

# Szolgáltatásminőségi paraméterek előrejelzésének javítása outlierek detektálásával és eltávolításával

KOVÁCS LÁSZLÓ, VASS DOROTTYA, VIDÁCS ATTILA

BME, Távközlési és Médiainformatikai Tanszék

[kovacs@alpha.tmit.bme.hu](mailto:kovacs@alpha.tmit.bme.hu), [vass@alpha.tmit.bme.hu](mailto:vass@alpha.tmit.bme.hu), [vidacs@tmit.bme.hu](mailto:vidacs@tmit.bme.hu)

**Kulcsszavak:** outlier detektáló algoritmusok, jellemzők számítása, szolgáltatásminőségi paraméterek előrejelzése

Az Interneten széles körben terjedő valós idejű alkalmazások szükségessé teszik, hogy a szolgáltatók megfelelő szolgáltatásminőségi paramétereket biztosítsanak a felhasználóknak. Ehhez szükséges a hálózati forgalom minél pontosabb előrejelzése. Ennek egy lehetséges megoldása a forgalom mérése, a mért adatok elemzése, majd ezt követően az adathalmazra megfelelő modell illesztése, végül ez alapján a kívánt paraméter meghatározása. Az előrejelzés pontosságát azonban nagymértékben rontják az adatsorban található kiugró minták (outlierek). Kifejlesztettünk egy eszközt, amely gyorsan és automatikusan képes detektálni és eltávolítani az outliereket az idősorokból. Ezenfelül javasolunk egy, az ARIMA (autoregresszív integrált mozgóátlag) modellt használó előrejelzési eljárást is az előzetes outlier detektálással és eltávolítással kiegészítve. Megmutatjuk, hogy ez az eljárás jelentősen növeli az előrejelzés hatékonyságát.

## 1. Bevezetés

A végfelhasználók számára fontos, hogy a szolgáltatók megfelelő szolgáltatásminőségi paramétereket tudjanak garantálni. A szolgáltatóknak szükségük van a mért adatokon alapuló minél pontosabb előrejelzésre, hogy meg tudják tervezni az erőforrások, az alkalmazások és a felhasználók összerendelését. Az előrejelzésnek figyelembe kell vennie, hogy a legtöbb mért adatsor tartalmaz outliereket, melyek helyi véletlen események által okozott extrém ingadozások. Ezek speciális minták (kiugró értékek), melyek nem követik az adatok többségének jellegzetes eloszlását. Az outliereknek jelentős hatásuk lehet a statisztikai becslésre az adatok analízise és a modellezése során.

Az általunk javasolt eljárás kombinálja az előzetes outlier detektálást és eltávolítást a szolgáltatásminőségi paraméterek ARIMA [11] modellen alapuló előrejelzésével. Ezt a modellezési eljárást gyakran használják különböző kutatásokban a hálózati forgalom és a szolgáltatásminőségi paraméterek leírására, előrejelzésére.

Hasonló megközelítéssel találkozhatunk közgazdasági területen [1], amely alapján pénzügyi idősorok előrejelzésére egy ingyenes szoftver is született [2]. A szoftver közgazdasági jellegéből adódóan csak kis adathalmazok kezelésére alkalmas, a távközlési hálózatokban végzett mérések során azonban gyakran nagy adathalmazok keletkeznek. Ezért kifejlesztettünk egy új szoftvert, amely képes nagy adathalmazok kezelésére is. Az általunk kifejlesztett alkalmazás gyorsan távolítja el az idősorokból az outliereket, és az előrejelzés számára megfelelőbb adatokkal helyettesíti azokat. A kifejlesztett algoritmus az úgynevezett L.O.C.I. [3] algoritmuson alapul. A program képes az algoritmus paramétereinek automatikus meghatározására is.

A működés helyességét mesterséges és valós adatsorokon is ellenőriztük, és az eredmények azt mutatják,

hogy az outlierek detektálása és eltávolítása gyors és hatékony.

A cikk felépítése a következő: a második fejezet az outlierek detektálásáról szól. Az általunk használt algoritmust a harmadik fejezet tárgyalja. A negyedik fejezet a kifejlesztett eszközt mutatja be, és vizsgálja a működés helyességét is. Az ötödik fejezetben az előzetes outlier detekció és eltávolítás hatását vizsgáljuk az előrejelzés jóságára. Végül a hatodik fejezet összegzi a fentiek és a továbbfejlesztési lehetőségeket.

## 2. Outlierek detektálása idősorokban

D. Hawkins szerint [10]: „Az outlier az a megfigyelés, amely annyira eltér a többi megfigyeléstől, hogy azt a gyanút kelti, hogy valamilyen más mechanizmus hozta létre.” Az outlierek az idősorban található, az előrejelzés szempontjából értelmetlen adatok, melyek különböző véletlen hibák hatására keletkeznek. Az outlierek jelentősen befolyásolhatják az idősorra illesztett modell paramétereit.

A fellelhető outlier detektáló algoritmusok az alábbi öt kategóriába sorolhatók:

- **Eloszlás alapú:** Ezek az algoritmusok néhány alapvető eloszlásmodellt tartalmaznak (például normális eloszlás, Poisson eloszlás stb.) és azokat a mintákat jelölik meg outlierként, amelyek az adott eloszlásmodellhez nem illeszkednek [4]. Ezen eljárások legnagyobb hátránya, hogy a mért adatok eloszlása nem mindig ismert. Gyakran sok és hosszadalmas vizsgálatot kell végrehajtanunk ahhoz, hogy megállapítsuk, melyik modell illeszkedik legjobban az adathalmazunkra. (Ezen túlmenően még az sem bizonyos, hogy egyáltalán létezik-e olyan modell, amelyre a mért adatok illeszkednek.)

- **Mélység alapú:** Ezek a geometrián alapuló algoritmusok az adatokat különböző héjakba sorolják, és a

legkülső héjon elhelyezkedő adatokat jelölik meg outlierként [5]. Az algoritmus sebessége a rétegszámmal fordítottan arányos.

- **Csoportosításon alapuló:** Ezen algoritmusok a bemenetként kapott adatokat valamilyen szempont szerint csoportosítják, osztályokba sorolják [6]. Outliernek tekintik azokat az elemeket, amelyek egymagukban alkotnak egyetlen csoportot. Mivel nem az outlierek meghatározása a fő feladatuk, így ezt a problémát nem képesek hatékonyan megoldani.

- **Távolság alapú:** Ezt az algoritmust először Knorr és Ng írta le [7,8]. Ennek alapján azt mondjuk, hogy egy  $P$  adathalmazban lévő  $x$  érték outlier, ha a  $P$ -ben lévő adatok legalább  $b\%$ -a adott távolságnál ( $r$ ) messzebb helyezkedik el a vizsgált  $x$  értéktől. Ez egy egyszerű, két paramétertől ( $r, b$ ) függő eljárás. Problémát az okozhat, ha az adatok jellemzői az adathalmaz különböző részein jelentősen eltérnek egymástól.

- **Sűrűség alapú:** Ezen algoritmust elsőként Breunig adta meg [9]. Ez az eljárás minden értékhez egy helyi outlier tényezőt (Local Outlier Factor, LOF) rendel hozzá, amelyet a szomszédok sűrűsége határoz meg. (Minél magasabb a LOF értéke, az adott pontnak annál kevesebb szomszédja van.) A magas LOF értékkel rendelkező adatokat outlierként jelöli meg. Ezen megoldás hátránya, hogy nagyon érzékeny a szomszédságot meghatározó paraméterekre.

### 3. Az algoritmus

Az általunk kidolgozott algoritmus az úgynevezett L.O.C.I. eljárás [3] alapul. A távolság alapú és a sűrűség alapú algoritmusok előnyeit kombináltuk. Az aktuális mintának csak egy kis szomszédságát vizsgáljuk, így nem okoz problémát, ha az adatok jellemzői jelentősen eltérnek egymástól az adathalmaz különböző részein. Ezen a szomszédságon belül az ebben található pontok statisztikai jellemzői alapján döntjük el, hogy a vizsgált pont outlier-e. Az algoritmusunk két különböző szomszédsággal dolgozik.

A mintavételezési szomszédság vagy más néven  $r$ -szomszédság (1. ábra) egy  $2r$  széles ( $r$  sugarú) intervallum. Az ábrán az aktuálisan vizsgált pontot vastag  $x$ -el jelöltük. A szaggatott vonal a vizsgált ponthoz tartozó mintavételezési tartomány határa. Ezen tartomány minden pontjára definiálunk egy kisebb szomszédságot, ez a számlálási szomszédság. Ezt néhány pontra pontozott vonallal jelöltük az ábrán. A számlálási szomszédság két paraméterrel jellemezhető:  $\alpha, \beta$ . E tartományban vizsgáljuk a szomszédok sűrűségét. Ha az aktuálisan vizsgált pont szomszédsűrűsége jelentősen eltér a mintavételezési tartományban található pontok átlagos szomszédsűrűségétől, akkor a pontot outlierként jelöljük meg. A szomszédsűrűséget alapvetően a számlálási szomszédságban található pontok száma határozza meg. Eddig az algoritmusunknak három fő paramétere van. Az  $r$  paraméter a mintavételezési szomszédság sugara, az  $\alpha$  és  $\beta$  paraméterek pedig a

számlálási szomszédság definiálásához szükségesek. Ezekon kívül még egy paraméterre van szükségünk ( $k$ ), ami a detektálás „szigorúságát” határozza meg. Az 1. táblázat a  $k$  paraméter definiálásához szükséges alapdefiníciókat tartalmazza.

Minden  $p_i, r, \alpha$  és  $\beta$  értékre definiáljuk a szórási tényezőt (DF) az alábbiak szerint:

$$DF(p_i, r, \alpha, \beta) = \frac{\hat{n}(p_i, r, \alpha, \beta) - n(p_i, \alpha, \beta)}{\hat{n}(p_i, r, \alpha, \beta)}$$

Megjegyezzük, hogy  $p_i$   $r$ -szomszédsága mindig tartalmazza  $p_i$ -t. Így a fenti tört nevezője mindig nagyobb nullánál, azaz a fenti kifejezés mindig értelmezett.

Definiáljuk  $\sigma_{DF}(p_i, r, \alpha, \beta)$ -et  $n(p, \alpha, \beta)$  normalizált szórásként a  $p \in N(p_i, r)$ -n:

$$\sigma_{DF}(p_i, r, \alpha, \beta) = \frac{\sigma_{\hat{n}}(p_i, r, \alpha, \beta)}{\hat{n}(p_i, r, \alpha, \beta)}$$

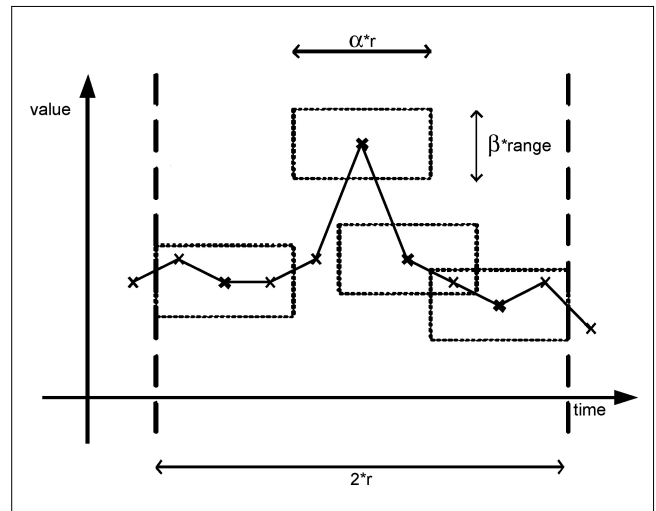
A pontot outlierként jelöljük meg, ha:

$$DF(p_i, r, \alpha, \beta) \geq k \cdot \sigma_{DF}(p_i, r, \alpha, \beta)$$

Tehát, ahogy már fent említettük, a  $k$  paraméter a szűrés „szigorúságát” határozza meg. A  $k$  paraméter értékét három körülinek választva és azt feltételezve, hogy a minták eloszlása normális, annak az esélye, hogy egy minta (amelyik nem outlier) kívül esik a háromszoros szóráson (és így outliernek tekintjük) kevesebb mint 1%. Automatikus paraméterbeállítások esetén  $k$  értéke 2.8. Ha az  $r$  paraméter értéke túl kicsi, nincs elég adatunk a szórást pontos megállapításához. Ha ez az érték túl nagy, az az eltérő helyi jellegekből fakadó hibás detekcióhoz vezethet. Alabeállítások esetén  $r$  értéke a minták számának 10%-a, de maximum 50.

A következő fejezetekben ismertetett eredmények során az  $\alpha$  és  $\beta$  paraméterek értéke megegyezett. Ha ezen paraméterértékek túl kicsik, előfordulhat, hogy helyes adatoknak sem lesz szomszédjuk, míg ha túl nagyok, majdnem minden minta beletartozik a számlálási szomszédságba, így az outlier nem esik kívül a  $k$ -szoros szóráson.

1. ábra Szomszédságok



<b>Jelölések és definíciók</b>	$P, p_i$	Az adathalmaz: $P = (p_1, \dots, p_i, \dots, p_N)$
	$N(p_i, r)$	$p_i$ r-szomszédjainak halmaza: $N(p_i, r) = \{p_x \in P \mid (i-r) \leq x \leq (i+r)\}$
	$n(p_i, r)$	r-szomszédok száma: $n(p_i, r) \equiv  N(p_i, r) $
	$n(p, \alpha, \beta)$	A p pont számlálási szomszédságába tartozó elemek száma
	$\hat{n}(p_i, r, \alpha, \beta)$	$n(p, \alpha, \beta)$ átlaga $p_i$ r-szomszédjain: $\hat{n}(p_i, r, \alpha, \beta) = \frac{\sum_{p \in N(p_i, r)} n(p, \alpha, \beta)}{n(p_i, r)}$
	$\sigma_{DF}(p_i, r, \alpha, \beta)$	$n(p, \alpha, \beta)$ szórása $p_i$ r-szomszédjainak halmazán: $\sigma_{\hat{n}}(p_i, r, \alpha, \beta) = \sqrt{\frac{\sum_{p \in N(p_i, r)} (n(p, \alpha, \beta) - \hat{n}(p_i, r, \alpha, \beta))^2}{n(p_i, r)}}$
	$DF(p_i, r, \alpha, \beta)$	Szórási tényező a pi pontra. Részletesen lásd lentebb.
	$\sigma_{\hat{n}}(p_i, r, \alpha, \beta)$	Normalizált szórás (hogy össze lehessen hasonlítani DF-el). Részletesen lásd lentebb.

1. táblázat

Alapbeállítások estén az a és b értéke 0.1 és 0.05 közé esik, az r paraméter értékétől függően. Az automatikus paraméterbeállítások még nem optimálisak, de már jó eredmények érhetőek el velük (lásd alább). Gyorsan változó idősorok esetén problémák jelentkezhetnek, ennek vizsgálata további terveink között szerepel.

#### 4. Outlier detektáló és eltávolító eszköz (ODRT Tool)

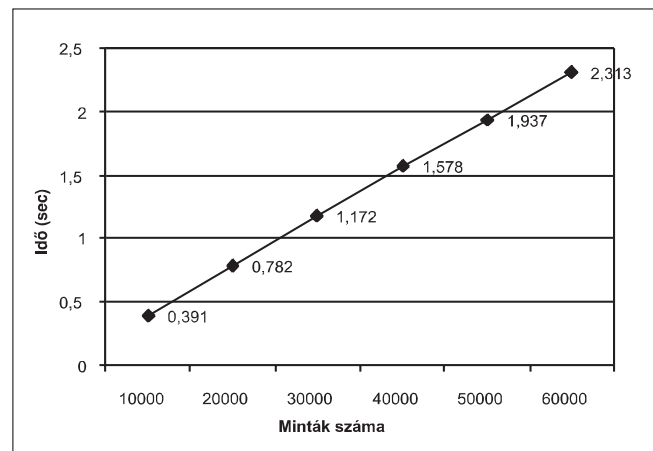
Kifejlesztettünk egy új szoftvert (ODRT, Outlier Detection and Remove Tool), amely gyorsan és hatékonyan képes kezelni nagy adathalmazokat is. Az alkalmazás jelenlegi megvalósítása Microsoft Windows alatt fut. A program automatikusan meg tudja határozni a paramétereket, így felhasználói beavatkozás nélkül is képes adathalmazokat kezelni. Az algoritmus helyes működésének ellenőrzésére számos mérést végeztünk.

A tesztek során az automatikus paraméterbeállításokat használtuk. Számos különböző paraméterű ARLMA folyamatot használtunk kiindulásként, majd ezekhez az idősorokhoz adtunk hozzá additív outliereket. Így meg tudtuk vizsgálni, hogy különböző esetekben mennyire hatékony az algoritmus, azaz a hozzáadott outlierek hány százalékát ismeri fel. Az ellenőrzés során vizsgáltuk az algoritmus sebességét (2. ábra), az outlierek számának hatását a detektálásra, az outlierek eloszlásának hatását a detektálásra, és néhány speciális esetet is (például: negatív outlierok, szinteltolódásos outlierok stb.).

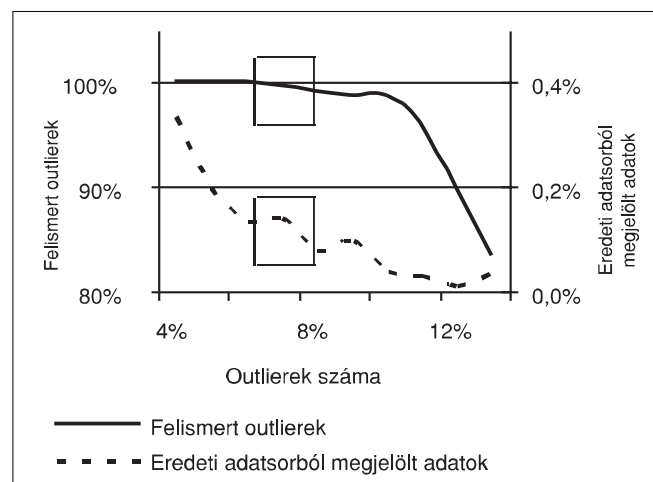
Az algoritmus sebességét automatikus paraméterbeállítások esetén vizsgáltuk. Ha a minták száma több mint 500, a paraméterek fix értékek, így a függvény lineáris. (A mérés során használt hardver: Intel P4 2,4Ghz processzor, ABIT BD7II alaplap, 256 MB RAM).

Amikor az outlierok számának a detektálás jóságára gyakorolt hatását vizsgáltuk, a hozzáadott outlierok száma a minták számának 4-13%-a volt. 200 mérést végeztünk, az eredményt a 3. ábra mutatja.

2. ábra Az algoritmus sebessége



3. ábra Az outlierok számának hatása



A folytonos vonal, és a baloldali ordináta tengely a felismert outliereket jelöli, míg a jobboldali tengely és a szaggatott vonal az eredeti adatsorban outlierként megjelölt értékeket. Az eredmények alapján megállapíthatjuk, hogy az 5-10%-os intervallumban az algoritmus a hozzáadott outlierok közel 100%-át felismeri, míg az eredeti adatsorból kevesebb mint 0.2%-ot jelöl meg outlierként.

### 5. Szolgáltatásminőségi (QoS) paraméterek előrejelzése

A széles körben terjedő valós idejű alkalmazások szükségessé teszik, hogy a szolgáltatók megfelelő szolgáltatás-minőségi paramétereket biztosítsanak a felhasználóknak. Az Internet-szolgáltatók számára fontos, hogy a szolgáltatásminőségi paramétereket valós méréseken alapulva és minél pontosabban tudják előrejelezni, hogy tervezni tudják az erőforrások és az alkalmazások, továbbá a felhasználók összerendelését.

#### 5.1. Előrejelzés az ARIMA modell alkalmazásával

Az ARIMA modell segítséget nyújt olyan problémák megoldásában, mint az idősorok előrejelzése és spektrális analízise. Box és Jenkins 1976-ban átfogó leírását adta az ARIMA modellnek [11], és napjainkban is számos szakirodalom tárgyalja (például [13,14,12]). Azt mondjuk, hogy a  $p$ -edrendű  $x_t$  folyamat autoregresszív, ha felírható a következő alakban:

$$x_t = c_1x_{t-1} + c_2x_{t-2} + \dots + c_px_{t-p} + \varepsilon_t$$

ahol  $\varepsilon_t$  egy  $\sigma$  szórású fehér zaj,  $c_1, \dots, c_p$  konstans.

Ekkor az idősor értékeit a saját korábbi értékeivel, és a véletlen ingadozást reprezentáló  $\varepsilon_t$  diszkrét fehér zajjal fejezzük ki.

A  $q$ -adrendű  $x_t$  folyamat mozgó átlag folyamat, ha felírható következő alakban:

$$x_t = d_1\varepsilon_{t-1} + d_2\varepsilon_{t-2} + \dots + d_q\varepsilon_{t-q} + \varepsilon_t$$

ahol  $d_1, \dots, d_q$  konstans.

A kombinált ARMA-modell autoregresszív és mozgóátlagolás tagokat is tartalmaznak, tehát a következő alakban írható fel:

$$x_t = c_1x_{t-1} + c_2x_{t-2} \dots c_px_{t-p} + d_1\varepsilon_{t-1} + d_2\varepsilon_{t-2} \dots d_q\varepsilon_{t-q} + \varepsilon_t$$

Az idősor gyakran magában hordozza a környezet változásának hatását. Ebben az esetben a megfigyelt értékek közti különbség lesz stacionárius, tehát az eddig leírt modellekkel jellemezhető.

Az ARIMA folyamatok az ARMA folyamatoknál általánosabb modellt írnak le. Ha egy ARIMA( $p,d,q$ ) folya-

matról beszélünk, akkor az azt jelenti, hogy az idősort  $d$ -szer differenciálva egy ARMA( $p,q$ ) folyamatot kapunk. (Az integrált modellek esetében tapasztalataink szerint a gyakorlatban általában elegendő egyszer differenciálni a sort ahhoz, hogy stacionárius modell legyen illeszthető rá.) Az autoregresszív paraméterértékeket a Yule–Walker egyenletrendszer [15] alapján számítottuk ki. A modellt az S-PLUS 2000 matematikai szoftver segítségével határoztuk meg [16].

#### 5.2. QoS paraméterek előrejelzése az outlierok előzetes eltávolításával

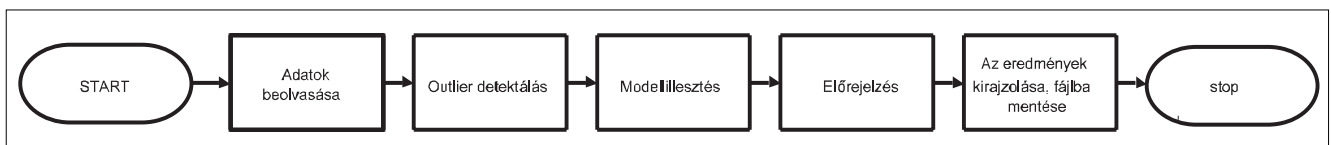
Az előrejelzésnek figyelembe kell vennie, hogy a legtöbb valós adatsor tartalmaz outliereket, amelyek a helyi véletlenszerű események hatására bekövetkező extrém ingadozások. Az outlierok detektálásával kiküszöbölhetjük ezen események hatását az előrejelzésre. A két eljárás, azaz az előrejelzés és az outlier detektálás együttműködését az 2. ábra szemlélteti.

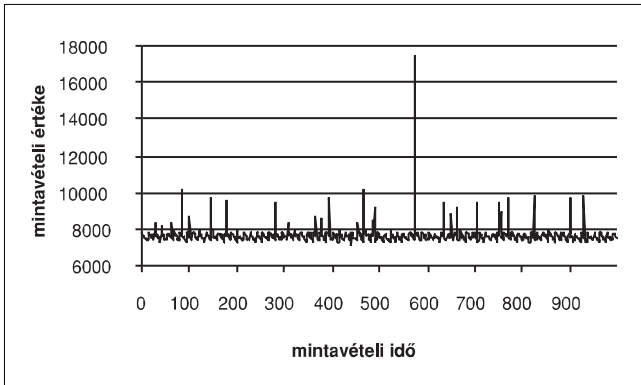
A következőkben az outlierok előzetes eltávolításának hatására az előrejelzésben bekövetkezett javulást tanulmányozzuk. Vizsgálatainkat valós, mért adatsorokon végeztük. Az adatok egy része a Salzburg Research-től származik. A mérések során a Salzburg Research hálózata és a szolgáltató (Telekom Austria) közötti végpont-végpont összeköttetések monitorozták az aggregált forgalom késleltetését. A késleltetés-mérés a GPS óraszinkronizáción alapult. A küldő és a fogadó fél közötti forgalom két útvonalválasztón haladt keresztül. Ez a hálózati kialakítás a mai szolgáltatók és felhasználók között tipikusnak tekinthető [18]. A kutatóintézet több százezer adatból álló idősort bocsátott a rendelkezésünkre, ebből használtunk fel részleteket a teljesítményelemzéshez. Az adatok másik része saját méréseken alapul, melyet a tanszék két gépe között végeztünk. Az előrejelzés és a modellillesztés során automatikus paraméterbeállításokat használtunk.

A teljesítményelemzés során 1000 adatból álló idősorokat vizsgáltunk. Az első 95% alapján jeleztük előre a következő 50 adatot. Az egyik előrejelzés során az outliereket előzetesen eltávolítottuk, a másik esetben ezt nem tettük meg. A modell rendjének meghatározásakor a rendre viszonylag magas (25-35) értékeket kaptunk. A mérések kiértékeléséhez az eredeti adatsort, az outlierok eltávolítása után és az előzetes outlier detekció nélkül kapott eredményeket egyetlen grafikonon tüntettük fel.

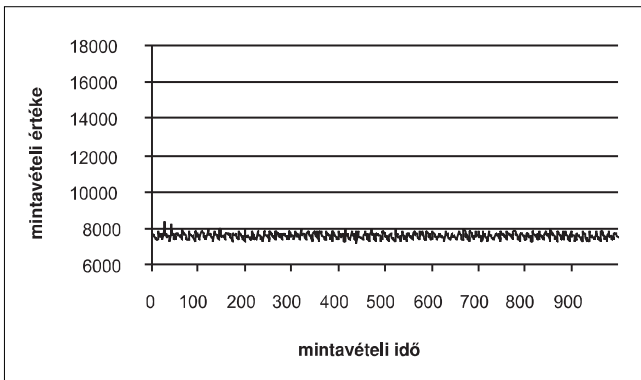
Először egy olyan esetet mutatunk be, amikor a vizsgált szakasz kihasználtsága valószínűleg alacsony, ezért a késleltetés kis tartományon belül ingadozik, de az adatsorban néha találhatók outlierok. Az eredeti és a szűrt idősort szemlélteti a 3/a. és 3/b. ábra.

2. ábra A két eljárás együttműködése





3/a. ábra Az eredeti adatsor

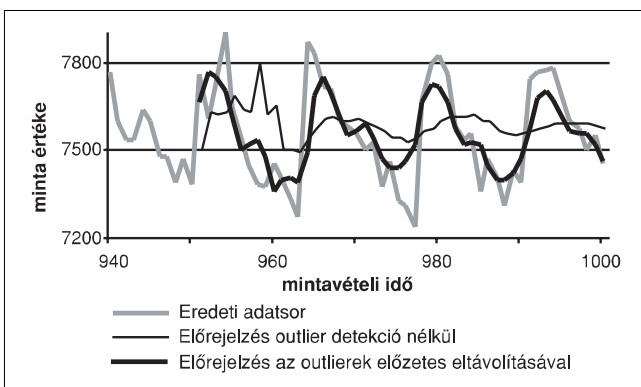


3/b. ábra A szűrt adatsor

A 4. ábrán láthatjuk az eredeti adatsort (vastag szürke vonal), az outlier detektálás nélküli előrejelzést (vékony fekete vonal) és az outlierek előzetes eltávolításával kapott előrejelzést (vastag fekete vonal). Jól látható, hogy a becült adatok az előzetes outlier detekció alkalmazásával még 50 adat után is jól követik az eredeti értékeket. Ha nem távolítjuk el az outliereket, a jóslás nem képes követni az eredeti adatsor ingadozását, ellaposodik, egy idő után konstans értéket vesz fel.

A második esetben a késleltetés értékei nagyobb intervallumban ingadoznak. Az 5. ábrán láthatjuk az eredeti adatsort (vastag szürke vonal), az outlier detektálás nélküli előrejelzést (vékony fekete vonal) és az outlierek előzetes eltávolításával kapott előrejelzést (vastag fekete vonal). Az ábra azt mutatja, hogy előze-

4. ábra Az outlier detekció hatása (1)



5. ábra Az outlier detekció hatása (2)

tes outlier detekció esetén az előrejelzés jól követi az eredeti adatsor főbb mozgásait, míg outlier eltávolítás nélkül a predikció ellaposodik, képtelen követni az adatsor változásait.

## 6. Összegzés és további teendők

A felhasználói igények kielégítéséhez valamint a hálózat kihasználtságának növeléséhez szükséges a hálózati forgalom és a szolgáltatásminőségi paraméterek minél pontosabb előrejelzése. Az előrejelzés pontosságát azonban nagymértékben rontják a mért adatsorokban található kiugró értékek, az outlierok. Egy olyan új eszközt mutattunk be, amely képes detektálni és eltávolítani az outliereket nagy méretű adatsorokból is gyorsan és hatékonyan. A modellezéshez és előrejelzéshez egy ARIMA folyamatokon alapuló módszert adtunk. Az eredmények alapján látható, hogy az előzetes outlier detekció hatására jelentősen javul az előrejelzés, ezért ez a módszer sokat segíthet a garantált szolgáltatásminőség biztosításában.

Miután eltávolítottuk az outliereket, a helyettesítési értéket a szomszédos adatok átlaga alapján határoztuk meg. Valószínűleg tovább javítana az eredményeken, ha a helyettesítési értéket valamilyen módon a megelőző értékek alapján előrejeleznénk (például lineáris predikcióval). Ennek vizsgálata további terveink között szerepel.

## Irodalom

- [1] V. Gómez and A. Maravall, „Programs TRAMO (Time series Regression with Arima noise, Missing observations, and Outliers) and SEATS (Signal Extraction in Arima Time Series). Instructions for the User”, Working Paper 9628, Servicio de Estudios, Banco de España, 1996.
- [2] A. Maravall and G. Caporello, „A tool for Quality control of time series data, Program TERROR”, Bank of Spain, Proc., Challenges to Central Bank Statistical Activities Conf., Irving Fisher Committee (ISI) and Bank for International Settlements, Basel, August 2002

- [3] S. Papadimitriou, H. Kitawaga, P. B. Gibbons, C. Faloutsos, „Fast Outlier Detection Using the Local Correlation Integral”,  
[http://www.intel-research.net/Publications/Pittsburgh/081620021325\\_99.pdf](http://www.intel-research.net/Publications/Pittsburgh/081620021325_99.pdf)
- [4] P. J. Rousseeuw and A. M. Leroy, „Robust Regression and Outlier Detection”, John Wiley and Sons, 1987.
- [5] T. Johnson, I. Kwok, and R. T. Ng, „Fast computation of 2-dimensional depth contours”
- [6] A. K. Jain, M. N. Murty, and P. J. Flynn, „Data clustering: A review”, ACM Computing Surveys, 31(3):264–323, 1999.
- [7] E. M. Knorr and R. T. Ng, „Algorithms for mining distance-based outliers in large datasets”, 1998;  
<http://www.courses.cs.uiuc.edu/~cs497jh/papers/knorrngvldb98.pdf>
- [8] E. M. Knorr, R. T. Ng, and V. Tucakov, „Distance-based outliers: Algorithms and applications”,  
[http://www.cs.ubc.ca/nest/dbsl/public/vldb\\_journal\\_feb2000.ps](http://www.cs.ubc.ca/nest/dbsl/public/vldb_journal_feb2000.ps)
- [9] M. M. Breunig, H. P. Kriegel, R. T. Ng, and J. Sander, „Lof: Identifying density-based local outliers”, 2000;  
<http://www.cs.ualberta.ca/~joerg/papers/LOF-final.pdf>
- [10] D. Hawkins, „Identification of outliers”, Chapman & Hall, London 1980.
- [11] G. Box, G. Jenkins, G. Reinsel, „Time Series Analysis: Forecasting and Control”, revised ed., Prentice Hall, 3rd Edition, 1994.
- [12] P. J. Bockwell, R. A. Davis, „Introduction to Time Series and Forecasting”, Springer Verlag, 2002.
- [13] Michelberger Pál, Szeidl László, Várlaki Péter, „Alkalmazott folyamatstatisztika és idősor analízis”, Typorex kiadó, 2001.
- [14] Vágó Zsuzsanna, „Idősorok sztochasztikus modelljei”, BKE oktatási segédlet, 1995.
- [15] P.J. Bockwell, R. A. Davis: Introduction to Time Series and Forecasting, Springer Verlag, 2002.
- [16] S-PLUS 2000, „Guide to Statistic”, Data Analysis Product Division Math Soft, 1999.
- [17] W. N. Venables, B. D. Ripley, „Modern Applied Statistics with S-PLUS”, Springer Verlag, 1998.
- [18] I. Milouceva, E. Müller and A. Anzalonli, „A practical approach to forecast QoS considering outliers”, Proc., Inter-Domain Performance and Simulation Workshop, pp.163–172, Salzburg, Austria, February 2003.

## Hírek

Az egyre növekvő **rádiófrekvenciás azonosítás (RFID)** számos új megoldást nyújtott a piac számára. A Sun, a SIS Technologies és az SSA Global termékeit és szolgáltatásait egyesítő Sun RFID-kezdményezéseit az EPCglobal konferencián is bemutatta.

A **Sun Java System RFID szoftver** (amely jelenleg Solaris operációs rendszeren érhető el, de a tervek szerint 2004 novemberében Linuxon is megjelenik) korszerűsíti az RFID-hardverek – a címkék és az olvasók – kezelését. A szoftver a Sun RFID rendszerstratégiájának fontos elemét jelenti, de ide tartoznak még a globális kliensszolgáltatások, a világ RFID tesztközpontjai, a Sun hardverek és tárolóeszközök, valamint az RFID-terület megoldási partnerei. A Sun Java RFID szoftverrel kapcsolatban további információk a [www.sun.com/rfid](http://www.sun.com/rfid) oldalon olvashatók.

A rádiófrekvenciás termékazonosítás (RFID) és az egyéb távérzékelő technológiák (SBS) kétségkívül versenyelőnyt és üzleti hasznot jelenthetnek, ha a vállalat információs infrastruktúrájának és üzleti folyamatainak szerves részévé válnak. Világszerte már számos gyártó döntött az RFID rendszerek bevezetése mellett. Sok vállalat előírta, hogy beszállítóik 2005-ig RFID jelzőcímkékkel lássák el raklapos szállítmányukat. Az IDC véleménye szerint a vállalatoknak nem szabad csupán azért bevezetni az új technológiát, mert a legnagyobb vevők ezt megkövetelik tőlük, hanem először meg kell érteniük az RFID-ban rejlő technológiai és hatékonyságnövelő képességeket, és ezek függvényében kell beépíteniük az új eszközt üzleti stratégiájukba.

A Meta Group piackutató cég szerint 2008-ra az iparcikkek 30%-át már RFID azonosítóval látják el, és ez az arány 2013-ra 80%-ra nőhet. Az amerikai Yankee Group is bízik a technológia hosszú távú sikerében: a vállalat szerint 2008-ra az RFID piaci értéke elérheti a 2,35 milliárd dollárt.