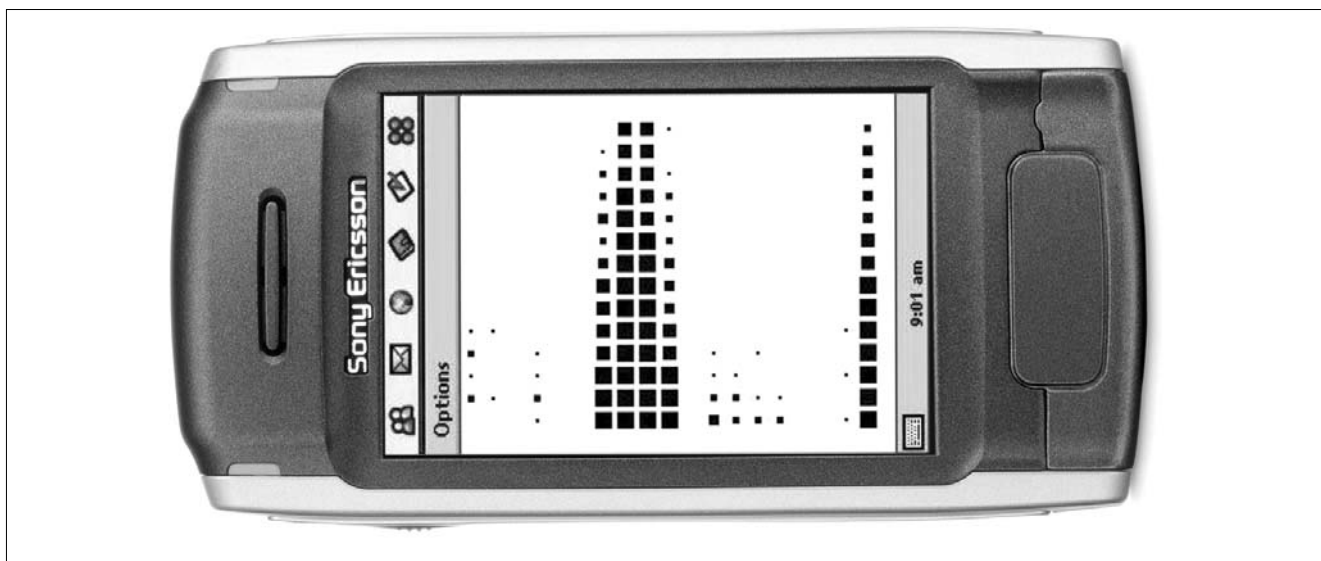
3.2. A programmotor elkészítésének folyamata

Az egyik sarkalatos pont a fejlesztési terv kidolgozása a hibakeresési módszerek kialakítása. Sajnos a Symbian-rendszerek, mint elsősorban grafikus rendszerek, az alfanumerikus kijelzésben a kijelző természete miatt korlátozottak, ráadásul a szöveges és numerikus adatok keverése a konzol alapú környezetekhez képest bonyolult programkódot kíván. Ezért az a döntés született, hogy a külön tesztelhető részeket PC-n készítjük el, és az így kapott programmotort utólag építjük be a rendszerbe. Így jött létre az FFT algoritmus is, amit DOS/Windows környezetben készítettünk el, majd a jól ellenőrzött kódot az MVC szerkezetnek megfelelően illesztettük a Symbian-ra írt programba. Ez a módszer zökkenőmentesen működött a mel-skála programozásánál is. Mivel a hangból képet előállító rendszert még kutatjuk, egyelőre a telefonokra nem készítettünk további elemző részeket.

4. A beszédelemző

A bejövő beszédjelet a programunk 40 ms hosszú átfedés nélküli ablakokra bontja. Hamming ablak alkalmazása után a mel-spektrum szerint 16 sávban átlagoljuk

3. ábra A spektrumelemző futás közben a SonyEricsson készülékén



az intenzitást. A kiszámított értékeket a kijelzőn a sötét színű négyzetek nagyságával érzékeltetjük, a *foto*n látható módon (3. ábra). Számítási szempontból a beszédelemző lelke a gyors Fourier-transzformáció. Ehhez a Decimation-in-time Radix-2 FFT algoritmust írtuk át Symbian-ra.

A Radix-2 algoritmus előnye a kis kód méret, ami mobiltelefonon fontos szempont. A Radix-4 és más FFT változatok csak néhány százalékkal gyorsabbak, viszont a kód méretük többszöröse a Radix-2-nek. További előny, hogy 2-hatvány méretű tömbökön tud dolgozni, és nem csak 4-hatvány méretűeken.

Az eljárás verzióinkban 32 bites lebegőpontos számábrázolást használ (TReal32), amely pontosság a céljainkhoz elegendő. A telefonok processzorait jellemzően nem a lebegő pontos műveletek gyors elvégzésére tervezték. Egy fix-pontos számokkal dolgozó FFT sokkal gyorsabban lefutna, de akkor a számértékek túlcsondulását megakadályozandó túl sokat kellene engedni a numerikus pontosságból.

A tárháshasználat csökkentése érdekében néhány DSP programozásban használatos trükköt vetettünk be. Az egyikre azért van szükség, mert az FFT komplex számokat vár bemenetként, viszont az ablakozott beszédjel valós. A legegyszerűbb megoldás az lenne, hogy az imaginárius részt feltöltjük mindenütt nullákkal, ami viszont a tárigényt fölöslegesen megkétszerezné. Ezért inkább a valós tömbünket egy fele annyi elemből álló komplex tömbként értelmezzük, és erre számolunk FFT-t. Ebből lineáris időben dolgozó algoritmussal egyszerűen ki lehet számítani az eredeti valós vektor Fourier-transzformáltját.

A másik speciális megoldás, hogy az algoritmus „helyben” számol. Ez azt jelenti, hogy az összes műveletet a bemenetként kapott tömbön hajtja végre, és ebbe kapjuk a végeredményt is, így nem kell átmeneti tömböket pluszban lefoglalni. Az ilyen, helyben dolgozó FFT-k jellemzően bitreverz sorrendben adják ki a Fourier-transzformáltat, amit utólag lehet egy egyszerű rutin meghívásával rendes sorrendbe átrendezni. Mindezeket a belső műveleteket egy C++ osztályba csomagoltuk, hogy az objektumot használó programozó számára láthatatlanok maradjanak.

5. Összegzés

A mobiltelefonok mérete, kijelzője, és programozási kapacitása elegendő ahhoz, hogy további hasznos alkalmazásokat lehessen a siketek számára fejleszteni. A további kutatásoktól várjuk a választ arra a kérdésre is, hogy van-e az erőforrásokkal dolgozó algoritmusok között olyan, melynek hibaaránya már elfogadhatóan csekély.

A következő kutatási fázisban egyrészt algoritmusokat fogunk összehasonlítani a pontosság és az erőforráshasználat szempontjából, másrészt a kijelző hatékonyságát kívánjuk növelni. A pontosságot személyi számítógépeken és mobiltelefonokon futtatott tesztek-

kel vegyesen, egy siketekből álló tesztcsoporton fogjuk meghatározni.

Ezzel a technológiával lehetségessé válhat a beszédalapú alkalmazások sora, akár beszélőazonosítás, akár felismerési feladatok területén. A folyamat teljesen a telefonkészüléken fut, a szolgáltatótól nem vesz igénybe többlet számítási kapacitást.

Köszönetnyilvánítás

A szerzők hálásak az NKTH támogatásáért, a SI-NOSZ közreműködéséért és hasznos tanácsaiért, valamint a T-Mobile-nak, hogy a készülékeket rendelkezésünkre bocsátotta. Külön köszönjük Dr. Takács Györgynek az értékes segítségét és a témavezetést.

Irodalom

- [1] Bárdi T., Feldhoffer G., Harczos T., Srancsik B., Szabó G. D.: „Audiovizuális beszéd-adatbázisok és alkalmazásai”, Híradástechnika, 2005.
- [2] Charaf H., Csúcs G., Forstner B., Marossy K.: Symbian alapú szoftverfejlesztés, Szak, 2004.
- [3] J. Stichbury: Symbian OS Explained, Wiley, 2005.



Least squares szupport vektor gépek adatbányászati alkalmazása

VALYON JÓZSEF, HORVÁTH GÁBOR

Budapesti Műszaki és Gazdaságtudományi Egyetem, Méréstechnika és Információs Rendszerek Tanszék
{valyon, horvath}@mit.bme.hu

Kulcsszavak: adatbányászat, Least Squares szupport vektor gépek, függvény approximáció, idősor előrejelzés

Napjainkban a számítástechnika, ezen belül is az adatbázisok és adattárházak elterjedt alkalmazásának köszönhetően az élet számos területén könnyedén nagy mennyiségű adat halmozható fel, melyek megfelelő feldolgozása, elemzése hasznosítható eredményekre vezethet. Az adatbányászat olyan eljárások és módszerek összessége, melyek segítségével feltárhatók felhalmozódott adatok közt rejlő, korábban ismeretlen összefüggések, rejtett trendek, szabályszerűségek. A klasszikus adatbányászat számos tudományterület, mint például a lineáris algebra, a gráfelmélet, az adatbázis elmélet, az algoritmus elmélet, a gépi tanulás, illetve a mesterséges intelligencia eszközeit felhasználja.

Cikkünkben a rendszerek, bemeneti és kimeneti adatai ismeretében történő fekete doboz modellezésével foglalkozunk, ahol a rendszer modelljét a rendelkezésre álló adatok alapján, ezek elemzésével kell meghatározni. Ezek általában nehezen algoritmizálható, specifikus megoldásokat igénylő feladatok, így gyakran intelligens rendszerek, illetve a „lágy” számítási módszerek (soft computing) kerülnek alkalmazásra. Kiemelten foglalkozunk a neurális hálózatok (Neural Networks – NN) egy különleges típusával, a szupport vektor gépek (Support Vector Machines – SVM) Least Squares változatával, az LS-SVM-mel, valamint bemutatjuk ennek egy módosított verzióját az LS²-SVM-et. Az előbbi módszerek felhasználásával cikkünkben a függvény approximációra és idősor adatok elemzésére, illetve előjelzésére is mutatunk megoldásokat.

1. Bevezetés

Az élet számos területén találkozunk olyan komplex rendszerekkel, melyek működése, jellemzői, belső összefüggései nem ismertek. Ilyen problémákkal találkozhatunk például az orvosi diagnosztika, az ipari és a gazdasági folyamatok területén. Ezen rendszerek működése során azonban általában rengeteg bemeneti és kimeneti adat gyűjthető, melyek feldolgozásával információkat nyerhetünk, vagy megalkothatjuk a rendszer modelljét. Ezt fekete doboz (black-box) modellezésnek nevezzük.

Az adatbányászat feladata, hogy a leggyakrabban adatbázisokban, illetve adattárházakban felhalmozott adatok alapján olyan összefüggéseket, információkat tárjon fel, melyek közvetlenül nem állnak rendelkezésre.

Az adatbányászat eszközei között megtaláljuk a gépi tanulást, a különféle „soft computing” technikákat és ezen belül a neurális hálózatokat is [1,2]. A neurális hálózatok többek között általános eszközt adnak a (nem-lineáris) regressziós feladat megoldására. Cikkünkben

ezek egy speciális fajtájával a szupport vektor gépekkel (SVM) foglalkozunk [3,4], melyek a gépek rejtett rétegében elhelyezkedő neuronjaiban található bázis-függvényekből állítják elő a kívánt függvényt. Az SVM-ek működésének lényege, hogy az eredeti megfogalmazásában még komplex nemlineáris megoldást igénylő feladatot, azaz a feladatból származó mintákat, nemlineáris transzformációk segítségével egy a bemeneti mintatér dimenziójánál több dimenziós térbe transzformálja, ahol az már lineárisan megoldható.

A módszer egyik legnagyobb előnye, hogy egy garantált felső korlátot ad az approximáció általánosítási hibájára. Egy másik fontos jellemzője, hogy a tanulási algoritmus törekszik a modell méretének minimalizálására (ritka modellt alkot), ami a hiba rováására történik, de mértéke egy paraméterrel szabályozható [9].

A hagyományos SVM alkalmazásának legnagyobb akadálya a módszer nagy algoritmikus komplexitása és a nagy memóriagigény, ami tipikusan a nagy adatmennyiség kezelését teszi lehetetlenné. A probléma megoldására számos megoldás született. Ezek az algoritmusok többnyire iteratív megoldások, melyek a nagy optimalizálási feladatot kisebb feladatok sorozatára bontják. Az egyes szeletelési „chunking” algoritmusok főként a feladat dekomponálás módjában – a részfeladatok meghatározásában – különböznek [5-7].

Az algoritmikus komplexitás legyőzésének egy másik lehetséges módja az LS-SVM alkalmazása, amikor is a számítási nehézségeket jelentő kvadratikus programozás helyett egy egyszerű mátrix inverziót használunk.

Az adatbányászati feladatokban általában nagy mennyiségű adattal kell dolgozni, ezért különösen fontos, hogy – ellentétben a hagyományos LS-SVM-el – a kiszámított modell komplexitása független legyen a tanítópontok számától. A következőkben az LS-SVM olyan kiterjesztésére adunk eljárásokat, amelyek lehetővé teszik, hogy egyszerűen, hatékonyan és kontrollálható módon hozzassunk létre ilyen modelleket.

A fekete doboz modellezési problémák körében általános feladatként két feladatcsoport jelölhető meg: (a) függvény approximáció (regresszió), illetve dinamikus rendszerek esetén (b) az idősor elemzése, előrejelzése.

Függvény approximáció esetén a rendelkezésre álló adatok közötti függvénykapcsolat elemzésére van lehetőség, azaz N adatból álló p -dimenzós mintakészlet esetén elemezhető, hogy a változók egyike (ez a rendszer kimenete, ami gyakran ismert) milyen függvénykapcsolatban áll a többi $p-1$ adattal (vagy annak bármilyen részalmazával). Ebben az esetben a minták sorrendjének nem tulajdonítunk jelentőséget, feltételezzük, hogy a rendszer statikus, azaz kimenete csak a pillanatnyi bemenettől függ.

Idősor előrejelzés esetén feltételezzük, hogy a rendszernek van emlékezete (pl. visszacsatolás, memória), azaz dinamikus rendszer. Az előrejelzési feladatban a bemeneti és/vagy kimeneti adatok egymást követő értékei alapján próbálunk egy olyan modellt konstruálni, ami jól reprezentálja a folyamat dinamikáját, így képes folytatni az idősort (jóslni a későbbi kimeneti értékeket). Ez a feladat is általánosítható, azaz az előrejelzés nem csak az időtengely, hanem tetszőleges változó mentén lehetséges (például annak monoton növekvő értékeire). A rendszerek jelentős része valamilyen dinamikát mutat, azaz a folyamat pillanatnyi értéke nem csupán az adott bemeneti jeltől, hanem az előzményektől – a rendszer állapotától – is függ. Ebben az esetben is valójában egy regressziós feladatot (függvény approximációt) kell megoldani, de úgy, hogy a függvény bemeneti változói a korábbi értékekkel kibővülnek.

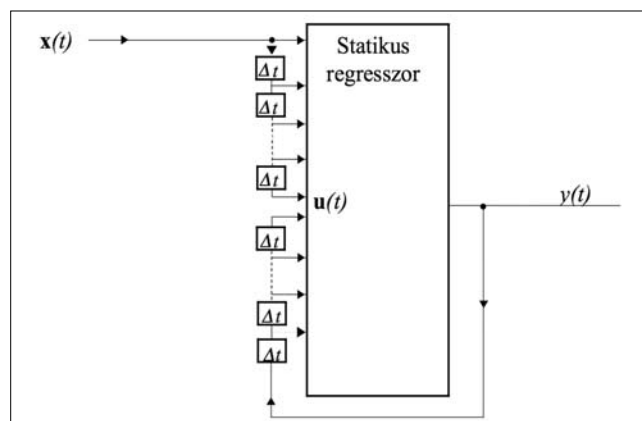
A következő rész bemutatja, hogy az idősor előrejelzés hogyan vezethető vissza a függvény approximációra. A 3. rész az LS-SVM regressziót és annak javasolt módosítását az LS-LS-SVM (LS^2 -SVM) regressziót mutatja be, majd a cikk a 4. részben szereplő alkalmazási példák utáni áttekintéssel zárul.

2. Bemeneti mintahalmaz előállítása dinamikus rendszerekhez

Az előrejelzési feladat során, egy időben változó értéksorozat valahány elmúlt értékének (a folyamat régebbi mintavételi értékeinek) ismeretében a következő értéket vagy értékeket kell előre jelezni, jóslni (a későbbi mintavételi értékekre becslést adni).

Természetesen ilyen esetekben a feldolgozó algoritmusnak vagy modellező eljárásnak is a megfelelő időfüggést kell mutatnia. Ezt legegyszerűbben úgy érjük el, hogy a statikus feladatok megoldására kialakított (rendszerint nemlineáris) modellt egészítjük ki dinamikus (rendszerint lineáris) komponensekkel (legegyszerűbb esetben egyszerű késleltetéssel, visszacsatolással).

Leggyakrabban a tisztán statikus modellt „kívülről” egészítjük ki dinamikus komponensekkel. Erre számos lehetőség van, de talán a legegyszerűbb és legelterjedtebb az 1. ábrán látható.



1. ábra
Késleltető sorokkal dinamikussá tett statikus rendszer

Az ábrán látható statikus modell úgy tehető dinamikussá, hogy a bemenő adathalmazt késleltetett értékekkel bővítjük. Ehhez a bemeneti N dimenziós x vektorokból $N+K$ dimenziós vektorokat állít elő úgy, hogy az N dimenzió mellé egyes bemenetek, illetve kimenet késleltetett értékét (értékeit) is felveszi.

$$u^T(t) = [x_1(t), x_1(t - T_{11}), \dots, x_1(t - T_{K_{11}}), \dots, x_i(t), x_i(t - T_{1i}), \dots, x_i(t - T_{K_{1i}}), \dots, y(t - T_{1Y}), \dots, y(t - T_{K_{YY}})] \quad (1)$$

ahol:

$x_i(t)$ az i -edik bemenet a t időpillanatban,

$x_i(t - T_{1i})$ az i -edik bemenet a $t - T_{1i}$ időpillanatban,

$[T_{1i}, \dots, T_{K_{1i}}]$ az i -edik bemenet késleltetéseinek halmaza ($i=1 \dots N$)

K_i az i -edik bemenet késleltetéseinek száma.

$u^T(t)$ az előállított $N+K$ dimenziós vektor a t időpillanatban.

$[T_{1Y}, \dots, T_{K_{YY}}]$ az y kimenet késleltetéseinek halmaza,

K_Y az Y kimenet késleltetéseinek száma,

y a kimenet.

Az itt bemutatott megadás meglehetősen általános, hiszen minden bemenet esetén egyedileg és tetszőlegesen lehet megadni a késleltetéseket. Ezt a gyakorlati megoldásokban gyakran korlátozzák:

- nem bemenetenként szabályozható késleltetések,
- nem tetszőleges késleltetések, hanem például ablakméret megadás (utolsó n minta) stb.

Az így átalakított, kibővített adatsorra már a statikus regresszió alkalmazható, melynek megvalósítására a következő fejezetben mutatunk megoldást.

A tanítás után a becslés ugyanígy a késleltetett értékekkel kibővített bemenetre történik. Előrejelzés esetén iteratív időlépésenként kell kiszámítani az eredményt, hiszen minden kimenetre szükség lehet a további kimenetek számításához.

3. Függvény approximáció

A regresszió, illetve a függvény approximáció megvalósítására számos különböző módszer áll rendelkezésre.

A módszerek több szempont szerint is rendszerezhetők, de a megoldható feladatok köre, illetve a megvalósítás szempontjából a következőket emeljük ki:

- Lineáris regresszió
- Nemlineáris regresszió

Ez a csoportosítás azért fontos, mert a későbbiekben bemutatásra kerülő bázisfüggvényes módszerek a nemlineáris regressziót a jóval egyszerűbb – hatékonyabban számítható – lineáris regresszióra vezetik vissza. A feladat mindegyik esetben a következő:

Adott $\{\mathbf{x}_k, d_k\}_{k=1}^N$ tanító ponthalmaz, ahol $\mathbf{x}_k \in \mathbb{R}^p$ egy p -dimenziós bemeneti vektor, illetve a $d_k \in \mathbb{R}$ a kívánt kimenet. A cél egy $y = f(\mathbf{x})$ függvény megadása, ami jól reprezentálja a tanítópontok által leírt kapcsolatot.

Az ismert analitikus megoldások (lineáris/polinomiális regresszió, spline-ok) mellett a függvény approximációs feladat neurális hálózatokkal is megoldható [1,2]. Az alábbiakban bemutatott szupport vektor módszer a Least Squares SVM, az LS-SVM is ide sorolható [8].

Mint említettük az SVM-ek működésének lényege, hogy az elsődleges térben még komplex nemlineáris megoldást igénylő feladatot, azaz az ezt leíró mintákat, nemlineáris transzformációk segítségével egy magasabb dimenziós térbe transzformálják, ahol az már lineárisan megoldható. További tulajdonságuk, hogy bizonyos válogatást végeznek a rendelkezésre álló bemeneti mintapárok között. Olyan mintákat választanak ki – ezek a szupport vektorok – melyek a megoldás szempontjából relevánsak, míg a többit eldobják. Így valójában egy ritka (sparse) megoldást hoznak létre [9]. A least-squares megoldás (az LS-SVM) a tanulás gyorsítása érdekében éppen ezt a ritkasági tulajdonságot áldozza fel. A cikkben javasolt részleges redukciós eljárás (az LS²-SVM) lehetővé teszi, hogy hasonlóan a hagyományos SVM megoldáshoz a modell mérete, illetve a hiba mértéke hangolható legyen, miközben az algoritmikus komplexitás tovább csökken [10,11].

3.1. LS-SVM regresszió

Az LS-SVM a hagyományos, Vapnik által bevezetett szupport vektor gépek (SVM) [3] egy módosított verziója, ami egy lineáris egyenletrendszer megoldására vezet, ellentétben a hagyományos módszerben alkalmazott idő és erőforrás-igényes kvadratikusan programozással. A fő előnye ennek a módszernek, hogy a számítási komplexitás ezáltal csökken.

Az SVM-eknél a megoldást $y(\mathbf{x}) = \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}) + b$ alakban keressük, ahol a $\boldsymbol{\varphi}(\cdot): \mathbb{R}^n \rightarrow \mathbb{R}^{n_h}$ egy nemlineáris leképezés egy többdimenziós térbe, ahol $n > n_h$, sőt sokszor $n \gg n_h$. A bemenetet tehát egy magasabb dimenziós térbe transzformáljuk, majd itt a tanítás során kiszámított lineáris megoldás együtthatói alapján számítjuk az eredeti feladat megoldását.

A szupport vektor gépeknél ehhez egy további feltételt fogalmazunk meg ($\mathbf{w}$ hosszának, $\mathbf{w}^T \mathbf{w}$ -nek minimalizálása), ami azt biztosítja, hogy a tanítópontokhoz minél jobban illeszkedő, de egyben minél simább meg-

oldást nyerjünk, így az általánosítási hibát minimalizáljuk. A regresszió számításához az alábbi optimálási feladat írható fel (2):

$$\min_{\mathbf{w}, b, e} J_p(\mathbf{w}, e) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + \gamma \frac{1}{2} \sum_{k=1}^N e_k^2,$$

a $d_k = \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_k) + b + e_k$, ahol $k = 1, \dots, N$ feltételekkel.

A fenti egyenletekből az alábbi Lagrange multiplikátoros egyenlet írható fel (3):

$$L(\mathbf{w}, b, e; \boldsymbol{\alpha}) = J_p(\mathbf{w}, e) - \sum_{k=1}^N \alpha_k \left\{ \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_k) + b + e_k - d_k \right\},$$

ahol az α_k értékek a Lagrange multiplikátorok.

A $\mathbf{w}$ és e kifejezése után a következő lineáris egyenletrendszer írható fel:

$$\begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & \boldsymbol{\Omega} + \gamma^{-1} \mathbf{I} \end{bmatrix} \begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{d} \end{bmatrix} \quad (4)$$

ahol:

$$\begin{aligned} \mathbf{d}^T &= [d_0, d_1, \dots, d_N], \quad \mathbf{1}^T = [1, \dots, 1], \quad \boldsymbol{\alpha}^T = [\alpha_0, \alpha_1, \dots, \alpha_N], \\ \boldsymbol{\Omega}_{i,j} &= \boldsymbol{\varphi}(\mathbf{x}_i)^T \boldsymbol{\varphi}(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j), \\ i, j &= 1, \dots, N. \end{aligned}$$

Az eredmény, hasonlóan a hagyományos SVM-hez a következő alakban írható fel:

$$y(\mathbf{x}) = \sum_{k=1}^N \alpha_k K(\mathbf{x}, \mathbf{x}_k) + b \quad (5)$$

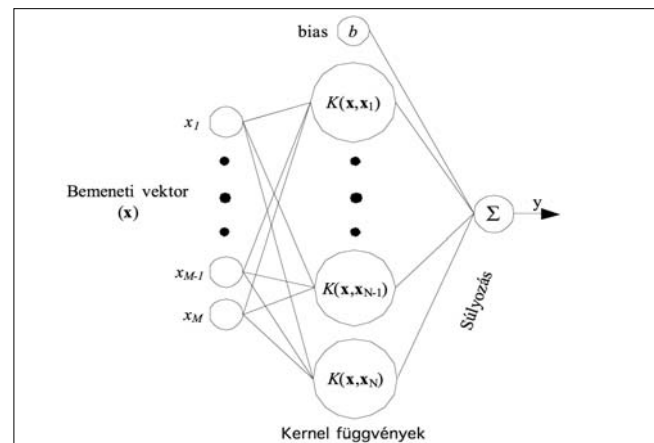
ahol az α_k és b a fenti egyenletrendszer megoldása.

Látható, hogy az eredeti $y(\mathbf{x}) = \sum_{i=1}^N w_i \varphi_i(\mathbf{x}) + b$ megoldáshoz hasonlóan egy lineáris megoldásra jutottunk, ahol a nemlineáris leképezést egy új, az eredeti $\boldsymbol{\varphi}(\cdot)$ leképezések szorzatként előálló $K(\cdot, \cdot)$ kernel függvénye váltja fel.

Bár a gyakorlatban az SVM-eket ritkán alkalmazzuk neurális hálózat alakjában, ez az értelmezés nagyon hasznos, mert egyszerűbb tárgyalásmódot tesz lehetővé a pusztán matematikai megközelítésénél. A szupport vektor gépek tanítása, és használata is matematikai számítások sorozata, a válasz számításának képlete pontosan megfeleltethető egy rejtett rétegű neurális hálózatnak (2. ábra). A rejtett réteg tipikusan nemlineáris neuronokat tartalmaz.

2. ábra

A szupport vektor gépnek megfeleltethető neurális hálózat



A bemenet egy M -dimenziós vektor. A nemlineáris kernel függvényeket a rejtett réteg neuronjai tartalmazzák. Ezen neuronok száma megegyezik a szupport vektorok számával (LS-SVM esetén a ritkaság hiányában N , azaz a minták száma). A hálózat válasza (y) a rejtett réteg neuronjai kimenetének súlyozott összege. Az α_k súlyok a tanítás során kiszámított Lagrange multiplikátor értékek. Ennek megfelelően, minél kisebb a hálózat, annál kevesebb számításra van szükség a válasz megadásához, így a cél a lehető legkisebb hálózat elérése.

3.2. LS^2 -SVM regresszió

Az LS^2 -SVM fő célja, hogy csökkentse a modell komplexitását, azaz a rejtett rétegben nemlineáris neuronok számát. A javasolt módszer egy járulékos előnye, hogy a modell számításának algoritmikus komplexitása is csökken. Az új megközelítés kiindulópontja a (4) képlet lineáris egyenletrendszer.

A javasolt módszer két fontos lépést tartalmaz:

- (i) Az *első lépés* átalakítja az LS -SVM megoldást úgy, hogy az a tanítópontoknak csak egy részhalmazát használja „szupport vektornak”. Megmutatjuk, hogy az így kapott túlhatározott egyenletrendszer is megoldható.
- (ii) A *második lépésben* a „szupport vektorok” (a kernel függvényeket meghatározó vektorok) automatikus meghatározására adunk módszert.

Túlhatározott egyenletrendszer

Ha a tanító készlet N mintapontot tartalmaz, akkor az egyenletrendszer $(N+1)$ ismeretlent, az α -kat és a b -t, $(N+1)$ egyenletet és $(N+1)^2$ együtthatót tartalmaz. Az együtthatók a $K(\mathbf{x}_i, \mathbf{x}_j)$ $i, j = 1, \dots, N$ kernel függvény értékei. A mintapontok száma tehát meghatározza az egyenletrendszer méretét, ami egyúttal a megoldás komplexitását, a hálózat méretét. Hogy ritka megoldást, azaz egy kisebb modellt kapjunk, az egyenletrendszert, illetve az együttható-mátrix méretét redukálni kell.

Vizsgáljuk meg közelebbről az LS -SVM regressziós problémát leíró egyenletrendszert és jelentését.

Az első sor jelentése:

$$\sum_{k=1}^N \alpha_k = 0, \quad (6)$$

míg a j . sor a

$$b + \alpha_1 K(\mathbf{x}_j, \mathbf{x}_1) + \dots + \alpha_k [K(\mathbf{x}_j, \mathbf{x}_k) + C^{-1} \mathbf{I}] + \dots + \alpha_N K(\mathbf{x}_j, \mathbf{x}_N) = d_j \quad (7)$$

feltételt, megkötést tartalmazza.

Az egyenletrendszer redukálásánál sorokat, illetve oszlopokat hagyhatunk el.

1) Ha a j . **oszlopot** hagyjuk el, akkor az ennek megfelelő α_j szintén „eltűnik”, így az eredményként megkapott hálózat mérete csökken. Az első sor által támasztott feltétel azonban automatikusan alkalmazkodik, hisz a megmaradó α_k -k összege 0 marad.

2) Ha a j . **sort** töröljük, akkor az $(\mathbf{x}_j, d_j)$ tanító pontnak megfelelő információ elvész, hiszen a j . megkötést (7) elveszítjük.

Ezek alapján az egyenletrendszert leíró mátrix legfontosabb része $\Omega + C^{-1} \mathbf{I}$ részmátrix, ahol Ω az összes lehetséges tanító vektor kombinációját tartalmazza ($\Omega_{ij} = K(\mathbf{x}_i, \mathbf{x}_j)$). A mátrix redukálása során abból sorokat, oszlopokat, illetve mindkettőt (sort a hozzá tartozó oszloppal) törölhetünk.

Minden oszlop egy-egy neuronnak felel meg, míg a sorok bemenet-kimenet relációkat, azaz a megoldás által teljesítendő feltételeket fogalmaznak meg. A mátrix az alábbi két módon redukálható:

Hagyományos teljes redukció – a $(\mathbf{x}_i, d_i)$ tanító pontot teljesen elhagyjuk, azaz mind a hozzá tartozó sort, mind pedig az oszlopot töröljük. Az alábbi egyenlet a tanító minták teljes elhagyásának hatását mutatja. A törölt elemek szürkével jelöltek.

$$\begin{bmatrix} 0 & \Omega_{00} + \frac{1}{C} & \Omega_{01} & \dots & \Omega_{0N} \\ \Omega_{10} & \Omega_{11} + \frac{1}{C} & \dots & \dots & \Omega_{1N} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \Omega_{(N-1)0} & \Omega_{(N-1)1} & \dots & \dots & \Omega_{(N-1)N} \\ \Omega_{N0} & \Omega_{N1} & \dots & \dots & \Omega_{NN} + \frac{1}{C} \end{bmatrix} \begin{bmatrix} b \\ \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix} = \begin{bmatrix} 0 \\ d_0 \\ d_1 \\ \vdots \\ d_N \end{bmatrix} \quad (8)$$

A hagyományos metszési (pruning) technika esetében pontosan ez történik, hisz a metszési algoritmus iteratívan kihagyja a tanító pontok egy részét. Az elhagyott tanító pontok által hordozott információ ezért teljesen elvész. A tanítópontok által hordozott információ megőrzésére a részleges redukciót javasoljuk.

Részleges redukció – a $(\mathbf{x}_i, d_i)$ tanító mintát csak részben hagyjuk el, úgy, hogy töröljük a ponthoz tartozó oszlopot, de megtartjuk a hozzá tartozó sort. Így a tanító minta által hordozott megkötés továbbra is érvényben marad, hisz az adott sor súlyozott összegének amennyire csak lehet egyezni kell a d_j kívánt kimenettel.

Ha kiválasztunk $M (M < N)$ „szupport vektort”, az egyenletrendszer túlhatározottá válik. A részleges redukció hatását a mutatja a (9) egyenlet, ahol az eltávolított részek szürkével jelöltek.

$$\begin{bmatrix} 0 & \Omega_{00} + \frac{1}{C} & \Omega_{01} & \dots & \Omega_{0N} \\ \Omega_{10} & \Omega_{11} + \frac{1}{C} & \dots & \dots & \Omega_{1N} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \Omega_{(N-1)0} & \Omega_{(N-1)1} & \dots & \dots & \Omega_{(N-1)N} \\ \Omega_{N0} & \Omega_{N1} & \dots & \dots & \Omega_{NN} + \frac{1}{C} \end{bmatrix} \begin{bmatrix} b \\ \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix} = \begin{bmatrix} 0 \\ d_0 \\ d_1 \\ \vdots \\ d_N \end{bmatrix} \quad (9)$$

A fentiek alapján egy túlhatározott egyenlethez jutunk, amit négyzetes hiba-minimalizálással oldhatunk meg. Egyszerűsítsük jelöléseinket az alábbiak szerint:

$$\mathbf{A} = \begin{bmatrix} 0 & \mathbf{I}^T \\ \mathbf{I} & \Omega + C^{-1} \mathbf{I} \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} b \\ \alpha \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} 0 \\ \mathbf{d} \end{bmatrix}. \quad (10)$$

A négyzetes hiba minimalizáló megoldás:

$$\mathbf{u} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{v}. \quad (11)$$

Az oszlopok törlése a sorok megtartása mellett biztosítja, hogy a neuronok száma csökkenjen, míg az összes ismert megkötést (mintapontot) figyelembe vesszük. Ez kulcsa annak, hogy a megoldás komplexitása a pontosság megtartása mellett csökkenthető legyen.

A módosított, részlegesen redukált egyenletrendszert least-squares értelemben oldjuk meg, ezért nevezzük ezt a módszert Least Squares LS-SVM-nek vagy röviden LS^2 -SVM-nek.

Az itt bemutatott részleges redukció hasonlít a hagyományos SVM kiterjesztéseként bevezetett redukált szupport vektor gép (Reduced Support Vector Machine – RSVM) alapelveihez [12]. Az RSVM esetén azonban, mivel az SVM eleve kisebb (ritka – sparse) modellt eredményez, ennek célja az algoritmikus komplexitás csökkentése, míg az LS-SVM esetén célunk a modell komplexitásának, a hálózat méretének csökkentése, azaz a ritka LS-SVM.

A kiválasztási eljárás

A fenti részleges redukció alkalmazása esetén szükség van valamilyen módszerre, ami meghatározza a szükséges szupport vektorokat. A kernel mátrix oszlopaiból kiválasztható egy olyan lineárisan független részhalmaz, melyek lineáris kombinációival a többi előállítható. Ez a kernel mátrix „bázisának” meghatározásával érhető el, hiszen az oszlopvektorok terének bázisa az a legkisebb vektorkészlet, amellyel a feladat megoldható. Az itt említett bázis csak egy bizonyos tolerancia értelmében határozható meg, hiszen célunk, hogy az N darab N dimenziós oszlopvektorból $M < N$ bázisvektort határozzunk meg.

A kiválasztási eljárás tartalmaz egy paramétert, ami kontrollálja a szükséges szupport vektorok számát (M). A szupport vektorok száma valójában nem függ a mintapontok számától (N), csak a probléma nehézségétől, hiszen M a mátrix lineárisan független oszlopainak száma. A gyakorlatban ez azt jelenti, hogy ha egy probléma nehézsége M neuront igényel, akkor a tanításhoz felhasznált mintapontok számától függetlenül a modell mérete nem növekszik.

A szupport vektorok kiválasztása az A^T mátrix reduced row echelon alakra hozása során részleges pivotálással végrehajtott Gauss-Jordan eliminációval történik [13,14]. A tolerancia figyelembevétele a pivot elem (p) ellenőrzésével történik. Eredetileg a pivotolás lényege, hogy az elimináció végrehajtása során szükséges osztásban a lehető legnagyobb elemet használjuk fel a numerikus stabilitás érdekében. A kiválasztott pivot elem nagysága azonban információt hordoz arról is, hogy mennyire fontos a hozzá tartozó oszlop a pontos megoldáshoz.

Ha $p \leq \varepsilon'$ (ahol ε' a tolerancia paraméter), akkor az adott oszlop elhagyható, ellenkező esetben az oszlopnak megfelelő bemenet egy szupport vektor. A módszer azon oszlop vektorok listáját adja meg, melyek az ε' toleranciaérték értelmében lineárisan függetlenek.

A megfelelő tolerancia helyes meghatározása hasonló a többi hyper paraméter (C , ε , kernel paraméte-

rek) megválasztási feladathoz. Egy lehetséges megoldás a kereszt kiértékelés (cross-validation) alkalmazása. Minél nagyobb a tolerancia, annál kisebb a hálózat, ellenben az approximáció hibája nő. Az ε' helyes megválasztása egy döntési feladat ahol a pontos approximáció illetve a modell komplexitása között kell dönten.

Fontos hangsúlyozni, hogy 0 tolerancia esetén az LS^2 -SVM és az LS-SVM megoldás azonos, mivel ebben az esetben a kiválasztási módszer a mátrix minden oszlopát, azaz minden tanító mintát megtartja.

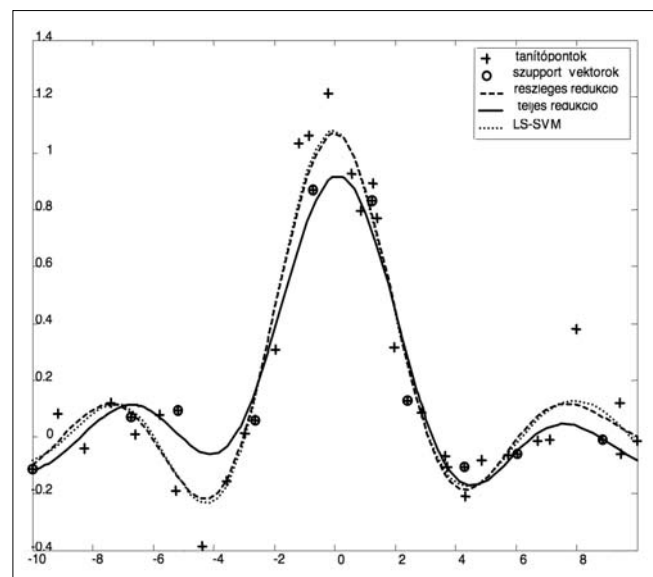
4. Vizsgálatok

A bemutatott algoritmusok értékeléséhez a szupport vektor regresszióval kapcsolatos irodalmakban legelterjedtebben használt benchmark feladatot a $\text{sinc}(\mathbf{x})$ függvényt használjuk a $[-10, 10]$ tartományban. Kernel függvénynek a Gauss kernel kerül alkalmazásra $\sigma = \pi$ paraméterrel. A fent bemutatott ε' tolerancia értéke 0.2, valamint $C = 100$. A bementi mintákat 0.01 szórású normál eloszlású zaj terheli.

Elsőként a részleges redukció hatását kerül bemutatásra. Később ugyanezen problémában az automatikus szupport vektor kiválasztó algoritmus is felhasználásra kerül. Ugyanez a probléma kerül felhasználásra a hagyományos módszerekkel (LS-SVM, és Pruned LS-SVM) való összehasonlításához is, de egy összetettebb idősor előrejelzési feladat is bemutatásra kerül (Mackey-Glass kaotikus idősor előrejelzése).

A redukciós módszerek összehasonlításához először egy nagyon egyszerű szupport vektor kiválasztási módszert használunk. A 40 véletlenszerűen generált bemeneti mintapont közül minden negyediket szupport vektornak választunk. A 3. ábrán a teljes és a részleges redukció eredménye együtt látható az eredeti, teljes LS-SVM-el.

3. ábra
A különböző redukciós módszerek eredménye azonos szupport vektor készlet esetén



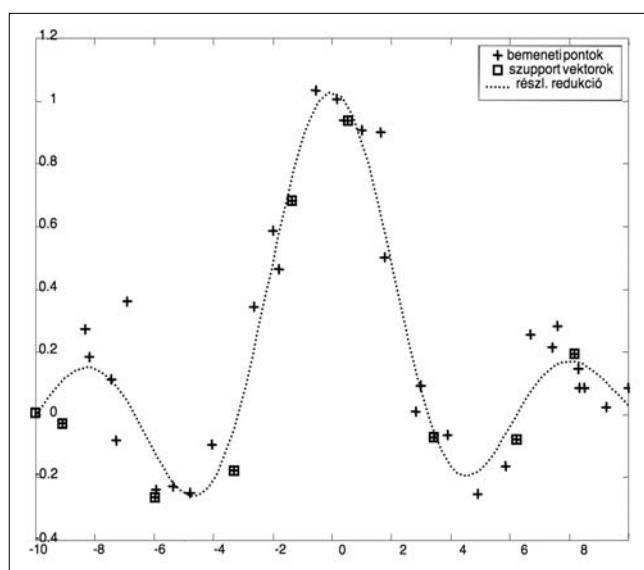
Látható, hogy a részleges redukció eredménye majdnem megegyezik az eredeti LS-SVM kimenetével, de a megoldás komplexitása a negyedére csökkent. Mivel a teljes redukció esetén az eredményt csak a szupport vektorok befolyásolják, ebben az esetben jóval rosszabb approximációt kapunk:

$$\text{MSE}_{\text{partial red.}}: 1.04 \times 10^{-3},$$

$$\text{MSE}_{\text{full red.}}: 6.43 \times 10^{-3},$$

$$\text{MSE}_{\text{LS-SVM}}: 1.44 \times 10^{-3}.$$

A bemutatott kiválasztási eljárással a szupport vektorok automatikusan is meghatározhatók. A 4. ábra egy ilyen, a tanulás során automatikusan meghatározott szupport vektor készleten alapuló megoldást mutat be.



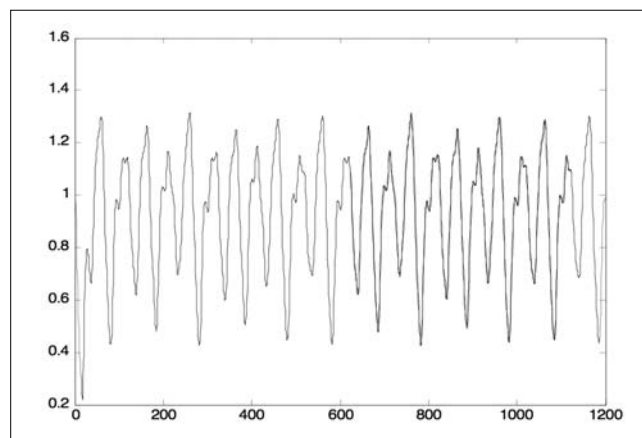
4. ábra

Egy részlegesen redukált LS-SVM, ahol a szupport vektorok a fent megadott kiválasztási algoritmussal kerültek meghatározásra ($\epsilon' = 0.2$)

Az ábrán látható, hogy míg az eredeti hálózat 40 neuront tartalmazna, a redukált hálózat csak 9-et. A kiválasztási módszer még fontosabb tulajdonsága, hogy a meghatározott szupport vektor halmaz számossága független a tanító pontok számától. Ha a feladat N rejtett rétegbeli neuronnal megoldható, akkor a felhasznált tanító pontok számával a hálózat mérete nem kell hogy változzon.

A táblázat (5. ábra) ugyanazon feladat ($\text{sinc}(x)$) különböző méretű tanítópont készletei mellett kapott szupport vektorok számát, valamint az approximáció hibáját mutatja be. Az approximáció hibája egy 100 zajmentes mintából álló ellenőrző készleten számított négyzetes

hiba. Látható, hogy a mintapontok számának növekedésével a hiba – a vártnak megfelelően – csökken, míg a hálózat mérete, azaz a megoldás komplexitása gyakorlatilag változatlan.



6. ábra

A Mackey-Glass idősor előrejelzési feladat megoldása LS-SVM-el

A 6. ábra a széleskörűen elterjedt Mackey-Glass idősor előrejelzési probléma megoldását mutatja be. A Mackey-Glass idősor előrejelzésében az $[-6, -12, -18, -24]$ késleltetésekkel használtuk fel, azaz a t . időpillanat $x(t)$ értékét négy korábbi érték alapján becsüljük (a Mackey-Glass folyamatban nincs külső bemenet, azaz a kimenet csak a korábbi értékektől függ). Az LS-SVM tanításhoz az idősor első feléből 500 mintát használtunk fel.

Az előrejelzés ellenőrzése során a következő értéket mindig a korábbi becsült idősor értékek alapján számítottuk. Egy másik megoldás, ha a későbbi becslések során a helyes korábbi értékeket használjuk fel. Az előbbi esetben a becslések hibája folyamatosan egyre nagyobb hibákat eredményez. Az utóbbi ellenőrzést akkor célszerű alkalmazni, ha a becslés során mindig csak a következő értéket kell meghatározni.

A fenti benchmark feladat alapján látható, hogy a bemutatott LS-SVM regresszió alapuló megoldás kiválóan alkalmas egy idősor előrejelzési probléma megoldására.

5. Összefoglalás

Cikkünkben a Least-Squares szupport vektor gépek adatbányászati alkalmazásainak lehetőségeit vizsgáltuk. Bemutattuk a hagyományos LS-SVM-et, majd javaslatot tettünk egy olyan kiterjesztésre, ami lehetővé teszi,

5. ábra
A javasolt módszerrel elért szupport vektorok száma és a négyzetes hiba ugyanazon feladatra különböző tanítóhalmaz készletekkel (toleranciaérték: 0.25)

Tanító minták száma (A hagyományos LS-SVM esetén ez megegyezik a neuronok számával)	A "szupport vektorok" száma (A neuronok száma a javasolt módszerrel)	Az approximáció négyzetes hibája (A javasolt módszer alkalmazásával elért hiba)
40	8	1.890×10^{-3}
80	9	0.877×10^{-3}
800	9	0.155×10^{-3}
1600	9	0.029×10^{-3}

hogy nagy mennyiségű adat esetén is a rendszer egy egyszerű modelljét alkothassunk meg. Tartalmazza, hogy egy előrejelzési feladat általános esetben hogyan vezethető vissza egy függvény regressziós problémára, melynek megoldására egy tisztán analitikus és egy neurális modell is bemutatásra került.

Ezek a módszerek és eljárások felhasználhatók különböző orvosi adatok elemzésére is, s így jól használható eszközt jelentenek, mint diagnosztikai, mind pedig egyéb kutatási célokhoz.

Irodalom

- [1] Horváth G. (szerk.):
„Neurális hálózatok és műszaki alkalmazásai”,
Műegyetem kiadó, Budapest, 1998.
- [2] S. Haykin:
„Neural networks. A comprehensive foundation”,
Prentice Hall, N. J., 1999.
- [3] V. Vapnik:
„The Nature of Statistical Learning Theory”,
New York, Springer-Verlag, 1995.
- [4] E. Osuna, R. Freund, F. Girosi:
„Support vector machines: Training and applications”,
Technical Report AIM-1602, MIT A.I. Lab., 1996.
- [5] C. J. C. Burges, B. Schölkopf:
„Improving the accuracy and speed of
support vector learning machines”
In M. Mozer, M. Jordan and T. Petsche, editors,
Advances in Neural Information Processing Systems 9,
pp.375–381, Cambridge, MA, MIT Press, 1997.
- [6] E. Osuna, R. Freund, F. Girosi:
„An improved training algorithm for
support vector machines”
In J. Principe, L. Gile, N. Morgan and E. Wilson, editors,
Neural Networks for Signal Processing VII –
Proceedings of the 1997 IEEE Workshop,
pp.276–285., New York, IEEE, 1997.
- [7] Thorsten Joachims:
„Making Large-Scale SVM Learning Practical”,
Advances in Kernel Methods-SV Learning’,
MIT Press, Cambridge, USA, 1998.
- [8] J.A.K. Suykens, T. Van Gestel, J. De Brabanter,
B. De Moor, J. Vandewalle:
„Least Squares Support Vector Machines”,
World Scientific, Singapore, 2002.
- [9] F. Girosi:
„An equivalence between sparse approximation
and support vector machines,”
Neural Computation, 10(6), pp.1455–1480., 1998.
- [10] J. Vallyon, G. Horváth,
„A generalized LS-SVM”,
SYSID’2003 Rotterdam, 2003, pp.827–832.
- [11] J. Vallyon, G. Horváth,
„A Sparse Least Squares Support
Vector Machine Classifier”,
Proceedings of the International Joint Conference
on Neural Networks IJCNN 2004, pp.543–548.
- [12] Yuh-Jye Lee, Olvi L. Mangasarian:
„RSVM: Reduced support vector machines”,
Proc. of the First SIAM International Conference
on Data Mining, Chicago, April 5-7, 2001.
- [13] W. H. Press, S. A. Teukolsky,
W. T. Wetterling, B. P. Flannery:
„Numerical Recipes in C”,
Cambridge University Press, Books On-Line, 2002.
www.nr.com
- [14] G. H. Golub, Charles F. Van Loan:
„Matrix Computations”,
Gene Johns Hopkins University Press, 1989.

Hírek

A **Sun Microsystems** a nyár folyamán új konszolidált tárolómegoldást alakított ki az OTP Banknál, amellyel a pénzügyi intézet több üzleti területéhez tartozó rendszert képes kiszolgálni. A heterogén rendszerek ellátására alkalmas, folyamatos rendelkezésre állást biztosító berendezéssel egyszerűsíthető és javítható az adatok felügyelete és védelme, valamint általa mérsékelni lehet az adatközpontok költségeit és bonyolultságát. Ezen megoldás a konszolidált rendszerek számára dinamikus háttérkapacitás-bővítést biztosít a jövőben, továbbá új rendszerek háttértár-konszolidációjának alapjait teremti meg. Az OTP Bank és a Sun folyamatos együttműködésének köszönhetően a rendszerek migrációja zökkenőmentes volt: az átállást minimális leállással, alacsony kockázati szinten sikerült megoldani.

A most bevezetett, jelentős bővíthetőségi tartalékkal rendelkező, nagy megbízhatóságú és magas rendelkezésre állást nyújtó – 16 terabájnyi adat megővására képes – tárolórendszer amellel, hogy a korábbi megoldásokhoz képest lényegesen magasabb szolgáltatási szintet biztosít, kedvező feltételekkel került értékesítésre. A színvonal emeléséhez többek között a nagyobb fokú adathozzáférési lehetőség és adatbiztonság, az emelt szintű teljesítmény, valamint több intelligens tárolóoldali szolgáltatás léte – például a pillanatfelvételi lehetőség, a távoli adatreplikáció, vagy az újgenerációs menedzsment framework – is hozzájárul. A kiépített rendszer további előnye még, hogy a pénzügyi intézetnél használt rendszerek teljes – a Sun által támogatott, StorEdge-t, tárolóhálózatot (SAN-t) és szervereket tartalmazó – infrastruktúrája így egy forrásból kap támogatást.

Statisztikai adat- és szövegelemzés Bayes-hálókkal: a valószínűségektől a függetlenségi és oksági viszonyokig

MILLINGHOFFER ANDRÁS, HULLÁM GÁBOR, ANTAL PÉTER

Budapesti Műszaki és Gazdaságtudományi Egyetem, Méréstechnika és Információs Rendszerek Tanszék
peter.antal@mit.bme.hu

Kulcsszavak: Bayes-statisztika, Bayes-hálók, tanulás, alkalmazási területek

Egy sokváltozós, akár több száz bizonytalan eseményt is tartalmazó tárgyterület szakértői háttértudáson, szakcikkeken és statisztikai adatokon alapuló valószínűségi modellezése több szintre és fázisra tagolódó feladat. Egyrészt tartalmazza a tárgyterület numerikus eloszlásának, a függetlenségi és az okozati relációknak, mint egymásra épülő szinteknek a modellezését. Másrészt felöleli a priori ismeretek szakértőtől, tudásbázisokból, a szemantikus webről és szabadszöveges forrásokból történő kinyerését és formalizálását, majd statisztikai adatokkal való kombinálását és egy döntéseméleti keretben való használatát, azaz a tudásmérnökség, a gépi tanulás és következtetés területét is. A cikkünkben a Bayes-háló modellosztályt (reprezentációt) mutatjuk be, amellyel ezek a feladatok sikerrel oldhatók meg. Ismertetjük a Bayes-statisztika keretrendszerét, amely a Bayes-hálók alkalmazásának nem szükségszerű, de gyakori környezete. A módszer gyakorlati alkalmazását az általunk kifejlesztett rendszer egy orvosi biológiai feladatra, a petefészekrák tárgyterületre történő alkalmazásán keresztül illusztráljuk, illetve áttekintjük a jelenleg létező ipari alkalmazásokat. Végül kitérünk az ismertetett modell gyengeségeire és vázoljuk az ezeket kiküszöbölni kívánó kutatási irányokat.

1. Bevezetés

A Bayes-háló alapú alkalmazások térhódítása a 90-es években kezdődött el [17], kezdetben főleg orvosi diagnosztikai és előrejelző rendszereknél. Az elmúlt néhány évben a felhasználási kör olyan változatos területekkel bővült, mint pénzügyi, telekommunikációs vagy hadszíntéri döntéstámogatás és hírszerzői információk integrálása.

A felhasználói viselkedéshez kapcsolódó alkalmazások két fő irányvonal mentén fejlődtek, egyrészt a személyre szabott információszolgáltatás terén, mint például a felhasználót segítő sugórendszerek [18], valamint az információs rendszerek biztonsága területén [32], ahol a rosszindulatú felhasználók kiszűrése a cél a viselkedésmintázatok vizsgálta alapján. Ehhez hasonló osztályozási feladat a spam levelek kiszűrése, melyre számos Bayes-hálón alapuló megoldás született [28].

Komplex rendszerek működtetésénél, legyen az mozgástervezés [25] vagy nyomtatórendszer karbantartás [29], ahol a diagnosztika a bonyolult felépítés és bizonytalanság miatt egyszerű szabályalapú módszerekkel nem követhető, szintén hatékonyan alkalmazható a bayesi megközelítés. Mindemellett különböző döntéstámogatási rendszerekben [30] és ezen belül, a kockázat-előrejelzés [26] terén is jelentős pozíciót töltenek be a Bayes-háló alapú alkalmazások. Egyes területeken, mint a Bayes-háló alapú adatbányászatonál [23] vagy az említett kockázat-előrejelzésnél a Bayes-háló tanulását maga a felhasználó irányíthatja.

A következőkben áttekintjük a Bayes-háló modellosztályt, annak egy gyakori alkalmazási környezetét, a Bayes-statisztikát és bemutatunk egy orvosi biológiai al-

kalmazást az integrált adat és szövegbányászat területén. A 2. fejezet áttekintést ad a Bayes-statisztikáról, majd ismertetjük a Bayes-háló reprezentációt és annak kézi konstruálásához, tanulásához és az azzal történő következtetéshez kapcsolódó fogalmakat, algoritmusokat és metodológiákat. A 4. fejezet alkalmazási területeket mutat be, illetve ismerteti kutatásainkat, végül pedig a Bayes-hálók továbbfejlesztési irányaira adunk kitékintést.

2. Bayes-statisztikai módszerek

2.1. A valószínűség bayesi értelmezése

A cikkben vizsgált Bayes-statisztika és a Bayes-háló modellosztály közös alapvető célja, hogy a bizonytalan háttértudáson és megfigyeléseken alapuló következtetések számára axiomatikus alapot és gyakorlati alkalmazhatóságot biztosítson. A fellépő bizonytalanságnak számos oka lehet, például a tudás kinyerése során alkalmazott módszer, az adatgyűjtési eljárás, vagy a tudás hiánya, esetleg figyelmen kívül hagyása.

A Bayes-statisztikai módszertan a bizonytalanság kezelésére a valószínűségi keretrendszert alkalmazza, a valószínűség szubjektivistá interpretációját elfogadva, szemben a mérnöki gyakorlatban elterjedtebb, a relatív gyakoriságok határértékein alapuló, úgynevezett frekventista értelmezéstől. A szubjektivistá értelmezésben a valószínűségeket az események bekövetkezésében való, adott kontextushoz tartozó a priori hiedelemnek, elvárásnak, egyfajta meggyőződési mértéknek tekintjük.

Az axiomatikus származtatásnál megmutatható, hogy egy döntési problémában minden eseményhez rendel-

hető egy pozitív valós szám, mely az adott esemény valószínűségeként értelmezhető és egy hasznossági érték, melyekkel a preferenciák egzaktul reprezentálhatók és racionális döntések hozhatók. (Egy döntési probléma egy $(E, C, A, <)$ négyessel definiálható, ahol 'E' az események, 'C' a következmények, 'A' a lehetséges cselekvések halmaza, '<' pedig az 'A' elemei feletti preferenciáinkat tükröző rendezés [6].)

2.2. A bayesi modell

A bayesi módszertan további axiomatikus alapját a reprezentációs tételek [6] jelentik, ezek megmutatják, hogy egy végtelen felcserélhetőséget teljesítő eloszlás (azaz amelyben bármely π permutációra $p(x_1, \dots, x_n) = p(x_{\pi(1)}, \dots, x_{\pi(n)})$), reprezentálható egy alkalmas adatgenerálási parametrikus modellosztállyal és egy e feletti eloszlással.

A valószínűség szubjektivista értelmezésére és a fenti tulajdonságú modellosztályok léteire alapozva javasolható a Bayes-statisztikai keretrendszer, amelyben a megfigyelési adatokat valószínűségi változók által paraméterezett modellegyüttesekből származtatjuk, azaz a megfigyelések és a modelparaméterek ugyanolyan modellezési szinten helyezkednek el.

Gyakorlati megközelítésekben az alkalmazott modellosztály paraméterezését hierarchikusan tagolják, leggyakrabban a következő módon, amit a cikkben is követünk: egyrészt a modelltér diszkrét elemek (a lehetséges *modellstruktúrák*) halmaza, másrészt hozzájuk numerikus paraméterek tartoznak.

2.3. Következtetés

A következtetés során a feladat, hogy megbecsüljük egy adott esemény, vagy egy modell feltételes valószínűségét az alapismereteink és a megfigyelési adatok szerint. Az első esetben *prediktív*, a másodikban *parametrikus* becslésről (a posteriori eloszlás számításáról) beszélünk. Mindkét esetben a Bayes-tételből indulunk ki, amelynek segítségével események feltételes valószínűségét számíthatjuk (a továbbiakban D az adatokat, G egy struktúrát, θ pedig egy paraméterezést jelöl):

$$P(G, \theta | D) = \frac{P(D | G, \theta)P(G, \theta)}{P(D)} \quad (1)$$

Az a posteriori eloszlást (röviden *posterior*) az előzőleg említett modelltér egy struktúrájának paraméterezésére, vagy magára a struktúrák terére is kiszámíthatjuk. A paraméterek esetén a képlet formailag megegyezik (1)-gyel, a struktúrákra vonatkozó pedig a paraméterek kiintegrálása után adódik:

$$P(G | D) = \frac{\int P(D | G, \theta)P(G, \theta)d\theta}{P(D)} \quad (2)$$

Predikció esetén még egy lépést teszünk: a keresett valószínűséget kiszámítjuk minden létező modellre, és ezeknek a modellek a posteriori valószínűségével súlyozott átlagát vesszük:

$$p(x | D) = \sum_k p(G_k | D) \int p(x | \theta_k) p(\theta_k | G_k, D) d\theta_k \quad (3)$$

2.4. Monte Carlo módszerek

A fenti posteriorok gyakran nem mintavételezhetők, ezért Monte Carlo módszereket kell alkalmaznunk, például a fontossági mintavételezést vagy a Markov-lánc Monte Carlo (MCMC) módszerek egyikét [14]. A posterior vizsgálata helyett a feladat gyakran egy vagy több $\bar{f} = E_{p(x|D)}[f(x)]$ alakú várható érték becslésére egyszerűsödik. Ez megtehető a következő lépésekkel, melyek helyességét az MCMC módszereknél igazolt nagy számok törvénye (2. pont) és centrális határeloszlás tétel (3. pont) biztosítja [14]:

1. a $\{x_i\}_{i=1}^N$ minta vételezése
2. $\bar{f}$ becslése az $\hat{f} = \frac{1}{N} \sum_{i=1}^N f(x_i)$ képlet alapján
3. konfidenciabecslés az $|\bar{f} - \hat{f}|$ eltérésre

A Monte Carlo mintavételezés mellett még gyakori a meghatározott számú, legnagyobb a posteriori valószínűségű modellstruktúra alapján történő kiszámítás, mely legegyszerűbb esetben egyetlen, úgynevezett MAP (maximum a posteriori) modell használatát jelenti. A legnagyobb valószínűségű modell(ek) meghatározását a tanulásról szóló fejezet tárgyalja.

2.5. Bayes-statisztikai megközelítés előnyei

A következő listában röviden összefoglaljuk a fentebb ismertetett bayesi módszertannak a klasszikus statisztikával szembeni előnyeit [27]:

- A paraméterek bizonytalanságát a felettük definiált eloszlással jellemezzük, így minden statisztikai következtetés egy direkt valószínűségi állítás, ami az automatizált többlépéses tanulási rendszereknél és tudásbázisok generálásánál igen előnyös.
- A paraméterbecslés egy inverziós feladatként fogható fel, hisz itt kizárólag az adatból következtetünk arra a paraméterre, amely annak generálását meghatározta. A Bayes-tétel pontosan ezt az inverziót formalizálja, így a következtetést a hipotetikus viselkedés figyelmen kívül hagyásával végzi, szemben a klasszikus statisztika egyes módszereivel.

- Az a priori eloszlások (röviden priorok) használata alkalmas az előismeretek összegzésére vagy akár a teljes ismerethiány kifejezésére is.

- A priorok – mivel leggyakrabban korábbi megfigyeléseken vagy vizsgálatokon alapulnak – az ismeretszerzési folyamat egyes fázisainak tekinthetők, hisz új tudásunkat (az a posteriori eloszlást) ez alapján szerezzük.

- A bayesi következtetés a Bayes-tétel segítségével egyenrangú módon, normatívan kombinálja az előismeretekben és az adatokban rejlő információkat. Így a Bayes-tétel használata az adatok és előismeretek egyfajta súlyozását valósítja meg: az adatok mennyiségének növekedtével azok befolyása is nő a posteriori eloszlásra.

- Az a posteriori eloszlás használata pontbecslés helyett a predikció során nem csak a legvalószínűbb konfiguráció alapján számol, hanem figyelembe veszi a kevésbé valószínű eseteket is, ami a modell komplexitásához képest kis mennyiségű megfigyelés esetén fontos.

3. Bayes-hálók

A valószínűségi megközelítésben bizonytalan tudásunkat sztochasztikus változók együttes eloszlásával reprezentáljuk. A szisztematikus struktúrával nem rendelkező tárgyterületek esetén (szemben például a kép- és hangfeldolgozással) az ilyen eloszlások modellezésére használt elsődleges eszközt ma a Bayes-hálók jelentik. Ezekben egy irányított körmentes gráfban (DAG – directed acyclic graph) reprezentálják a változókat és a köztük lévő összefüggéseket: minden csomópont egy-egy változót jelöl, és minden csomóponthoz tartozik egy lokális feltételes valószínűségi modell, amely leírja a változó függését a szüleitől (a pontos definíciót a következő fejezet tartalmazza).

Mint reprezentációs eszköz, egy Bayes-háló háromféle értelmezést kaphat, ezek a felsorolás sorrendjében egyre erősebb modellezési, értelmezési lehetőséggel bírnak:

- Tekinthető egyszerűen az együttes eloszlás egy hatékony ábrázolásának, hisz a csomópontonkénti feltételes valószínűségi modellekre való faktorizálással a felhasznált paraméterek száma jelentősen csökken.
- Egy adott struktúra meghatározza, hogy az ábrázolt eloszlásban milyen feltételes függések és függetlenségek lehetnek, azaz az élek tekinthetők a közvetlen valószínűségi összefüggések reprezentációjának, míg a teljes gráf a reprezentált eloszlás függési térképének.
- Az előzőnél is erősebb a kauzális értelmezés, melyben minden élt az érintett két csomópont közötti ok-okozati összefüggésként értelmezzük.

3.1. A valószínűségi definíció: szintaxis és szemantika

Egy Bayes-háló struktúrája és a reprezentálni kívánt eloszlás közti kapcsolatot az alábbi négy feltételre alapozhatjuk, melyekről belátható [9], hogy ekvivalensek.

- A $P(X_1, \dots, X_n)$ eloszlás *faktorizálható* a G DAG szerint, ha:

$$P(X_1, \dots, X_n) = \prod P(X_i | Pa(X_i)),$$

ahol $Pa(X_i)$ az X_i csomópont szülői halmaza.

- A $P(X_1, \dots, X_n)$ eloszlásra teljesül a *sorrendi Markov-feltétel* G szerint, ha

$$\forall i = 1..n : I(X_{\pi(i)} | Pa(X_{\pi(i)}) \setminus \{X_{\pi(j)} \mid j < i\} \setminus Pa(X_{\pi(i)}))_P$$

ahol az $I(X|Y|Z)$ reláció az X feltételes függetlenségét jelenti a Z -től Y feltétellel, π pedig a struktúra egy topologikus rendezése.

- A $P(X_1, \dots, X_n)$ eloszlásra teljesül a *lokális (szülői) Markov-feltétel* G szerint, ha bármely változó független nem-leszármazottaitól, feltéve szüleit.
- A $P(X_1, \dots, X_n)$ eloszlásra teljesül a *globális Markov-feltétel* G szerint, ha

$$\forall x, y, z \subseteq \{X_i\} : I(x | z | y)_G \Rightarrow I(x | z | y)_P,$$

vagyis, ha z d-szeparálja x -et y -től a G gráfban, akkor x független y -től, feltéve z -t.

Egy elfogadott definíció a Markov-feltételek által adott függőségi rendszer tulajdonságaira épít [24]: A ' G ' *irányított körmentes gráf* a ' $P(U)$ ' *eloszlás Bayes-hálója* (U az összes változó halmaza), akkor és csak akkor, ha minden $u \in U$ változót a gráf egy csomópontja reprezentál, a gráfra teljesül valamelyik (és így az összes) Markov-feltétel, és a gráf minimális (azaz bármely él elhagyásával a Markov-feltétel már nem teljesülne).

Míg ez a definíció egyértelműen a valószínűségi függetlenségek rendszerének reprezentációjaként tekint a Bayes-hálóra, addig a mérnöki gyakorlatban közkedvelt az alábbi, praktikus meghatározás:

Az ' U ' *valószínűségiváltozó-halmaz Bayes-hálója* a (G, θ) *páros*, ha ' G ' *irányított körmentes gráf*, amelyben a csomópontok jelképezik U elemeit, θ pedig csomópontokhoz tartozó ' $P(X|Pa(X))$ ' *feltételes eloszlásokat leíró numerikus paraméterek összessége*.

Fontos megjegyezni, hogy a definiált modellosztályban a lehetséges struktúrák száma a csomópontok számában szuperexponenciális, ez pedig például a később tárgyalandó tanulás komplexitását is befolyásolja.

Bár egy Bayes-háló egyaránt tartalmazhat diszkrét és folytonos változókat is, mi a továbbiakban kizárólag diszkrét, véges változókkal foglalkozunk, feltéve továbbá, hogy minden lokális feltételes valószínűségi modell a multinomiális eloszlásokhoz tartozik, így a paraméterek úgynevezett feltételes valószínűségi táblák (FVT-k) elemei.

Egy adott Bayes-háló struktúrája meghatározza, hogy az milyen függéseket írhat le (például külön komponensekben lévő változók közt nem lehet függés), azonban különböző struktúrákhoz is tartozhat azonos implikált függési rendszer. Ha két struktúrából ugyanazok a feltételes függetlenségek olvashatók ki, a két gráfot *megfigyelés-ekvivalensnek* mondjuk. Belátható [24] hogy két gráf akkor és csak akkor megfigyelés-ekvivalens, ha irányítás nélküli vázuk, illetve v -struktúráik (az $A \rightarrow B \leftarrow C$ típusú részgráfok úgy, hogy A és C közt nincs él) megegyeznek.

A megfigyelési ekvivalencia segítségével a struktúrákat diszjunkt osztályokba sorolhatjuk. Minden ilyen ekvivalencia osztályt egy úgynevezett esszenciális PDAG² gráffal reprezentálhatunk. Az esszenciális gráf váza megegyezik az osztályba tartozó gráfokéval, és csak azok az élei irányítottak, amelyek iránya mindegyik gráfban megegyezik (ún. kényszerített – compelled – élek).

3.2. Kauzális definíció

Az előző, tisztán valószínűségi definíciók bevezetése után formálisan könnyen áttérhetünk a Bayes-hálók kauzális értelmezésére: egy (G, θ) *páros kauzális Bayes-hálója* a $P(U)$ *eloszlásnak*, ha egyrészt a tárgyterület valószínűségi modellje az előző értelmezések szerint, továbbá minden él közvetlen ok-okozati viszonyt jelképez.

¹ ' z ' d-szeparálja ' x '-et és ' y '-t a ' G ' gráfban ($x, y, z \subseteq V(G)$), ha minden ' x ' és ' y ' között menő irányítatlan ' p ' utat blokkol, azaz, ha (1) ' p ' tartalmazza ' z ' egy elemét nem összefutó élekkel, vagy (2) ' p ' tartalmaz egy ' n ' csomópontot összefutó élekkel, hogy ' z ' nem tartalmazza sem ' n '-t, sem valamelyik leszármazottját.

² Egy PDAG (partially directed acyclic graph) gráf vegyesen tartalmaz irányított és irányítatlan éleket.

Hasonlóan, itt is létezik egy Markov-feltétel: egy $P(U)$ eloszlás és egy kauzális relációt leíró G gráf teljesíti a kauzális Markov-feltételt, ha G és $P(U)$ teljesíti a lokális Markov-feltételt.

A Markov-feltétel teljesülése biztosítja, hogy minden (kauzális) függés kiolvasható a gráfból, a másik irányhoz, ahhoz tehát, hogy minden a gráfból kiolvasott függés teljesüljön az eloszlásban, annak stabilnak kell lennie. Egy $P(U)$ eloszlás stabil, ha létezik olyan G gráf, hogy $P(U)$ -ban pontosan a G -ből d -szeparációval kiolvasható függések és függetlenségek teljesülnek benne (például megfelelő paraméterezés mellett előfordulhat, hogy egy $A \rightarrow B \rightarrow C$ struktúrában A és C függetlenek).

A fenti kauzális definíció a modell és a tárgyterület összefüggéseinek értelmezését illetően igen erős, a megfigyelési adatok statisztika elemzésének kereteit meghaladó eszközt szolgáltat. Alkalmazásakor figyelembe kell vennünk, milyen nem kauzális kapcsolatok okozhatnak valószínűségi összefüggést két változó között, azaz milyen korlátai vannak a kauzális értelmezésnek.

Ilyenek lehetnek például:

- Zavaró változók: a két változó közti függést okozhatja egy közös ősz (úgynevezett zavaró változó) is.
- Kiválasztási bias: a változók közti függés lehet az adatgyűjtési mód következménye is (például ha egy orvosi adatbázisba csak a komolyabb megfázással kezelt betegek kerülnek be, akkor a láz és torokfájás között direkt függést figyelhetünk meg).
- Az ősz-ok, leszármazott-okozat megfeleltetés és a DAG gráfstruktúra kizárja a mechanizmusokban lévő visszacsatolások (ciklikusságok), illetve az oda-vissza ható okozatiság lehetőségét.
- A modellter maga (azaz, hogy milyen változók szerepelnek, illetve azok milyen értékkel rendelkezik) szintén befolyásolja, hogy milyen direkt függések jelennek meg (azaz a gráf struktúrát).

3.3. Bayes-hálók és a tudásmérnökség

A fentebb definiált Bayes-háló a tudásmérnökség eszközeként jelent meg a 80-as években, konstruálása jellemzően a szakértőktől származó adatokból történt manuálisan. A kézi konstruálás még napjainkban is jelentős súlyt képvisel a Bayes-hálók alkalmazásában, másrészt ahol az adathoz viszonyítva jelentős a priori tudás áll rendelkezésre, ott a Bayes-hálók tudásmérnöki alkalmazása a bayesi keretrendszer alkalmazásának egy kezdeti fázisát jeleni, nevezetesen a prior konstruálást.

A tudásmérnökség metodikájára nagy hatással volt a nagy mennyiségű elektronikus tárgyterületi információ megjelenése, a megfelelő mennyiségű statisztikai adat elérhetősége, valamint a Bayes-statisztikai alapú gépi tanulási módszerek elterjedése.

A felépített tudásbázissal szemben követelményként jelent meg a bayesi módszerek alkalmazásakor, hogy támogassa a priorok konstruálását, hiszen a valószínűségekkkel leírt a priori tudás és a rendelkezésre álló adatok bayesi frissítéssel történő kombinációja szolgáltatja a végső tudásmodellt. Mindemellett fontos, hogy

a tudásbázis segítse komplex, akár szabad szöveges háttérismereteket is tartalmazó valószínűségi állítások megfogalmazását, valamint tegye lehetővé a szakértőktől származó szubjektív információ tárolását, mely releváns a bayesi, a priori tudásmodell megalkotásánál.

Egy tudásbázis megépítéséhez olyan környezetben, ahol rendelkezésre áll elektronikus tárgyterületi tudás, elegendő statisztikai adat, valamint a megfelelő bayesi módszerek, az alábbi lépések szükségesek (amelyekből a specifikusokat részletezzük):

1) *Célok, alkalmazási terület és modellezési szintek identifikációja*

Terminológia és ontológia elfogadása.

2) *Nem rendszerezett tudás begyűjtése*

Ehhez a lépéshez tartozik az összes releváns elektronikus és egyéb szövegalapú információforrás feldolgozása, ami magába foglalja az a priori információ kinyerését különféle szövegbányászati módszerek alkalmazásával, mint például az általunk kifejlesztett módszer, amit a későbbiekben mutatunk be.

3) *Struktúra kinyerése*

A G DAG struktúrák feletti $p(G)$ priorok konstruálása, melyek egyesítik a szakértők által megadott információkat az elektronikus forrásokból kinyert információkkal. A $p(G)$ a priori eloszlást többnyire normalizálatlan formában lehet előállítani: például egy adott referencia struktúrától való eltérés alapján

$$P(G|\xi) \propto \kappa^\delta; 0 < \kappa < 1,$$

ahol δ a referenciától való tetszőlegesen definiált strukturális tulajdonságokbeli eltéréseknek a száma.

4) *Paraméter és hiperparaméter kinyerése*

A valószínűségi paraméterek számos módon nyerhetők: adatbázisok, szakirodalom vagy szakértők szubjektív véleménye alapján. A $p(\theta|G)$ paraméter prior specifikációja az általunk vizsgált diszkrét, véges esetben egy egyszerű módszerrel megtehető, ha feltehetjük az egyes változókhoz és szülői értékkonfigurációkhoz tartozó paraméterek függetlenségét:

$$P(\theta|G_0, \xi) \propto \prod_{i=1 \dots n} \prod_{j=1 \dots q_i} P(\theta_{i,j} | G_0, \xi).$$

Egy szinte kizárólagosan használt eloszláscsalád az adott változó, adott szülői értékkonfigurációjához tartozó $P(\theta_{i,j} | G_0, \xi)$ megadására a Dirichlet eloszlás $\text{Dir}(\theta_{i,j} | \alpha_{i,j}, \xi)$, ahol az $\alpha_{i,j}$ hiperparaméter jelentése a paraméterhez tartozó szülői értékkonfiguráció korábban megfigyelt eseteinek számait jelenti [9].

Megmutatható, hogy a Dirichlet család az egyetlen lehetséges választás, ha az ugyanazon megfigyelési ekvivalencia osztályba tartozó G struktúrákhoz ekvivalens priorokat szeretnénk megadni, ami kauzális modellezésnél nem szükségszerű [16].

További feltevések mellett az is bizonyítható, hogy az összes struktúrához konzisztens $p(\theta|G)$ definíciója ekvivalens egy teljes modellhez tartozó pontparametrizációnak és egyetlen korábban megfigyelt összesetszámmal jelentő hiperparaméternek a megadásával. E kettő együtt valójában egy a priori adathal-

mazt definiál, ami korábban megfigyelt eseteket tartalmazza, így az összesetszámot virtuális vagy a priori mintaszámnak nevezünk.

5) Érzékenységi analízis, verifikáció és validáció

A modellek posteriorjának vizsgálata magába foglalja egyrészt az a priori eloszlásokra való érzékenység vizsgálatát (ami különösen fontos a több szakértőt és tudásbázist is magában foglaló automatizáltan származtatott prioroknál), másrészt referencia priorokkal való összehasonlítást. Mindkét esetben gyakran szükséges a modellosztály komplexitása miatt, egyrészt hogy modell jegyeket használjunk, másrészt hogy MAP modellre alapozzuk a vizsgálatot.

Mint ahogy az látható, a tudásbázis építése a bayesi modellkiértékeléssel zárul. A kiértékelés tartalmazza az adat és a modell kompatibilitásának vizsgálatát és az a posteriori valószínűségek vizsgálatát, azaz a tudásmérnöki folyamat lényege az a priori modell konstruálása a későbbi tanulási folyamat számára.

3.4. Következtetés Bayes-hálókbán

Egy konkrét Bayes-hálóban való következtetés alapfeladata a $P(X = x|Y = y, G, \theta)$ mennyiség kiszámítása, azaz adott egy struktúra és paraméterezése valamint ismert a bizonyítékváltozók (Y) behelyettesítése, kérdés a lekérdezőváltozók (X) egy adott konfigurációjának valószínűsége.

Könnyen belátható [15], hogy a feladat NP-teljes (hiszen például visszavezethető a kielégíthetőségi problémára), számításigénye a csomópontok számában exponenciális. Ezért a gyakorlatban vagy szimuláción alapuló, közelítő eredményt adó Monte Carlo módszereket [14], vagy a gráfot másodlagos struktúrába transzformáló úgynevezett junction-tree algoritmusokat [19] alkalmaznak.

Hogy $P(X = x|Y = y)$ a mennyiséget kiszámíthassuk, azaz valódi bayesi predikciót végezzünk, a (3) képlet szerinti összegzést és integrálást kell elvégezni. Ilyenkor az 2.4. fejezet közelítései alkalmazhatók.

3.5. Bayes-háló tanulása

Mivel a teljes bayesi következtetés annak komplexitása miatt csak különleges esetekben hajtható végre, gyakran a teljes modellter helyett csak egyetlen modellt használunk. Ha elegendő statisztikai adat áll rendelkezésre, a fent bemutatott manuális konstruálás mellett szerepet kaphat az optimális modell keresése, a *tanulás*, mely végezhető a szakértői modelltől kiindulva, annak finomításával, vagy tabula rasa alapon is. A tanulás, mint az optimális modell keresése, a parametrikus következtetés alkalmazásának tekinthető és megmutatható, hogy NP-teljes bonyolultságú [10], az adatok szükséges mennyiségére kívánt közelítési hiba mellett [15] ad képletet.

A tanulás két szinten lehetséges: kereshetjük adott struktúra mellett az optimális paraméterezést (parametertanulás), vagy a legjobb struktúrát és annak paraméterezését (struktúratanulás). Az optimalitás valamilyen

mérték szerint értendő, ez legegyszerűbb esetben a modell a posteriori valószínűsége.

A MAP modell keresése mellett elképzelhető más kritériumfüggvény is, amely leggyakrabban az a posteriori valószínűség egyenletes priorral, kiegészítve valamilyen, a struktúra bonyolultságát büntető taggal. Az ilyen büntetés alkalmazása felfogható a prior módosításának: minél erősebb a büntetés, annál kisebb a bonyolult struktúrák valószínűsége. A leggyakoribb ilyen minősítési függvény a bayesi információ-kritérium függvény (BIC – Bayesian information criterion), a képlete [11]:

$$BIC(G, D) = \log P(D|G) - \frac{\log N}{2} |\theta|, \quad (4)$$

ahol 'N' a tanító minták, $|\theta|$ pedig a háló paramétereinek száma. A $\log N$ -nel arányos mellett még elképzelhető N-ben lineáris vagy polinomiális büntetés is.

Számításigényét tekintve a tiszta a posteriori kritériumfüggvény, és a teljes, független, azonos eloszlású minták alapján végzett tanulás a legegyszerűbb. Ekkor, Dirichlet eloszlású paraméterprior feltéve, adott struktúra a posteriori valószínűsége egyszerű, zárt formában számítható [8]:

$$P(G|D) \propto P(G, D) = P(G) \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{(N'_{ij} + r_i - 1)!}{(N_{ij} + N'_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} \frac{(N_{ijk} + N'_{ijk})!}{N'_{ijk}!}, \quad (5)$$

ahol az N_{ijk} az i. változó j. szülői konfigurációjának és k. értékének az előfordulását, q_i az i. változó szülői konfigurációinak a számát és r_i az értékeinek számát jelenti (N_{ij} a megfelelő marginális). Az N'_{ijk} a megfelelő virtuális mintaszámokat jelöli (ezek előismeretek hiányában 1-nek választhatók).

Parametertanulás esetén az optimális paraméterezés az FVT-k külön-külön, relatív gyakoriságokkal való kitöltésével elérhető, struktúratanulás esetén pedig minden csomóponthoz külön megkereshető az optimális szülői halmaz, feltéve hogy ismert a csomópontok egy kauzális rendezése. (Egy kauzális rendezésben a csomópontok szülei csak az őket megelőző változók közül kerülhetnek ki. A kauzális rendezés a reprezentáns DAG csúcsainak egy topologikus rendezése.) Ha ilyen információ nem áll rendelkezésre, ügyelni kell, hogy a DAG tulajdonság ne sérüljön, például úgy, hogy minden lehetséges sorrendet külön megvizsgálunk.

3.6. Bayes-háló tanulása hiányos adatok alapján

Amennyiben a tanító adatok hiányosak, azaz bizonyos változók értéke nem minden esetben ismert a tanulás feladata jóval nehezebbé válik. Ilyenkor a parametertanulásban iteratív eljárások használhatók, a legismertebbek ezek közül a gradiens alapú közelítő eljárások vagy ezek robosztusabb változatai, a konjugált gradiens és a skálázott konjugált gradiens algoritmusok [7], vagy az expectation maximization algoritmus [13].

Struktúratanulás esetén, mivel a szülői halmazok nem tanulhatók külön még adott sorrendnél sem, a teljes struktúratertet bejáró keresésre van szükség. Mivel

a lehetséges struktúrák száma a csomópontok számával szuperexponenciálisan nő, a gyakorlatban nem teljes keresési eljárásokat kell alkalmazni, például mohó keresést vagy szimulált lehűtést (ekkor az elemi lépés pl. egy él törlése, beszúrása, vagy megfordítása lehet).

Ezek az eljárások is azonban csak akkor működnek, ha az adatokra teljesül a véletlenszerű eltűnés (MAR – missing at random) feltétele, azaz ha a bejegyzések eltűnése nem függ az eltűnt értéktől [13].

3.7. Jegytanulás

A jegytanulás során bizonyos részstruktúrák (jegyek) meglétének valószínűségét keressük. Ilyen jegy lehet a legegyszerűbb esetben például egy adott él megléte, vagy Markov-határ keresése. Egy X csomóponthalmaz Markov-takarója egy olyan Y halmaz, melyre igaz, hogy $I(X|Y|U \setminus (X \cup Y))$ (azaz Y d-szeparálja X -et a háló többi részétől). Egy csomópont vagy csomóponthalmaz Markov-határa annak minimális Markov-takarója. Ez lehetővé teszi egy szimmetrikus, páronkénti reláció definiálását a Markov-határbeliséget, az egymás Markov-határában való előfordulást (MBM(X, Y) – Markov boundary membership). A jegytanulás alternatív megoldást jelenthet a struktúratanulással szemben, mivel ha segítségével meg tudjuk állapítani a fent említett viszonyok meglétének valószínűségét (azaz, hogy egy csomópont beletartozik-e egy másik Markov-határába), akkor ezzel a MAP modell egy jó közelítését konstruálhatjuk.

A kérdéses valószínűségek számítása, a bayesi következtetés sémáját követi, amiből következően összegeznünk kell azon struktúrák a posteriori valószínűségét, amelyek rendelkeznek a kívánt jeggyel:

$$P(x|D) = \sum_{k:1_x[G_k]=1} P(G_k|D) \quad (6)$$

Természetesen itt is alkalmazhatók közelítő Monte Carlo módszerek, mivel a struktúrák feletti összegzés túl számításigényes, hacsak nincsenek rendkívül pontos a priori ismereteink a lehetséges struktúrákról.

4. Egy alkalmazási terület: petefészekrák-diagnosztika

A petefészekrák biológiájának és preoperatív diagnosztikájának kutatása inspirációként szolgált az integrált szöveg és adatelemzés általános problémáinak a vizsgálatában és elvezetett egy Bayes-hálót alkalmazó rendszer kifejlesztéséhez.

A leuveni egyetem (KUL) villamosmérnöki karának (ESAT) egy csoportjában (SCD/SISTA) az egyik szerző részvételével (A.P.) 1998-tól folynak a kutatások a petefészekrák preoperatív diagnózisával és általános biológiai modellezésével kapcsolatban, együttműködve az egyetem kórházával (Univ. Hospital Gasthuisberg). A kezdeti kutatások célja 1998 és 2000 között a petefészek daganatok preoperatív diagnosztikájában használható matematikai, statisztikai modellek kifejlesztése volt, a klinikán meglévő szakértői tudás és az ott gyűjtött adatok alapján. A második fázisban 2000 és 2002

között egy nemzetközi konzorcium alakult, amely a világ vezető petefészekrák kutatóit és diagnosztáit tömöríti, az International Ovarian Tumor Analysis (IOTA) konzorcium [31]. Ennek célja nagy mennyiségű, azonos protokoll szerint beszerzett és jelenlegi tudásunk alapján igen részletes betegleírás összegyűjtése, illetve a létrejött adatbázis alapján a tárgyterület átfogó statisztikai elemzése. A harmadik fázisban 2002-től folytatódik az IOTA konzorcium adatainak gyűjtése és elemzése, illetve a leuveni egyetem génchip laborjának közreműködésével 2003-tól megindult a daganatok genetikai profiljának elemzése is. Jelenleg a második fázis adatainak elemzése folyik, azonban a kifejlesztett módszerek, különösen, amelyek az integrált szöveg és adatelemzést támogató Bayes-hálókon alapulnak, már a harmadik fázis számára készültek, a génaktivitás mintázatok és a klinikai adatok együttes elemzésére.

4.1. A probléma leírása

A petefészekrák korai diagnosztikája kiemelkedő fontosságú, mivel jelenleg a páciensek kétharmadát már csak előrehaladott állapotban sikerül diagnosztizálni, ami a kezelések esélyeit nagyban lerontja. A petefészekrákhoz kapcsolódó a priori információk nagy mennyisége és sokszintűsége jól illusztrálja a „integrált adat és tudás” elemzés kihívásait általános problémákban is.

A rosszindulatú daganat kialakulásának magyarázatára több elmélet is létezik, amelyek az ovulációk számához, a gonadotropinok szintjéhez, a karcinogén anyagokhoz, illetve az örökletes vagy szerzett genetikai rendellenességekhez kapcsolódnak. A kockázatot befolyásoló ismert faktorok például a szülések száma, terméketlenség, a teherbe esést segítő hormonális kezelések, a szoptatási időszak hossza, hormonális fogamzásgátlók, karcinogének, mell- és petefészekrák családi előfordulása, életkor, méheltávolítás. További elérhető orvosi mérések és megfigyelések például a daganat alaktani és eresedési leírói, vagy a tumormarkerek szintjei (például CA 125). A faktorok egy részének a hatását kvantitatívan is ismerjük (bizonyos genetikai rendellenességek esetén a kockázat megnövekedését), más faktoroknak azonban már a megállapítása, mérése is erősen szubjektív [31].

4.2. A priori információk

A petefészekrák preoperatív diagnosztikájához kapcsolódó, klinikai gyakorlatban használt változók átfogó modellezéséhez a következő információforrások álltak rendelkezésre:

1. Az IOTA konzorcium által kidolgozott terminológia és adatgyűjtési protokoll, amely a petefészekrák ultrahangos diagnosztikájához kapcsolódó, a klinikai gyakorlatban használt fogalmak elméleti és gyakorlati meghatározását tartalmazza (egy tárgyterületi részontológia).
2. Elektronikusan elérhető teljes publikációk és kóvontatok, amelyek közül a legfontosabb cikkek száma ezres, a potenciálisan releváns cikkek száma már tíz-

ezres nagyságrendű. További természetes nyelvű, részben strukturált információforrások az orvosi lexikonok, amelyek közül felhasználtuk az Online Medical Dictionary és CancerNet Dictionary szócikkeit, és részleteket a Merck's Manual-ból. Kiemelt fontosságú dokumentumok a már említett IOTA adatgyűjtési protokoll.

3. Általános orvosi szótárak, taxonómiák, tezaurszok, mint a Medical Subject Heading (MeSH).
4. Részleges statisztikák: általános demográfiai adatok, petefészekrákhoz kapcsolódó általános statisztikák (például az USA NCI SEER adata), korábban publikált petefészekrák kutatások statisztikái.
5. Szakértői ismeretek az IOTA konzorcium résztvevőitől.

Az előző információforrások igen sokrétű és sokféle típusú a priori információt tartalmaznak explicit vagy implicit módon a problémára, a változókra, azok kvalitatív és kvantitatív relációira vonatkozóan. A munka során a következő explicit a priori információkat hoztuk létre vagy származtattuk.

4.3. Szótárak

Egy hétszáz szavas szótárt, egy ehhez tartozó szinonima listát és szakkifejezések listáját. Ezek részben az IOTA konzorcium terminológia meghatározásából és az IOTA adatgyűjtési protokollból, illetve szóstatistikák szakértői elemzése alapján lettek kézilleg összeállítva. Automatikus eszközökkel, illetve a MeSH általános orvosi szótár felhasználásával több nagy méretű, egymillió szószám feletti szakszótárt is előállítottunk.

4.4. Dokumentum gyűjtemények

Elsőként két orvosi szakértő az elektronikusan elérhető MEDLINE dokumentumgyűjteményből kiválasztotta az IOTA kontextusnak leginkább megfelelő hivatkozásokat az egyes szakterületi változókhoz. 42 illetve 22 különböző szakcikk került így kiválasztásra, 3-5 cikk változónként. E dokumentumoknak, mint a szakterületre és feladatra leginkább specifikusnak az úgynevezett relevancia faktorát a legmagasabb állítottuk be.

A szakértők kiválasztották az IOTA kontextushoz legrelevánsabb szaklapokat (2 db), az igen releváns (3 db), közepesen releváns (33 db) és a releváns újságokat (93 db). Ezek alapján létrehoztunk öt egymásba ágyazott dokumentumgyűjteményt a MEDLINE 1982 és 2003 közti kivonatai alapján, amelyek így 45, 5.367, 71.845, 231.582 és 378.082 kivonatot tartalmaznak.

Létrehoztunk egy további dokumentumgyűjteményt az On-line Medical Dictionary és a CancerNet Dictionary alapján, amelyek együttesen 67.829 szócikket tartalmaznak és a változók leírásai szintén tartalmaznak hivatkozásokat az itteni szócikkekre.

Végül még három technikai jellegű dokumentumgyűjteményt hoztunk létre az IOTA protokoll, egy petefészekrák diagnosztikájáról szóló Ph.D tézis és a Merck Manual alapján. Ezek a gyűjtemények szakértők által kiválasztott szócikkeket tartalmaznak az egyes változókhoz, illetve azok csoportjaihoz (részletesebb leírások az [1] és [4]-ben).

4.5. Változók közötti relációk

Az a priori információforrásokból a következő explicit relációkat, illetve relációkra vonatkozó ismereteket származtattuk:

- változók csoportosítása (például alaktani változók, eresedéssel kapcsolatos változók)
- változók értékeire vonatkozó szükségszerű logikai összefüggések,
- páronkénti, közvetlen statisztikai függőségek, okozati, kvalitatív monotonitási és hatáserősségi információval,
- többváltozós okozati mechanizmusok, kvalitatív hatáserősségi információval
- részleges statisztikák, függőségek kvantitatív jellemzése.

4.6. Adatok

A későbbiekben bemutatott eredményekben egyrészt az IOTA projekt által gyűjtött adatok egy előzetes, részleges adathalmazát használtuk fel, amely 782 esetet tartalmaz[4], másrészt a klinikai adatok mellett felhasználtuk a dokumentumgyűjteményekből származtatott bináris szakirodalmi adatokat, amelyekben egy bejegyzés a tárgyterület változóinak explicit előfordulását vagy egy küszöbértékhez kötött implicit relevanciáját reprezentálja.

4.7. Integrált adat- és szövegelemzés Bayes-hálókkal

A felsorolt a priori tudáselemeket és az adatokat egy „annotált” Bayes-hálós tudásbázisban reprezentáltuk, amit a kifejlesztett rendszer tárol (1. ábra).

A rendszer az akadémiai és kereskedelmi Bayes-hálókhoz kapcsolódó szoftverekhez képest amellelt, hogy tartalmazza a megszokott tudásmérnöki, következtetési és tanulási támogatást, a következő egyedi tulajdonságokkal bír:

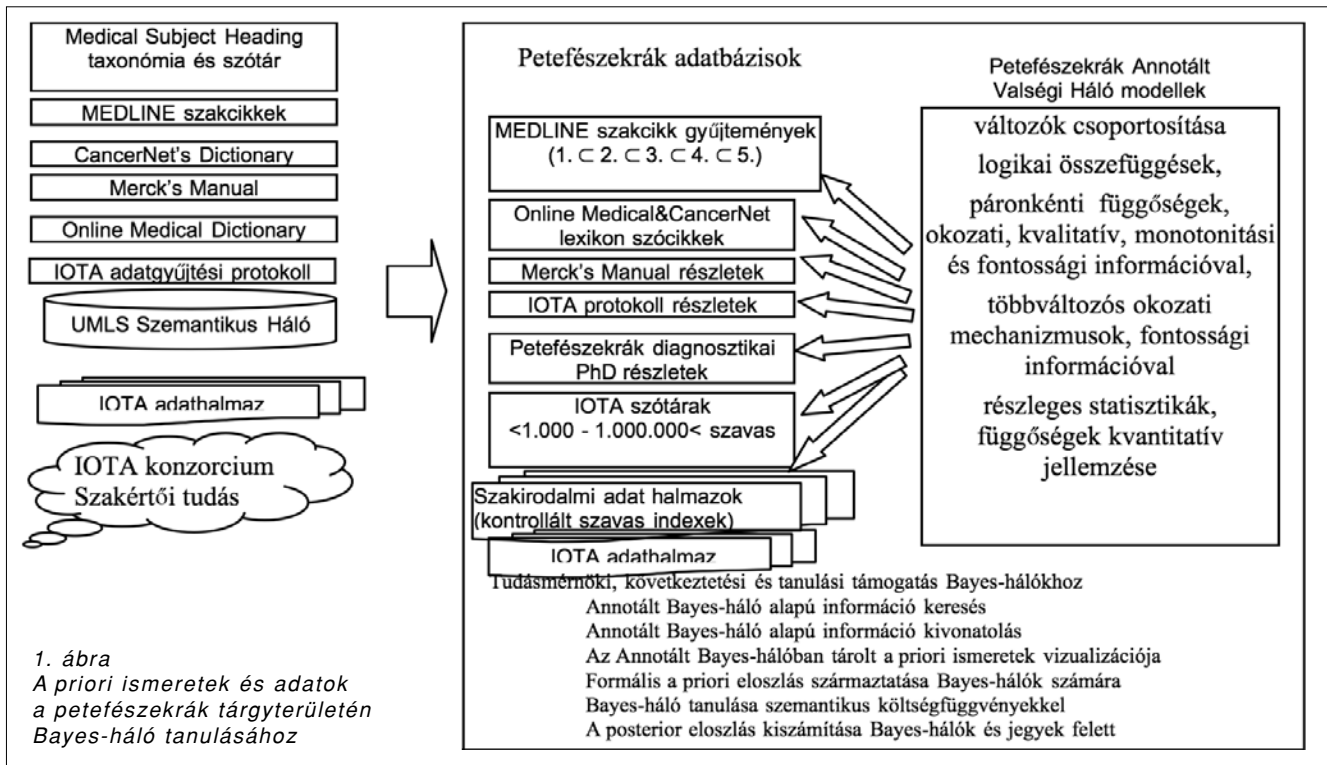
- Tárgyterületimodell-alapú és személyre szabott információkeresés, amelyben egy kifejlesztett lekérdezősi nyelv segítségével az épített vagy tanult annotált Bayes-háló alapján a tudásmérnöki kontextusnak megfelelő relevanciamérték definiálható az illeszkedő szakcikk megtalálására [1].

- Statisztikai információkivonatolás, amely az egyes szakcikk releváns fogalmait tartalmazó adatbázis elemzésén alapul Bayes-hálós modellekkel. Az alkalmazott modellek lehetnek a fogalmak előfordulását leíró valószínűségi modellek, illetve a szakcikk keletkezésének és írásának generatív (okozati) modelljei [4,5].

- Tárgyterület specifikus modelltanulás, mivel az annotált Bayes-hálós tudásbázist felhasználva háttérismereteket is tartalmazó költségfüggvény definiálható a kiválasztott modellre ($L(G^A, G)$), ami az posteriorral együtt definiálja a modellek várható jószágát).

- Egyszerű és komplex Bayes-hálóbeli struktúrális jegyek a posteriori eloszlásának kiszámítását vagy Monte Carlo becslése.

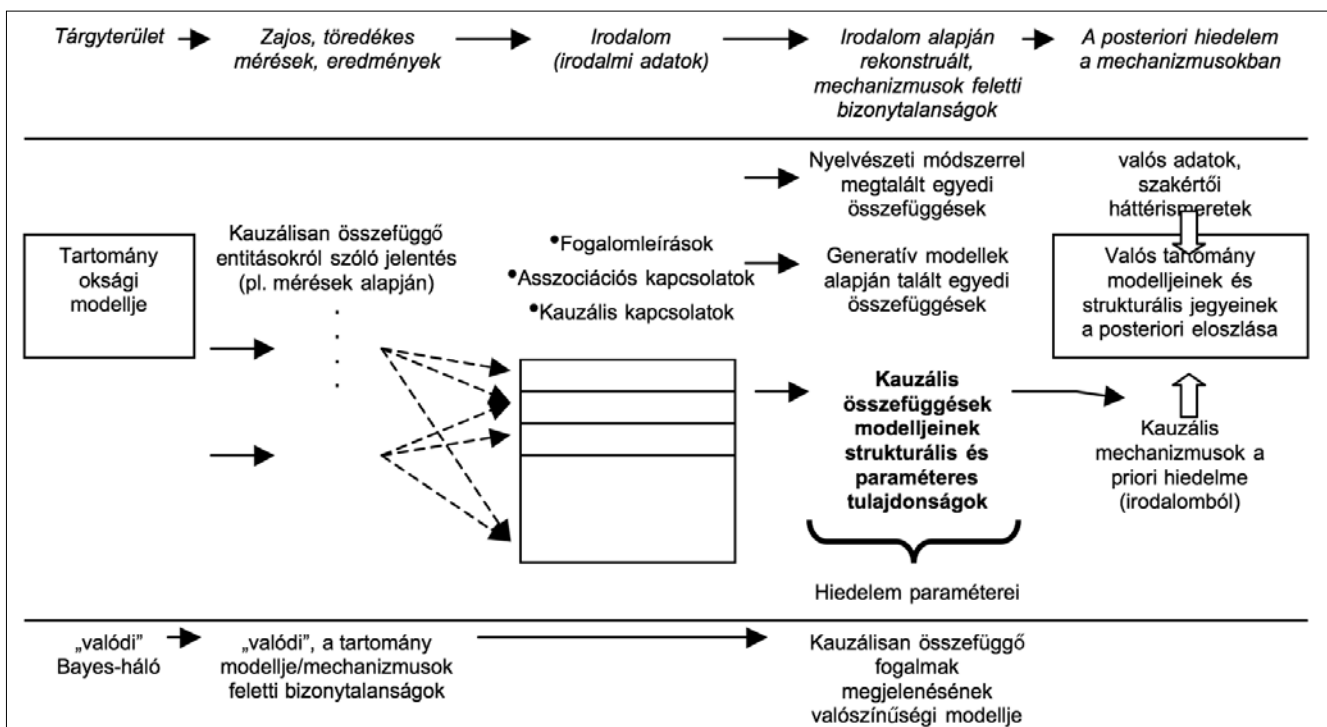
- Osztályozó konstruálás támogatása a priori eloszlások indukálásával osztályozós modellstruktúrákra és paraméterekre [3].



Ezek a kutatások főként az IOTA projekthez kapcsolódva fejlődtek. Rájuk épülve vagy részben kapcsolódva új kutatási irányok a Bayes-hálóbeli strukturális jegyek elsőrendű valószínűségi logikán belüli kezelése és lokális kauzális algoritmusok vizsgálata a teljes bayesi megközelítés mellett [20]. A szakirodalmi „adat” elemzése mindegyik esetben központi helyet foglal el, akár mint teszt terület vagy cél. Az integrált adat- és szövegelemzést a 2. ábra mutatja be.

A szakirodalom statisztikai elemzésére, a Bayes-háló Bayes-statisztikai keretrendszerben történő felhasználására két eredményt mutatunk be, amelyek az (5)-(6) egyenlet szerinti posteriorokat mutatják sorrendi alapú Monte Carlo Markov Chain módszerekkel megbecsülve [12]. A 3. ábra (köv.old.) baloldalán az irányítatlan élek azokat a páronkénti Markov-határbeliséget mutatják, amelyek a posteriori valószínűsége egy adott küszöbérték feletti, illetve a szakértőtől származó priori

2. ábra Az integrált adat- és szövegelemzés Bayes-hálókkal



valószínűség szerint. A jobb oldalon a Markov-határbeliség posteriori valószínűségének az alakulását mutatjuk be a nagy Medline dokumentumgyűjteményt használva, ahol minden év esetén az előző öt év publikációt használtuk fel adatként.

5. Kitekintés

Az eddigi fejezetek rövid áttekintést adtak a monolitikus Bayes-hálók használatáról. A monolitikus jelző ez esetben arra utal, hogy egy adott problémára konstruált hálóban nincsen hierarchikus vagy moduláris dekomponálás. A következőkben rövid áttekintést adunk a Bayes-hálók kiterjesztésére törekvő irányzatokról.

Az első lépést ebben az irányban az annotált Bayes-hálók vizsgálatával tettük meg, ami lehetőséget adott tetszőleges szemantikai információ bevitelére és automatizált felhasználására. A következő lépés a már említett jegytanulás volt, mivel ennek felhasználásával felfedezhetők reguláris hálórészletek (bizonyos területeken gyakori az ok-okozati mechanizmusokban felfedezhető, ismétlődő mintázat, például a biológiában egyes gének aktivációs sémái).

A modularizációs igényre adott formális válasz az objektumorientált Bayes-hálók (OBN) megjelenése volt [22]. Mint nevük is mutatja, a programozástechnikában ismert objektumorientált paradigmához hasonlóan terjesztik ki a Bayes-hálókat. Egy objektumorientált Bayes-hálózat objektumokból áll, melyek szintén tovább bonthatók objektumokra, vagy egyszerű valószínűségiválasztó-csomópontokra. Ezzel a többszintű hierarchiával a teljes rendszer funkcionálisan különálló részei elszigetelhetők egymástól, valamint lehetővé válik előre felépített részhalóknak a teljesbe építése. Hasonló koncepció áll a valószínűségi relációs modellek mögött is [21].

6. Összegzés

A cikkben bemutatott Bayes-hálók Bayes-statisztikabeli alkalmazása mögött a következő általános trendek azonosíthatók be.

A számítási kapacitás növekedésével a Bayes-statisztika gyakorlatban is fontos, komplex modellek felett is alkalmazhatóvá vált, elsősorban a Monte Carlo módszerek alkalmazásával. Az elektronikus elérhető a priori ismeretek mennyiségének növekedése szintén a Bayes-statisztikai megközelítést helyezte előtérbe, hiszen az adatok mennyiségének általános növekedése gyakran még mindig nem elegendő a szükséges modell komplexitásához képest. A két trend eredményeképpen a Bayes-statisztika egy normatív tudás és adat integrálást tesz lehetővé a számítási erőforrások intenzív, de az MCMC módszerek miatt egységes alkalmazásával.

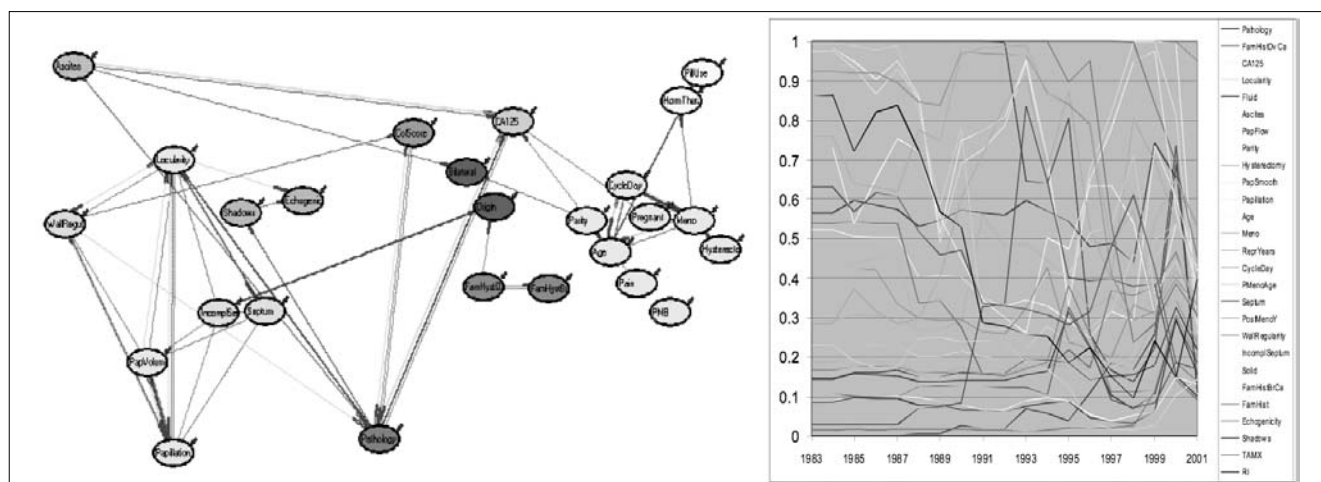
A Bayes-hálók szintén ebbe a két trendbe illeszthetők, egyrészt mint számításigényes modellosztály, másrészt mint az a priori ismereteket és megfigyeléseket vagy kísérleti adatokat integráló modellosztály. További előnye, hogy három kapcsolódó szinten is értelmezhető a modell, mint az együttes eloszlás hatékony faktORIZÁLÁSA, mint az együttes eloszlás feltételes függetlenségeinek explicit reprezentálása és mint a tárgyterület okozati kapcsolatainak az ábrázolása.

A bemutatott orvosi biológiai alkalmazás mellett ezek más területeken is megmutatkozó, általános trendek. A jelenlegi kutatások a reprezentáció dekomponálását, hierarchizálását és strukturált információkkal történő formális kiegészítését célozzák.

Irodalom

- [1] P. Antal, T. Mészáros, D. Timmerman, B. De Moor, T. Dobrowiecki: Domain knowledge based information retrieval languages: an application of annotated Bayesian networks, Fifteenth IEEE Symposium on Computer-Based Medical Systems (CBMS 2002), June 3-7, Maribor, Slovenia, pp.213–218.
- [2] P. Antal, P. Glenisson, G. Fannes, Y. Moreau, B. De Moor: On the potential of domain literature for clustering and Bayesian network learning, The Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (ACM SIGKDD), 2002, Edmonton, pp.405–414.

3. ábra Balra az adott küszöbérték (0.5) feletti valószínűségi Markov-határbeli relációk, jobbra a Markov-határbeliség posteriori valószínűségének alakulása a Szövetten és a többi IOTA változó között



- [3] P. Antal, G. Fannes, D. Timmerman, Y. Moreau, B. De Moor: Bayesian Applications of Belief Networks and Multilayer Perceptrons for Ovarian Tumor Classification with Rejection, *Artificial Intelligence in Medicine*, 2003, vol.29, pp.39–60.
- [4] P. Antal, G. Fannes, Y. Moreau, D. Timmerman, B. De Moor: Using Literature and Data to Learn Bayesian Networks as Clinical Models of Ovarian Tumors, *Artificial Intelligence in Medicine*, 2004, vol.30, pp.257–281.
- [5] P. Antal, A. Millinghoffer: Learning Causal Bayesian Networks from Literature Data, In Proc. of the 3rd Int. Conf. on Global Research and Education, Inter-Academia'04, Budapest, 6-9. September 2004.
- [6] J. M. Bernardo: *Bayesian Theory*, Wiley & Sons, Chichester, 1995.
- [7] C. M. Bishop: *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford, 1995.
- [8] G. F. Cooper, E. Herskovits: A Bayesian Method for the Induction of Probabilistic Networks from Data, *Machine Learning*, 1992, vol.9, pp.309–347.
- [9] R. G. Cowell, A. P. Dawid, S. L. Lauritzen, D. J. Spiegelhalter: *Probabilistic Networks and Expert Systems*, Springer Verlag, 1999.
- [10] N. Friedman, Z. Yakhini: On the Sample Complexity of Learning Bayesian Networks, *Proc. of the 12th Conf. on Uncertainty in Artificial Intelligence*, 1996, pp.274–282.
- [11] N. Friedman: Learning Belief Networks in the Presence of Missing Values and Hidden Variables, *Proc. of the 14th Int. Conf. on machine learning*, 1997, pp.125–133.
- [12] N. Friedman, D. Koller: Being Bayesian about Network Structure, *Journal of Machine Learning Research*, Kluwer Academic Publ., Dordrecht, Netherlands, 2002. vol.2, pp.1–30.
- [13] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Rubin: *Bayesian Data Analysis*, Chapman & Hall, London, 1995.
- [14] W. R. Gilks, S. Richardson, D. J. Spiegelhalter (edit.): *Markov Chain Monte Carlo in Practice*, Chapman & Hall, 1995.
- [15] C. Glymour, G. F. Cooper: *Computation, Causation, and Discovery*, AAAI Press, 1999.
- [16] D. Heckerman, D. Geiger, D. Chickering: Learning Bayesian networks: The Combination of Knowledge and Statistical Data, *Machine Learning*, vol.20, 1995, pp.197–243.
- [17] D. Heckerman, A. Mamdani, M.P. Wellman: Real-world applications of Bayesian networks, *Communications of the ACM*, vol.38, issue 3, 1995, ACM Press, New York, pp.24–26.
- [18] E. Horvitz, J. Breese, D. Heckerman, D. Hovel, K. Rommelse: The Lumiere Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users, *Proc. of the 14th Conf. on Uncertainty in Artificial Intelligence*, July. 24-26. 1998, Madison, Wisconsin, USA, pp.256–265.
- [19] C. Huang, A. Darwiche: Inference in Belief Networks: A procedural guide, *Int. Journal of Approximate Reasoning*, vol.15, 1996, pp.225–263.
- [20] Hullám G.: Bayes-hálók strukturális tulajdonságainak tanulása kényszer alapú módszerekkel, diplomamunka, BME-MIT, 2005.
- [21] D. Koller, A. Pfeffer: Probabilistic Frame-Based Systems, In *Proc. of the 15th National Conf. on Artificial Intelligence (AAAI)*, Madison, Wisconsin, 1998. pp.580–587.
- [22] K. Laskey, S. Mahoney: Network Fragments: Representing Knowledge for Constructing Probabilistic Models, In *Proc. of the 13th Conf. on Uncertainty in Artificial Intelligence (UAI-1997)*, Morgan Kaufmann, 1997, pp.334–341.
- [23] P. Myllymaki, T. Silander, H. Tirri, P. Uronen: Bayesian Data Mining on the Web with B-Course, In *Proc. of The 2001 IEEE Int. Conf. on Data Mining*, IEEE Computer Society Press, 2001, pp.626–629.
- [24] J. Pearl: *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, San Francisco, 1988.
- [25] K. W. Przytula, T. Lu: Bayesian Networks Based Diagnostic Tools for Locomotives: Model Development and Inference, *The 2nd Bayesian Modeling, Applications Workshop During UAI-04*, July 7th, 2004.
- [26] S. Ramamurthy, H. Arora, A. Ghosh: Operational Risk and Probabilistic Networks – An Application to Corporate Actions Processing, Banking & Capital Markets Solutions Consulting, Infosys Techn. Ltd, www.hugin.com/cases/Finance/Infosys/oprisk.article
- [27] C. P. Robert: *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, Springer-Verlag 2001.
- [28] M. Sahami, S. Dumais, D. Heckermann, E. Horvitz: A Bayesian Approach to Filtering Junk E-mail, *AAI Workshop on Learning for Text Categorization* July 1998, Madison, Wisconsin, AAAI Technical report WS-98-05.
- [29] C. Skaanning, F.V. Jensen, U. Kjærulff, L. Parker, P. Pelletier, L. Rostrup-Jensen: SACSO – A Bayesian-Network Tool for Automated Diagnosis of Printing Systems, Machine Intelligence Group, Aalborg University, Technical Report, 1998.
- [30] A. P. Tchangani: Decision Support System with Uncertain Data: Bayesian Networks Approach, www.ici.ro/ici/revista/sic2002_3/art2.htm
- [31] D. Timmerman, L. Valentin, T. H. Bourne, W. P. Collins, H. Verrelst, I. Vergote: Terms, definitions and measurements to describe the sonographic features of adnexal tumors: a consensus opinion from the International Ovarian Tumor Analysis (IOTA) Group, *Ultrasound Obstetrics Gynecology*, vol.16, 2000, pp.500–505.
- [32] E. Wright, J. Fitzgerald, D. Barbara, T. Shakelford, K. Laskey, G. Alghamdi, X. Wang: Detecting insider threats in information systems, *The Second Bayesian Modeling Applications Workshop During UAI-04*, July 7th, 2004.

Komplex hibamenedzsment megoldás VoIP-szolgáltatások felügyeletéhez

VARGA PÁL, MOLDOVÁN ISTVÁN, MOLNÁR GERGELY*

Budapesti Műszaki és Gazdaságtudományi Egyetem, Távközlési és Médiainformatikai Tanszék
{pvarga,moldovan}@tmit.bme.hu

* Ericsson Magyarország Kft., gergely.molnar@ericsson.com

Kulcsszavak: VoIP, szolgáltatásminőség, hibamenedzsment

A jó minőségű VoIP (Voice over IP) megoldások elterjedésével egyre növekszik a VoIP-szolgáltatásokat használók száma. A szolgáltatás-minőség biztosításához megfelelő felügyeleti rendszerre és a VoIP-szolgáltatásokra specializálódott hibamenedzsmentre van szükség. Ebben a tanulmányban egy olyan átfogó hibamenedzsment rendszert mutatunk be, amely – a VoIP-szolgáltatók igényeire szabottan – mind a hibadetektálás, mind pedig a hibaok-meghatározás során felmerülő kérdésekre képes választ adni. A rendszer a hagyományos passzív, szabály alapú hibafeldolgozási algoritmusok használatát egy aktív, hibakereső ellenőrzés támogató módszertannal egészíti ki. Az aktív hibaok-keresési lépések ütemezését egy újszerű, a Petri-háló leírásán alapuló módszer teszi áttekinthetővé és – a feladatok párhuzamos végrehajtásának köszönhetően – gyorsá.

1. Bevezetés

VoIP szolgáltatásokat nem csak akkor veszünk igénybe, amikor számítógépünkről egy klienssel az Internet lehetőségeit használva telefonálunk, hanem sokszor „észrevétlenül” is, amikor hagyományos telefonunkkal hívást kezdeményezünk. Ahhoz, hogy a VoIP-hálózatok által hordozott hívások elfogadható minőségűek legyenek, a megfelelő hálózati méretezés mellett a szolgáltatás üzemeltetésére is kellő figyelmet kell fordítani.

A VoIP hálózatok szolgáltatásainak fennakadásmentes működtetéséhez célszerű átfogó hibamenedzsment rendszert (Fault Management System – FMS) üzemeltetni. Ez a cikk egy, a VoIP szolgáltatások (és nem csak a VoIP hálózat) üzemeltetését segítő FMS működését mutatja be. A VoIP-szolgáltatási hibák (faults) orvoslása a VoIP-szolgáltató feladata, míg a hordozóhálózat hibáinak kezelése a hálózati infrastruktúra üzemeltetőinek hatásköre.

A rendszer a hálózati elemek által jelzett események folyamatos gyűjtése és feldolgozása során képes kiszűrni az egyes hibákat, majd javaslatot tud adni a hibaokok helyére és a hibaelhárítás lépéseire. A hálózat működését felügyelő operátor munkája ezzel jelentősen egyszerűsödik, ám sohasem válik feleslegessé, hiszen az összetettebb hibák felismerése és kiküszöbölése továbbra is feladata marad.

A VoIP-szolgáltatás minőségi mutatóit a hálózati elemek, az IP-hálózat mint üzemeltetendő entitás, valamint a VoIP-hoz kötődő alkalmazások nem kívánt működéséből fakadó hibák is leonthatják. A fenti elemek állapotában bekövetkező változások „normális” és „hibát jelző” eseményeket is generálnak, ezek az FMS-be jutnak. Az egyes események önálló kiértékelésével az FMS-ben nem mindig lehet eldönteni, hogy az pontosan a hiba okáról, valamilyen hiba továbbterjedéséről, vagy csak egy normálisnak tekinthető állapotváltozásról számol-e be. Mivel egy hibajegy-típus – például „al-

kalmazás-szerver nem elérhető” –, több helyen és módon is rögzítésre kerülhet, az eredendő hiba feltárása ezen redundáns adatokból igen összetett feladat.

Hibamenedzsment alatt azt a folyamatot értjük, mely során a hibasokaságot az operátor a hibadetektálás, hibajel-feldolgozás, hibaok-meghatározás és hibajavítás lépésein keresztül megszünteti. A hibadetektálás folyamata során a hálózat különböző eseményüzenetforrásainak kimenetét gyűjtjük egybe, és alakítjuk azonos formátumúvá. A hibajel-feldolgozó céljainknak megfelelő szűréseket, eseménykorrelációs- és trendanalízis-műveleteket végez ezen az üzenethalmazon. Az így kialakult hibajegyek mindegyike egy-egy meghatározott hibajelenséget ír le, melynek eredetét a hibaok-meghatározás során derítjük fel. A hibamenedzsment folyamat végső eredménye a hibaok javítására adott javaslattevél. Az egyes lépéseknek megfelelő hibamenedzsment-funkciók kialakítása során felmerülő kérdéseket a későbbi fejezetekben ismertetjük.

A távközlésben hagyományosan használt hibamenedzsment módszerek egyik legjobban dokumentált, átfogó keretrendszere a TMN (Telecommunications Network Management) [11]. Ennek a magyar viszonyokra adaptált és több nagyszerű ötlettel bővített változata a [12]. Ezek a rendszerek hierarchikusan épülnek fel, és talán ez az oka, hogy nem elég rugalmasak – nem kezelik elég hatékonyan a topológia-változásokkal járó következményeket. Az utóbbi évtizedekben használt FMS keretrendszerek és az újabb hibaok-analízis módszerek rövid bemutatása megtalálható a [6] forrásban.

Ezen VoIP hibamenedzsment-rendszer prototípusát a BME Távközlési és Médiainformatikai Tanszékének próbahálózatán üzemeltük be, az Ericsson Magyarország Kft. és a Kovax'95 Kft. munkatársainak közreműködésével. A hívásadatokból kinyerhető hiba-információ elemzését a NIIF VoIP-hálózatából származó anonim adatsorokon végeztük. Jelen cikkben részletezett módszertan a [10] workshopon is bemutatásra került.

2. Hibamenedzsment

Egy komplex hibamenedzsment rendszernek az alábbi funkciókat kell megvalósítania:

- észleli, és összegyűjti a hálózat működése során kialakuló hibaüzeneteket,
- segít a hibaüzenetek szűrésében,
- diagnosztikai tesztek futtat a hibaforrás felderítésére, javaslatot tesz a hibajavítására.

A hálózatmenedzselő alkalmazás a felügyeleti központban figyeli a hálózatot, hibajelenségekre „vadászva”. A megjelenő hibákat a hibadetektálás, hibajel-feldolgozás, hibaok-meghatározás és hibajavítás lépésében keresztül kezeli (1. ábra). A folyamat eredménye egy hibajavítási javaslattevél. Az operátor ennek alapján elkezdheti a hiba okának megszüntetését. Ha a hibaok-meghatározás során a rendszer nem képes egészen pontosan meghatározni a hiba okát, a folyamat naplójának vizsgálatával az operátor látja az elvégzett lépéseket és azok eredményeit, így ezek ismétlése nélkül további irányokban keresheti a hiba okát.

A folyamat elemeire való hivatkozás megkönnyítése érdekében érdemes definiálni a hibajel és a hibajegy fogalmát. *Hibajel* alatt a hibadetektálás során keletkező esemény-jelzéseket értjük. Ez magában foglalja a hálózati elemek által küldött státusz-változások eseményjelzéseit (nem csak a figyelmeztetéseket vagy konkrét hibajelzéseket), valamint a hibamenedzsment rendszer aktív elemeinek eseményjelzéseit. *Hibajegy* alatt a hibafeldolgozó alrendszer kimeneti jelzéseit értjük, azaz azokat a hibára utaló figyelmeztetéseket, amelyekkel a hibaok-meghatározó rendszernek valamint az üzemeltető operátornak feltétlenül foglalkoznia kell.

3. Hibadetektálás

A hibadetektálás célja a VoIP szolgáltatást veszélyeztető hibajelenségek minél hamarabbi érzékelése és az ezekről szóló jelentések továbbítása hibafeldolgozásra.

A hibajel-adatbázis kialakításához ezen státusz-információt tartalmazó üzenetek mellett a hibamenedzsment-rendszer aktív ellenőrző elemeinek jelzéseit is felhasználjuk. Ezen adatbázis kialakításánál azzal a nehézséggel találkozunk, hogy a különböző hibajel-források által küldött hibajelek különböző formátumban érkeznek. A későbbi munka megkönnyítése érdekében ezeket a hibajeleket egységes formátumba kell rendeznünk, továbbá kezelői felületet kell biztosítanunk a vizsgálatukra. Az így átalakított hibajeleket a hibafeldolgozás során már könnyen tudja kezelni a rendszer.

A VoIP-szolgáltatások hibamenedzsmentje során a következő típusú *hibajelekkel* találkozhatunk:

- a hálózatban használt hibadetektáló elemek által kiszűrt specifikus VoIP hibaüzenetek (Syslog, QoS monitor (próbahívó) [2,3]),
- a VoIP hívások során keletkező hívás rekordok (üzenetek) (Radius rekordok [4,5] – ezekből leginkább a hibásnak minősülő bontási ok-kódok gyakori előfordulása utalhat súlyos hibára),
- a monitorozó elem használata közben keletkezett üzenetek (Aktív monitor),
- felhasználók által jelzett hibák (HelpDesk rendszer).

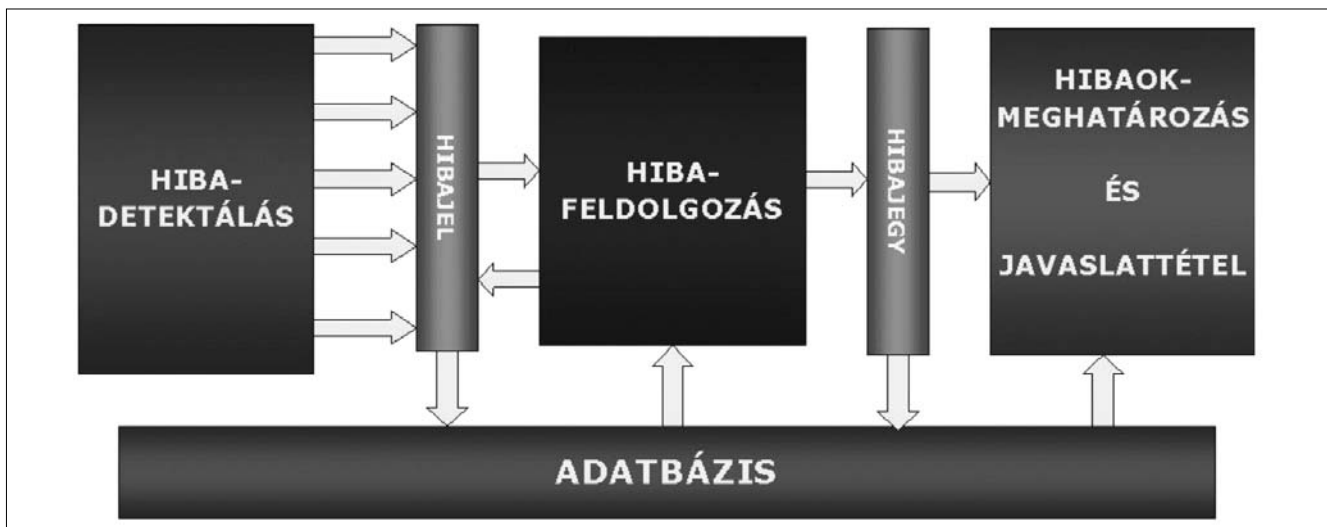
A hibamenedzsment rendszer kizárólag a VoIP-szolgáltató által javítandó hibákat értékeli. Ez azt jelenti, hogy bár a rendszer alkalmas egyes hordozóhálózati hibák jelzésére, nem feladata pontos javaslatot tenni a hordozóhálózat hibáinak javítására. Ehhez alacsonyabb szintű hálózati hibák figyelésére kialakított hálózatmenedzsment-alkalmazások szükségesek.

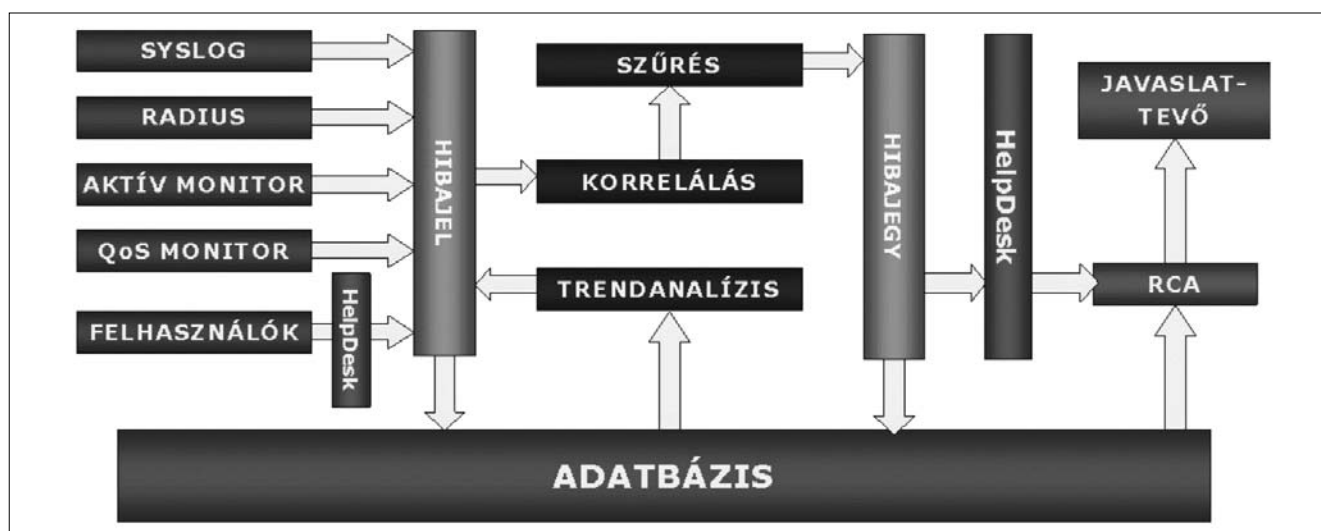
4. Hibafeldolgozás

A hibadetektáló alrendszer üzenetei a hibajel-adatbázisba (lásd az 1. ábrán) kerülnek. Ezek az esemény-leírás típusú rekordok képezik a hibafeldolgozás bemeneti adathalmazát.

Az adatbázis manuális feldolgozása meghaladja az üzemeltető operátor lehetőségeit, hiszen egy absztrakt

1. ábra A hibamenedzsment folyamatának elemei és ezek egymásrahatása





2. ábra A hibajelek feldolgozásának folyamata

és valódi értelemben is többszörözött, másodpercenként akár több tucatnyi rekorddal növekvő adathalmazról van szó.

A hibafeldolgozás egy automatizált adatfeldolgozó folyamat, melynek eredményeként hibajelek helyett csak a *hibajegyeket* (alarmokat) tárjuk az operátor elé. A hibajelek feldolgozásának menetét a 2. ábra szemlélteti.

A hibajeleket a későbbi hasznosítás érdekében az adatbázis tárolja. Feldolgozásuk a korrelátor modulban kezdődik. Az adatbázisba került hibajeleket a trendanalizátor folyamatosan elemzi – ennek eredményeképpen új, trend-jellegű hibajelet generál. A redundancia-mentesítést, és a „lényeg kiemelését” a szűrő modul végzi. A hibafeldolgozó modulok (szűrő, korrelátor, trendanalizátor) működését a későbbiekben vázoljuk.

4.1. Szűrő

A legegyszerűbb hálózat-felügyeleti algoritmus a szűrés. A jelenlegi hibamenedzsment-rendszerekben használt szűrési algoritmusok közül [6] a szabály alapú szűrést [7] választottuk. Ennél a módszernél a szűrőmodul egyenként megvizsgálja a beérkező hibajeleket, hogy talál-e rájuk alkalmazható szabályt. Amennyiben ilyen szabály nincs, a hibajelből alapértelmezés szerint hibajegyet generál. Minden más esetben az alkalmazott szűrőszabálynak megfelelően jut tovább, vagy nyomódik el a hibajel. A szabályok a leírásukban szereplő alapparaméterek (érvényességi időintervallum, hibajel kód, darabszám) értékének és az adott szabály típusának (elnyomó, számláló, redundancia gátló, dominancia) megfelelően fejtik ki hatásukat [1]. Az aktív szabályrendszert az operátor szabadon változtathatja: az új szabályok összeállítása mellett a létező szabályokat módosíthatja, vagy törölheti is.

Az operátori munka megkönnyítéséhez a szabályrendszernek legalább a következő szűrési igényeket kell kielégíteniük:

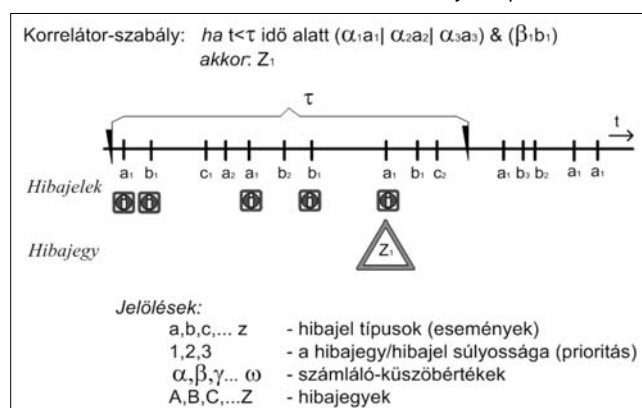
- a folyamatosan beérkező nem súlyos hibajelek esetében ne generáljon mindig egy újabb hibajegyet (számláló és elnyomó típusú szűrő);

- ha egy hibajegy keletkezett a rendszerben, akkor addig ne engedjünk több hibajegyet generálódni, amíg tudjuk, hogy a hasonló hibajelből keletkezett hibajegy még aktív (redundancia gátló);
- a hatásukban súlyosabb (nagyobb prioritású) hibajelek nyomják el a „gyengébb” (kisebb prioritású) hibajeleket, melyek ugyanarra a jelenségre, hibára utalnak (dominancia szűrő).

4.2. Korrelátor

Az eseménykorreláció a szűrésnél magasabb szintű feldolgozási módszer, amely különböző hibajelek közti összefüggések felderítésére hivatott. Több hibajel rövid időn (néhány percen) belüli beérkezése esetén módunk van egy sokkal specifikusabb hibajel-leírást adni. A korrelátor alkalmazása során nem veszítünk hibajeleket: a szabálynak megfelelően mind a bementi, mind az új, specifikus hibajelet továbbküldi (a szűrő modulba, amely megfelelő módon nyomja majd el ezeket). A szabályokat az operátor a szűrőszabályokhoz hasonló módon szabadon változtathatja. Az esemény-korrelátor végső célja, hogy lehetőség szerint átfogóbb hibajelet generáljon, számos hibaok-elemet vonjon össze egy eredményes okká. A korrelációs szabályok alkalmazásának hatását a 3. ábra szemlélteti [8].

3. ábra Szabály alapú korrelátor



4.3. Trend-analizátor

A trend-analizátor alkalmazásával elérhetjük, hogy adott hibajelek előfordulásából előre jelezni tudjuk a rendszer rendellenes viselkedését. A trend-analízis rejt magában némi bizonytalanságot, hiszen a trend meghatározására nincs bizonyítottan jó automatikus módszer. Az aktuális folyamatra kell találni egy jó trendfüggvényt, amivel lehetőségünk nyílik jósolni. Jelen esetben valamelyest más a helyzet, itt az algoritmus előre megírt szabályoknak eleget tevő hibamintákat keres.

A hibabekövetkezés előrejelzésénél a szabály által meghatározott trendváltozót vizsgáljuk, például ennek kell egy bizonyos küszöbértéknél nagyobblnak lennie. A hibáknak egy adott időn belül kell bekövetkeznie, sorrendiségük mellékes tényező. Tapasztalataink szerint a trend-analízis segítségével alapvetően telítődéses (a hiba folyamatos romlás eredményeképpen, fokozatosan alakul ki) hibákat lehet jósolni. A hirtelen bekövetkező, katasztrofikus hibák esetében a módszer használhatatlan: nincs miből jósolni.

5. Hibaok-meghatározás és javítás

A VoIP szolgáltatások hibamenedzsmentjéhez készült keretrendszer fejlesztése során a hibaok-analízis (Root Cause Analysis – RCA) újszerű megközelítését valósítottuk meg. A hagyományos hibamenedzsment-rendszerek a hibajelek passzív feldolgozására szorítkoznak. Az így kialakult hibajegyek közül az egyszerűbbek képesek voltak a hibaok leírására, az összetettebbek további feldolgozását azonban az operátorokra bízták, akik aktív ellenőrzéseket végeztek a felügyelendő rendszeren, így kutatva az eredendő hibaok után.

A *passzív* RCA rendszerek tehát csak a hibajeleken dolgozva próbálják megkeresni a hibaokot, az *aktív* módszer pedig az, hogy folyamatosan ellenőrzéseket végzünk. Mi ötvöztük a topológia változását rugalmasan követő, modell alapú, passzív esemény-korrelációs módszereket az operátori munkában mindennapos aktív ellenőrző lépések automatizált ütemezésével, így komplex, passzív és aktív (ellenőrző / beavatkozó) algoritmusokat is használó hibamenedzsment-rendszert kaptunk.

A kizárólag passzív esemény-korrelációval operáló, szabály alapú rendszerek egyik alap gondolata a következő: megfelelően összetett szabályrendszer alkalmazásával elérhető, hogy a hibafeldolgozás kimenete ne csak a specifikus hibajegy legyen, hanem paraméterként tartalmazza a hiba helyét, és a valószínűsíthető okot. Ehhez összetett, és egyben paraméterezett korrelációs szabályokra van szükségünk. Az ilyen hibajegykből származó paramétereket a hibaok-meghatározó rendszer kimeneteinek is tekinthetjük.

Ez a módszer azonban kizárólag passzív hibakorreláció esetében használható hibaok-meghatározásra. A passzív szabályrendszer legnagyobb hátránya, hogy nem ad lehetőséget a hálózati csomópontok aktív lekérdezésére – így (kulcsfontosságú információk hiányá-

ban) nem tudjuk az összes hibajegy RCA elemzését automatikusan elvégezni. Az ilyen esetek megoldásához algoritmikus hibaok-meghatározásra van szükség.

5.1. Algoritmikus hibaok-meghatározás

A hibaok-analízis során a hálózathoz kinyerhető információk (topológiai, eszköz adatok) felhasználásával határozzuk meg a legvalószínűbb hibát. A folyamat eredménye a hibajavítási javaslatlét, melynek figyelembe vételével nagyban leegyszerűsíthető a hálózatot üzemeltető operátor munkája.

A VoIP-szolgáltatások felügyeletére kialakított RCA-keretrendszerünkkel szemben támasztott követelmények a következők voltak. A rendszer képes legyen

- különböző esemény-adatbázisokban való keresésre,
- aktív ellenőrzések beiktatására, ennek ütemezésére,
- az ellenőrzések és keresésének párhuzamos futtatásának kezelésére,
- kezelni a topológiai információ dinamikus változását.

Emellett szem előtt tartottuk, hogy a rendszer modulárisan fejleszthető legyen, és a fejlesztés viszonylag gyorsan eredményt hozzon.

A hibaok-meghatározási módszereket három nagy csoportba rendezhetjük: *hibakorrelációs módszereken alapuló, statisztikai alapon működő* valamint *modell alapú rendszerek*. Az aktív lekérdezések indítását, majd ezek eredményeinek kiértékelését is támogató módszerünket egy modell alapú RCA-módszerrel ötvöztük.

A modell alapú RCA rendszerek jellemzően egy rugalmas modell segítségével írják le a hálózati topológiát. Új elem bekerülése esetén a korrelációs szabályrendszer generálható, és nem kell manuálisan újraírni. A topológiai leírás (VoIP hálózatot feltételezve) tartalmazza a fizikai csomópontokhoz rendelt saját IP-címet és az adott csomópontokhoz egy hop-on belül csatlakozó csomópontok IP-címeit. A szabályrendszer modelljének kialakításánál a topológiai információt rugalmasan kell kezelni (fel kell készíteni a topológia változására). A modell a hibajelek egymáshoz való viszonyát hierarchikusan írja le.

5.2. Petri-hálóval vezérelt ellenőrző lépések

A hibajelek szűrésével és korrelálásával a hiba jellegét (és esetleg helyét) aránylag jól leíró hibajegy(ke)t kapunk. Ezekben a hibajegyekben változóként szerepel a hibajegy azonosítója, típusa, és szerencsés esetben – további paraméterek mellett – a hibát okozó eszközök halmazszerű leírása. A hibajegyek kiértékelésének célja a hibaok(ok) meghatározása, majd javaslatlét a hiba javítására. Ez a folyamat a 4. ábrán követhető végig (a következő oldalon).

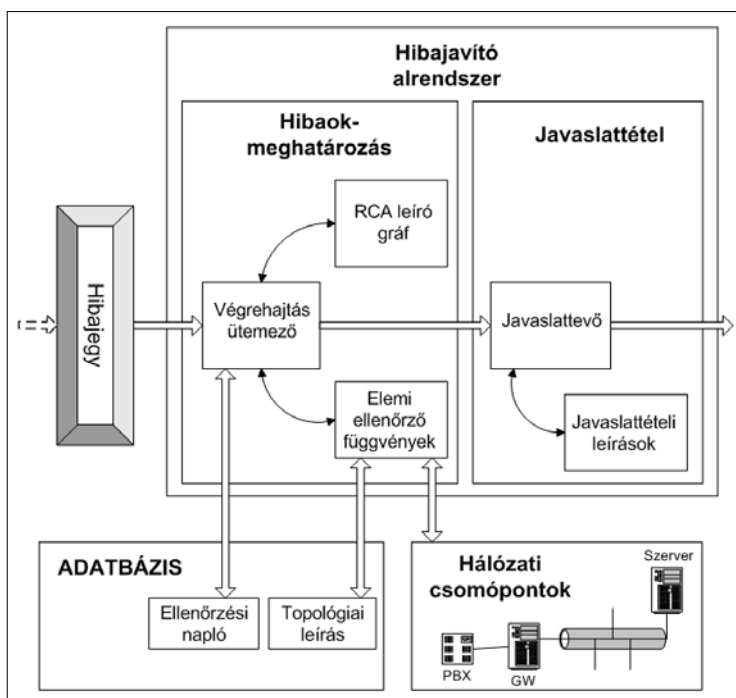
A hibajegyek a hibafeldolgozó alrendszerből érkeznek a hibajavító alrendszerbe. Minden hibajegy-típushoz tartozik egy Petri-háló formájú *RCA-leíró gráf*. Ez grafikusán ábrázolja az hibaok kiértékeléshez szükséges lépések (*elemi függvények*) kapcsolatát: mely adatok rendelkezésre állása esetén mely lépéseket lehet

végrehajtani (erre mutat példát a később részletezendő 5. ábra). Az aktuális ütemben végrehajtható ellenőrző lépések (elemi függvények) indítását a *végrehajtás ütemező* végzi. Amint minden ellenőrző lépés eredményét megkapta, és nincs több végrehajtandó lépés, a hibaok-meghatározási folyamat eredményét a javaslattevő felé továbbítja. Ez a végeredményhez rendeli a megfelelő javaslattevői leírást, és az operátor rendelkezésére bocsátja azt.

A fenti folyamat központi eleme az RCA-leíró gráf. Ezt a hibamenedzsmentben újszerű módszerrel, egy Petri-hálós leírással valósítottuk meg [9]. A megoldás előnye, hogy az aktív ellenőrzési feladatok (a Petri-háló „átmenetei”) konkurens módon, az adatok (a Petri-háló csomópontjai) rendelkezési állásának ütemében hajtódnak végre. Ha egy adat(halmaz) rendelkezésre áll, az ezt reprezentáló csomópontba jelzés, „token” kerül. Egy adott ellenőrzés akkor kerül végrehajtásra, ha az összes bemeneti adata rendelkezésre áll (minden bemeneti csomópontjában van „token”). Az ellenőrzés végrehajtásának eredményeképpen előállnak a kimeneti adatok, azaz a kimeneti csomópont(ok) „token”-nel lesznek ellátva. A gráfba külön adatlekérdező lépéseket lehet bevezetni annak érdekében, hogy egy adott ellenőrző-függvény összes bemeneti változója rendelkezésre álljon.

Az 5. ábra segítségével tekintsük át a „magas hálózati csomagvesztés” hibajegyhez tartozó RCA-leíró Petri-háló alkalmazásának jellegzetességeit. Abban a pillanatban, amikor a hibajavító alrendszerhez megérkezik a hibajegy, a végrehajtás ütemező aktiválja a jegyhez tartozó RCA-gráfot. Az első ütemben mindazon „átmenetek” (jelen esetben konkrét elemi ellenőrző függvények) „tűzelnek”, amelyek minden bemenetén van a kiindulási állapotban „token”.

4. ábra Hibajavító alrendszer architektúráis felépítése



Ezek az elemi ellenőrző függvények a VoIP szolgáltatásokhoz kapcsolható hibamenedzsment feladatoknak megfelelően a következő csoportokba sorolhatók:

- interfész-állapotot lekérdező függvények,
- konfigurációval kapcsolatos lekérdezések,
- aktív tesztek,
- egyéb (például hasonló hibajegy keresése).

Amikor egy adott elemi függvény kiértékelésre kerül, a „token” megjelenik a kimenetén (az ehhez tartozó adat más elemi függvények bemenete lehet). A végrehajtás ütemező minden ütemben megvizsgálja, hogy van-e az RCA-gráf átmenetei (elemi függvényei) között olyan, amelynek minden bemenetén van „token”, s ha talál ilyet, elindítja az adott elemi függvény kiértékelését. Ha az utolsó (tipikusan a végső kiértékelést végző) elemi függvény kimenetén is megjelenik a „token”, a végrehajtás ütemező átadja az eredményeket a javaslattevőnek, és megszünteti az RCA-entitást.

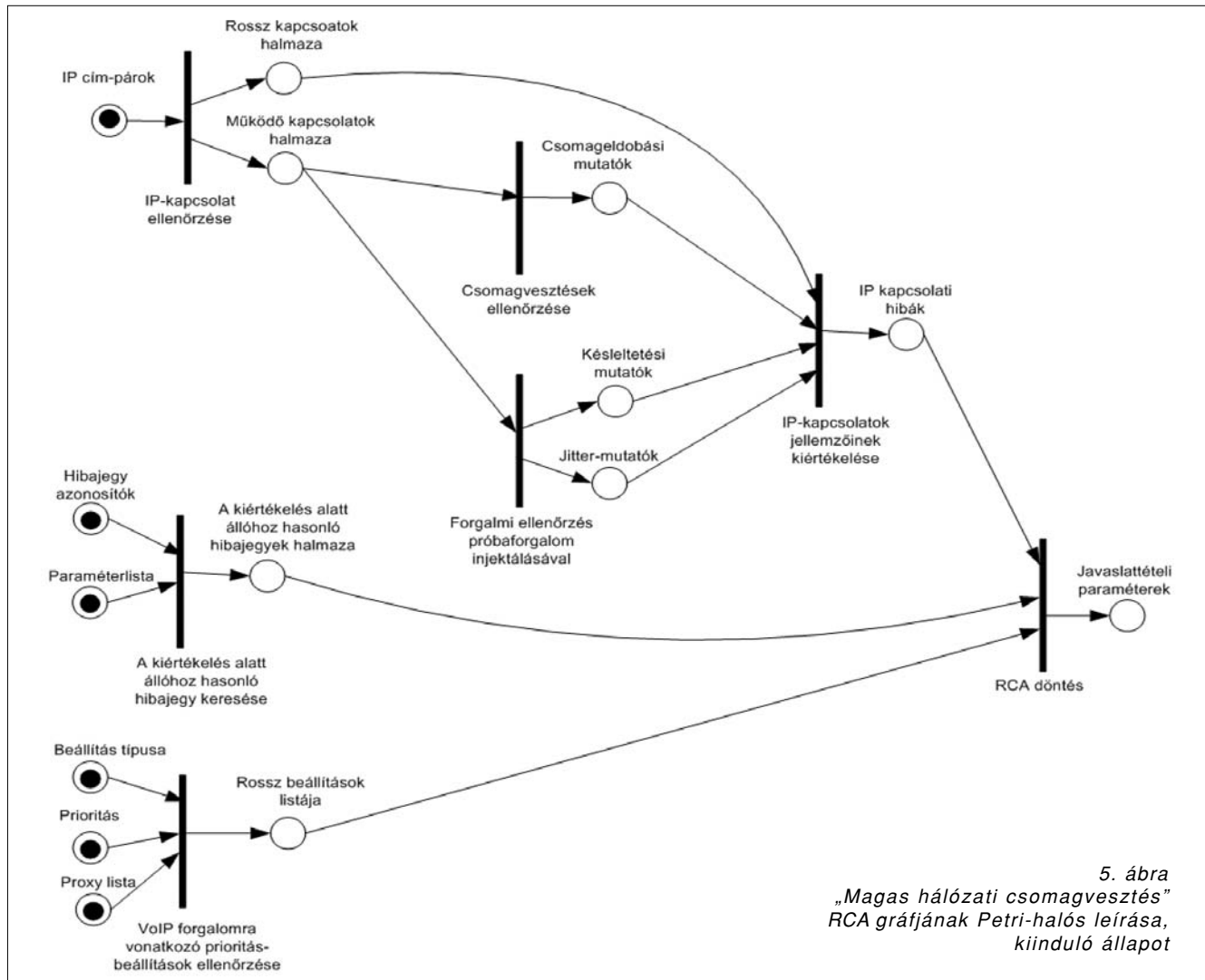
A Petri-hálós RCA-leírásnak köszönhetően egyes RCA-folyamatok végrehajtási ideje akár felére is csökkenthető [8], ráadásul az ellenőrzések végrehajtásához nem kell előre meghatározni a végrehajtási sorrendet, csak azt, hogy az adott RCA kiértékeléséhez milyen feladatokat kell végrehajtani. A feladatok a Petri-háló összeköttetéseit követve az adatok rendelkezésre állásának ütemében hajtódnak majd végre.

6. Összefoglalás

A szolgáltatást felügyelő operátor munkáját nagyban megnehezíti a kiterjedt VoIP-hálózatban hibajelek megjelenésének követhetetlen üteme, a hibajelek mennyisége és nehezen strukturálható megjelenése. Megfelelően kialakított szűrőszabályokkal megakadályozhatjuk, hogy a VoIP hibamenedzsment-rendszert feleslegesen árasssák el (azonos típusú) hibajegyek. A beérkező hibajeleket időlegesen teljesen elnyomhatjuk, számukat korlátozhatjuk, prioritásokat határozhatunk meg közöttük.

Ha a hibajeleket megfelelően csoportosítjuk és szűrjük, lehetőségünk nyílik a hibát átfogóan leíró hibajegyek összeállításáig. Ha csak ezt tárjuk az operátor elé, akkor nagyban megkönnyítjük a munkáját. Megfelelő „tudás” birtokában trendanalízis segítségével képesek vagyunk bizonyos hibákat előre is jelezni. A művelet során mintaillesztéssel kutatunk a hiba-adatbázisban – sikeres találat esetén komolyabb hibák előrejelzésére is lehetőségünk nyílik.

A szűrő-, korrelátor- és trendanalízis-szabályok alkalmazásával az eredendő hibaok nem mindig mutatkozik meg. A hibamenedzsment rendszernek ilyenkor el kell indítania egy hibaok-analizáló folyamatot. Az RCA során a hibamenedzsment rendszer aktív lekérdezésekkel ellenőrizheti az esetleges hibaforrásokat. Az eredmények kiértékelése után javaslatot tesz a hibákra és helyére a hálózatban, majd a javításra.



5. ábra
 „Magas hálózati csomagvesztés”
 RCA gráfjának Petri-hálós leírása,
 kiinduló állapot

Az ellenőrzések sorrendezésére és kiértékelésére több módszer létezik. Az IKTA-00092-2002 számú OM-projekt keretében a NIIFI szakembereinek bevonásával kifejlesztettünk egy új, Petri-hálós leírás alapuló módszert. Ennek alkalmazásával az RCA ellenőrző lépései az adatok rendelkezésre állásának függvényében (párhuzamosan) kerülnek végrehajtásra, a folyamat leírása leegyszerűsödik, a végrehajtás felgyorsul.

Irodalom

- [1] Felügyeleti rendszer integrált hang-adat hálózatok számára (műszaki melléklet, I. és II. munkaszakaszok), BME-TMIT, Ericsson Magyarország Kft., Kovax'95 Kft., IKTA-00092-2002 beszámoló, 2003.
- [2] ETSI TS 101 329-5 V1.1.2; Telecommunications and Internet Protocol Harmonization Over Networks (TIPHON) Release 3; End-to-end Quality of Service in TIPHON systems; Part 5: Quality of Service (QoS) measurement methodologies – ETSI, 2000.
- [3] ITU-T Recommendation P.862; Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs – ITU-T, 2001.
- [4] RFC 2866 – RADIUS Accounting – C. Rigney, 2000.
- [5] RFC 2865 – Remote Authentication Dial In User Service (RADIUS) – C. Rigney, S. Willens, A. Rubens, W. Simpson, 06/2000.
- [6] A Model For Alarm Correlation in Telecommunications Networks – Dilmar Malheiros Meira, 1997.
- [7] Discovering Rules for Fault Management – Roy Sternitt, Eighth Annual IEEE International Conference and Workshop on the Engineering of Computer Based Systems (ECBS'01), 2001.
- [8] Hibamenedzsment VoIP hálózatokban – Szijártó Tamás, BME-VIK TDK konferencia, 2004.
- [9] Petri Net Theory and the Modeling of Systems – James Lyle Peterson, Prentice-Hall, 1981.
- [10] VoIP-szolgáltatások hibamenedzsmentje – Varga Pál, Moldován István, Molnár Gergely, Networkshop 2005., Szeged.
- [11] Recommendation M.3000 Series – TMN: Telecommunications Management Network, ITU-T, 1992-1997.
- [12] OSS Concept of the Hungarian Telecom, Reznák Roxán

IPV6 – THE PROTOCOL OF FUTURE INTERNET?

Keywords: Internet, IP, IPv4, IPv6

The hook that holds the Internet together is the network layer protocol called IP or Internet Protocol. Unlike most other network layer protocols, it was designed from the beginning with internetworking in mind. Its job is to provide a best-effort way to transport datagrams from source to destination in the network. The currently used version of IP is known as IPv4 which is an old and widely deployed protocol in the Internet. However, it seems that IPv4 cannot keep pace with the recent enormously growing number of devices connected to the Internet and the increasing routing complexity, even in spite of the continuously introduced add-on solutions to IPv4. What will happen with the Internet when some day IPv4 won't be able to accomplish its job anymore? The proposed solution is a new and more powerful version of IP, called IPv6, which is supposed to replace IPv4 soon. But is this idea realistic? If yes, when this switch-over will take?

SDL: A STANDARDISED FORMAL METHOD TO MODEL COMMUNICATION PROCESSES

Keywords: SDL, UML2.0, FMT, ITU-languages

The goal of the paper is the presentation of modeling capabilities of the formal methods Specification and Description Language (SDL). The main motivation of the paper is the continued convergence of UML since 1999 towards the SDL language developed in the telecommunications world. This convergence appears for the software engineers in standard UML2.0, which has completely integrated the implementation part of SDL to describe the dynamic behavior of the system. Although the software tools providers publish the description of domain-specific implementation properties of UML2.0, to better understand UML2.0, which presents a radical change compared with the old versions, it is useful to get acquainted with the capabilities of SDL of modeling the dynamic behavior.

DVB-T MEASUREMENT EXPERIENCES IN GYŐR AND BUDAPEST

Keywords: digital TV, DVB-T, OFDM, measurements

The terrestrial digital broadcasting of television programs has begun in Hungary as well, starting from Budapest and Kabhegy stations. This article presents the DVB-T reception measurements conducted at Széchenyi University in Győr and Telecommunications Polytechnic Puskás in Budapest, and outlines the resulting experiences.

AUDIO-VISUAL SPEECH DATABASE AND ITS APPLICATIONS

Keywords: speech synthesis, people with hearing impairment, speech-to-animation conversion

The paper presents an audio-visual database to be used for the development of a lip-motion animation program for people with hearing impairment. Design considerations are explained and the test series with deaf people which the design is based on.

MOBILE PHONE APPLICATIONS FOR DEAF PEOPLE

Keywords: mobile phone, Symbian OS, deaf people

In the paper, a system is described that facilitates speech recognition in cases when it is not possible to lip read. Our objective is to develop a portable device, preferably a mobile phone that produces the modeled mouth or head motions from the speech.

APPLICATION OF LEAST SQUARES SUPPORT VECTOR MACHINES FOR DATA MINING

Keywords: Support Vector Machines, time series prediction, function approximation

Data mining is a collection of methods and algorithms that can effectively be used to discover implicit, previously unknown, hidden trends, relationships, patterns in masses of data. Classical data mining uses many methods from different fields, like the foundations of linear algebra, graph theory, database theory, machine learning and artificial intelligence. In this paper we address the problem of black-box modeling, where the model is built based on the analysis of input-output data. These problems usually cannot be addressed analytically, so usually they require the use of soft computing methods. We focus on the use of a special type of Neural Network (NN), the least squares version of Support Vector Machines (SVMs), the LS-SVM. We also present an extension of this method, the LS²-SVM, which enables us to deal with large quantities of data. Using this method we present solutions for function approximation, time series analysis, and time series prediction.

BAYESIAN NETWORK BASED ANALYSIS OF STATISTICAL DATA AND SCIENTIFIC LITERATURE

Keywords: Bayesian inference and networks, learning

The probabilistic modeling of a high dimensional domain based on expert knowledge, scientific literature and statistical data is multi-phased, multi-level task. First, it includes the modeling of the joint distribution over the domain variables on numeric, qualitative and possibly causal levels. Second, it includes the acquisition and extraction of prior domain knowledge from experts, from the web and from free-text resources, then its combination with statistical data and usage of the result in a decision theoretic framework. We overview the Bayesian network representation and the Bayesian statistical framework, which are successfully applied tools for these challenges. The application is demonstrated in a biomedical field related to the diagnosis of ovarian cancer and to the general problem of integrated analysis of statistical data and scientific literature.

A COMPLETE FAULT MANAGEMENT SOLUTION FOR MAINTANANCE OF VOIP SERVICES

Keywords: VoIP, quality of service, fault management

To ensure quality of service in VoIP networks, an adequate management system and a specific fault management method is necessary. The paper presents a thorough fault management system which addresses both the error detection and the identification of fault sources.