



„Podporujeme výskumné aktivity na Slovensku/Projekt je spolufinancovaný zo zdrojov EÚ

EXTRAPOLÁCIA NEDOSTATOČNÝCH DÁTOVÝCH VZORIEK.

*PREDPOKLADY A PODMIENKY, KTORÉ MUSIA BYŤ NAPLNENÉ,
METÓDY, NÁSTROJE EXTRAPOLÁCIE.*

Predmetom šiestej analýzy v rámci projektu je zhrnúť použitie extrapolačných štatistických techník v kontexte estimácie odchýlky výroby a spotreby elektrickej energie. Táto správa je len formálnym výstupom analýzy vhodnosti rôznych prístupov k procesu extrapolácie dostupných dát. Cieľom analýzy bolo popísať predpoklady a podmienky, ktoré musia byť naplnené pri extrapolácii dát v energetike, prínosy a riziká použitia extrapolačných techník, popísať typy dát použiteľných na extrapoláciu, uviesť príklady extrapolačných techník, popísať postup extrapolácie v časových krokoch a uviesť možnosti umelého rozšírenia vzorky v prípade nedostatočných vzoriek dát.

1. Extrapolácia

Extrapolácia je štatistická technika inferencie nepoznaného zo známych stavov - okolností. Pomocou nástrojov extrapolácie sa snažíme predvídať budúce dátové body, pri čom sa spoliehame na historické dáta, ktoré máme k dispozícii v určitej granularite (napr. odhad veľkosti populácie za niekoľko rokov na základe medziročného tempa rastu populácie od začiatku 19. storočia). Inak povedané, dátová extrapolácia je predikcia dát mimo hranice tabuľky. Tento proces sa dá jednoducho popísať ako vzorku dátových bodov s hodnotami x_1, \dots, x_n , ku ktorým sa snažíme aproximovať bod s hodnotou mimo rozsah daných bodov vo vzorke. Ak tento novo identifikovaný bod započítame do vzorky, môžeme iteráciami popísať ďalšie body a vytvoriť tak prediktívny trend. Každý ďalší bod, bude mať menšiu mieru pravdepodobnosti odhadu jeho hodnoty.

2. Podmienky a limity sledovania trendov pomocou extrapolácie

Ako každá štatistická technika, extrapolácia má podmienky, ktoré musia byť splnené za účelom jej efektívnej aplikácie a metodologické limity, v kontexte ktorých sa pohybuje. Predpovede na princípe extrapolácie môžu byť veľmi efektívne, pokiaľ kontext, z ktorého analyzované dáta pochádzajú, nenaznačuje žiadne prerušenie, alebo **nepravidelné výkyvy v dlhodobom vyskytujúcom sa trende**, či minulých dátovo paralelných trendoch. Príkladom takéhoto výkyvu sú hodnoty premenných na začiatku ekonomickej krízy v dátach finančnej burzy alebo zníženie globálnej teploty po výbuchu sopky v meteorologických dátach. Samozrejme, výkyvy v trende nám môžu pripadať ako nepravidelné alebo neočakávané z dôvodu nedostatočnej vzorky (napr. krátkych časových radov).

Avšak, takzvaná **extrapolácia z priamky** (pri ktorej predikujeme ďaleko do budúcnosti na základe krátkodobého trendu) so sebou vždy nesie riziko, pretože niektoré nepredvídateľné faktory doň takmer vždy zasiahnu. Extrapolácie môžeme použiť ako orientačnú evidenciu, napr. pre zistenie smeru vzťahu trendu a konštantnej premennej (napr. v porovnaní s terajším priemerom zrážok bude v blízkej budúcnosti pršať viac či menej?), určite však nie na predikciu presných hodnôt. Ak chceme aplikáciu extrapolačných metód správne použiť, je treba zvážiť nasledujúce okolnosti:

- **Chyba predpovede sa zvyšuje so znižovaním vzorky dát**, z ktorej chceme extrapolovať a naopak, pravdepodobnosť kvalitnej predpovede sa zvyšuje s objemom historických dát.
- Extrapolácia funguje obzvlášť dobre, pokiaľ sú **testovacie sety dát zbierané z konštantne bežiacего signálu v pravidelných intervaloch** a tieto dáta sú použité na predpovedanie budúcich dátových bodov pri zachovanom intervale (napr. v takom istom intervale, aký bol interval zberu).
- Pravdepodobnosť chyby predpovede sa zvyšuje s požiadavkou predpovedať dlhšie časové horizonty a naopak, **čím kratší horizont, čím bližší k súčasnosti, tým máme väčšiu pravdepodobnosť generovať presnú predpoveď**.
- Technikám extrapolácie, úspechom priamo naviazaným na veľkosť vzorky, **štandardne pomáha aj širšie pole zberu dát**. Napríklad v energetike a meteorológii pôjde o rozdiel medzi v extrapolovaním dát pre celé Slovensko verzus pre jeho lokálnu oblasť.
- Pri aplikácii extrapolačných metód musíme mať na pamäti, že sme vzali za svoj predpoklad, že aktuálny a budúci pohyb trendu sledovaného javu má priamy vzťah k minulým stavom, sa odvíja od jeho minulých stavov. Esenciálne informácie o jeho pohybe v budúcnosti sú teda uložené v dátach o jeho minulosti a sú extrahovateľné. **Predpokladáme, že minulé trendy budú pokračovať do budúcnosti**. Aby sme zabezpečili, že je tento predpoklad dlhodobu správny, je treba vytvoriť systematický

prístup k identifikácii elementov v dátach, ktoré predpokladáme, že zostanú stabilné a dlhodobo sledovať rozdiel medzi stabilnými a dynamickými zložkami trendu.

3. Výhody a prínosy vs. nevýhody a riziká extrapoláčnych techník

Výhody extrapoláčnych techník im pomohli udržať sa pri zavádzaní do praxe v oblastiach, kde je dôležité sledovať a spresňovať trendy v krátkych a častých intervaloch. Uvádzame najväčšie výhody aplikácií týchto metód:

- Extrapoláčne techniky vynikajú svojou **komputačnou jednoduchosťou**. To znamená, že pri dobrom nastavení mechanizmov kumulujúcich surové dáta do použiteľných agregátov, je systém, postavený na extrapoláčnych technikách, schopný bežať v reálnom čase a sledovať trendy intenzívnych dátových tokov.
- Druhou výhodou je **metodologická transparentnosť** extrapoláčnych systémov. Na rozdiel od predikcií generovaných pomocou neurónových sietí, pri extrapoláčnych technikách neexistuje „black box“, máme pri nich **úplnú kontrolu nad vývojom predikcie**. Na rozdiel od metód algoritmov posilňovaného učenia, má extrapoláčná štatistika **menšie nároky na iterácie**.
- Extrapolácie majú dobrú schopnosť **pracovať s agregovanými dátami** v kontexte času (s nižšou granularitou dát), bez umelého zníženia ich relevancie.
- Vďaka jednoduhosti, je jej **nasadenie bezproblémové**.

Extrapolácia je vhodná pre spracovávanie dát z veľkých oblastí, vykresľovanie trendov v krátkych horizontoch a na sledovanie premenných, v ktorých predpokladáme pomalé rasty a poklesy (teda nie náhle a extrémne výkyvy v hodnotách). Samozrejme, extrapolácia má aj nevýhody a nedostatky, ktoré paradoxne ako jej výhody, plynú z jej jednoduhosti:

- Štatistická extrapolácia neprináša odpoveď na otázku, čo je príčinou sledovaného trendu. Ide o čisto deskriptívnu techniku. Aj ako deskripcia, má nedostatky, nakoľko pracuje s agregátmi a teda **nesleduje špecifické zmeny stavu štruktúry** dátových vzoriek.
- To znamená, že napríklad tvorba nových kohort v dátach (napr. nový sektor populácie, nová skupina rovesníkov) je pre interpretátora vykresleného trendu neviditeľný. Na sledovanie dát o štruktúre a systematickom kontexte **musíme použiť pomocné frekvenčné a cluster analýzy**, ktoré nám pomôžu zväžiť zmyslupnosť aplikácie extrapoláčnych štatistických techník. Takúto analýzu štruktúry je potrebné robiť opakovane.
- Aplikácia extrapoláčnych techník ako primárneho nástroja tvorby predikcií je možná len vtedy, pokiaľ môžeme **vylúčiť prípadné externé vplyvy**, mimo systému, s ktorým počítame. Ide o neznáme premenné, ktoré by mohli posunúť budúce

hodnoty mimo škálu historických hodnôt (napríklad energetika je jedným z takýchto systémov).

- Automatizácia rozhodovacieho systému, založeného na dátach z extrapoláčnych predikcií, sa v praxi často nepoužíva. Výsledky extrapolácie **je nutné vizuálne skontrolovať a posúdiť**, či zodpovedajú našim poznatkom o správaní sa hodnôt v danej doméne. Tento typ analýzy má teda skôr verifikačný ako exploratívny charakter.
- Pri extrahovaní testovacích vzoriek dát za účelom extrapolácie je potrebné zozbierať dva typy vzoriek, nakoľko musíme **porovnávať výsledky extrapolácie málo a viac roztrúsenej vzoriek** dát (citlivosť na dátovú štruktúru).
- Podobný problém ako vyššie popísaný limit **extrapolácie z priamky** nastáva ak má systém počítať s veľkým množstvom premenných. Tento problém je však oveľa ťažšie detekovať, ako problém krátkych časových sledov dát. Riešením je často agregácia dát cez premenné, na základe identifikácie vzťahov medzi nimi a ich následnej korelácie.
- Pri sledovaní trendu je dobré dopredu vedieť aký typ trendu (model) máme očakávať. Ak máme informácie o očakávanom tvare trendu pre daný časový úsek (trend môže mať podobu lineárnej, kvadratickej, exponenciálnej funkcie atď.), môžeme použiť **metódu least-squares** (metóda najmenších štvorcov) za účelom nájdenia optimálne vykreslenej krivky trendu. Chyba extrapolácie krivky generovanej pomocou **least-squares** je často signifikantne menšia ako chyba extrapolácie z priamky pomocou interpolovaného polynomiálu (matematickej funkcie odpovedajúcej priebehu setu dát, modelu alebo trendu). Treba si však uvedomiť, že fakt, že sme pomocou least-squares identifikovali krivku veľmi dobre fitujúcu (najlepšie kopírujúcu) na originál, nie je dôkazom toho, že extrapolácia je metóda predikcie na základe konkrétnej vzorky dát v jej kontexte a kvalite – t.j. predikcia dosiahnutá pomocou takejto extrapolácie nemusí dosiahnuť kontext a kvalitu konkrétnej zdrojovej vzorky.
- Vo väčšine prípadov, **systémy v reálnom svete** nemajú tendenciu nasledovať aktuálne trendy dlhodobo.

4. Dáta použiteľné na extrapoláciu

Kvalita extrapolácie závisí priamo od kvality vzorky dát, z ktorej extrapolujeme, preto by sme radi zbežne popísali typy dát, vhodných priamo na extrapoláciu alebo aspoň jej podporu. Vieme využiť štyri zdroje dát: historické dáta, dáta z analogických kontextov, simulácia dát - dáta z laboratórnych alebo experimentálnych situácií a simulácia dát - dáta z priameho (poľného) zberu.

Historické dáta sú často priamou a prvou voľbou pri extrapolovaní. Pokiaľ máme dlhšie časové obdobie priamu skúsenosť (evidenciu) s javom, ktorý sa snažíme predikovať, je to jednoznačný zdroj dát, z ktorého chceme pri predikcii vychádzať. Presnosť extrapolácie na

základe historických dát závisí na dvoch premenných: a) presnosť merania historických dát a ich všeobecná kvalita, b) premenlivosť trendu.

Dáta z analogických kontextov sa používajú ako podporné dáta alebo v prípade, keď historické dáta nie sú k dispozícii (sú neúplné). Steinov paradox naznačuje, že pridaním analogických dát vieme značne spresniť výsledok extrapolácie. Podmienkou je evidencia o tom, že jav, ktorý vieme sledovať, je naozaj analogický k javu, ktorý sa snažíme predikovať. Tieto dva javy by teda mali byť v stabilnom vzťahu, či už korelačnom alebo kauzálnom. Na sledovanie - nakoľko sú dva javy analogické a dáta z nich použiteľné pre jednu extrapoláciu - existujú rôzne indexy.

Simulácie môžu byť použité na generovanie dát v prípade, že nemáme k dispozícii iné zdroje dát, alebo keď sa chceme uistiť o tom, že jav v historických dátach sa správa štandardne. Simulácie môžu prebiehať experimentálne v laboratórnych podmienkach alebo v poľnom výskume. Laboratórne podmienky nám dávajú lepšiu kontrolu nad externými vplyvmi a poľné zbery majú lepšiu ekologickú validitu (lepšie simulujú podmienky reálneho sveta). V tabuľke ponúkame hodnotenie jednotlivých zdrojov dát v kontexte cieľa predikcie:

HODNOTENIE JEDNOTLIVÝCH ZDROJOV DÁT V KONTEXTE CIEĽA PREDIKCIE

(1 = najpriaznivejšia alternatíva)

Zdroj dát	Odhad súčasného o stavu trendu	Očakávam že jav je reprezentovaný		Efekt výskumníkov ej zaujatosti (bias)	Cena zberu dát
		Malými zmenami	Veľkými zmenami		
		mi	mi		
Historické dáta	1	1	4	1	1
Dáta z analogických kontextov	2	4	3	2	2
Laboratórna simulácia	4	3	2	4	3
Poľný výskum	3	2	1	3	4

5. Dva príklady extrapolačných techník

Popíšeme si dva z mnohých nástrojov extrapolácie. Vybrali sme základné techniky, ktoré sa ale často používajú v praxi.

a. Technika Markov Chains

Pri technike Markov Chains používame vzorec nedávneho správania sa javu ako základ pre tvorbu prognózy. Pri tomto prístupe sa predpokladá sa, že správanie javu v budúcnosti sa dá odvodiť zo znalosti súčasného stavu javu a z analýzy toho, ako sa jav presunie z jedného stavu do druhého. Ide teda o progresívnu a nie regresívnu metódu, lebo sa primárne zameriava na analýzu krokov vpred ($n(t) - n(t+1)$) a nie na sledovanie historickej stability trendu. Dôležitým krokom je tu kvalitne zadefinovať stavy a akcie, ktorými sa jav presúva do ďalšieho stavu (to môžu byť napríklad jednotlivé zbery, alebo zmena ročného obdobia).

b. Technika Exponential Smoothing

Exponencial Smoothing je pomerne jednoduchá, intuitívna, lacná, a dobre známa metóda. Nadväzuje na filozofiu dekompozície. Predpokladáme pri nej, že časové rady sa skladajú z niekoľkých základných prvkov: priemer, trend, sezónnosť a chyba (odchýlka). Prvé tri zložky predstavujú historické dáta a sú tak aj vpisované do grafov. V extrapolačnom grafe sa tak dá znázorniť odhad aktualizovaného priemeru, predpoveď po tom, ako sme započítali trend do odhadu aktuálneho stavu a celkovú prognózu, ktorá bude kombináciou súčasného stavu, trendu a sezónnosti. K tejto kombinácii vieme vypočítať absolútnu a relatívnu chybu odhadu.

Exponencial Smoothing je podobný metóde kľzavého priemeru, ale novším údajom priradzuje väčšiu váhu. Priradené váhy klesajú exponenciálne od najaktuálnejších dát smerom k historickým, takže čím staršie údaje, tým menší je ich vplyv. To, že váhovanie historických dát dáva zmysel, je podporené viacerými empirickými štúdiami. Exponencial Smoothing môže byť použitý spolu s dekompozičnými technikami na analýzu historických dát.

K popisu základných metód prikladáme všeobecný popis hlavných krokov a princípov, ktoré musia byť dodržané pri extrapoláciách rôznych druhov. Kroky sú radené podľa časovej priority:

Výber a príprava dát:

- Získaj dáta reprezentujúce jav, ktorého správanie sa snažíme extrapolovať;
- Použi všetky dostupné dáta, najmä pri dlhodobých predpovediach;
- Popíš štruktúru javu a kontext, v ktorom sa jav nachádza (predikcia odchýlky výroby a spotreby elektrickej energie na Slovensku), aby si mohol na spresnenie extrapolácie využiť doménové znalosti;
- Vyčisti dáta a zníž chybu merania;
- Prispôb dáta časovým sériám, v akých boli zbierané;

- Prispôsob dáta relevantným historickým zmenám.

Sezónne úpravy na dátach (napasovanie dát na intervaly, ktoré poznáme):

- Prispôsob dáta sezónnym zmenám v správaní javu, ak tvoje doménové poznatky poukazujú na existenciu fluktuácií javu v ich kontexte a pokiaľ je vzorka na takúto úpravu dostatočná;
- Použi multiplikatívne faktory pre stabilné časové úseky a javy, ktoré majú presné pomerovo škálovateľné dáta;
- Vypusti sezónne úpravy dát, ak čelíš vyššej miere štatistickej neistoty.

Extrapolovanie:

- Kombinuj odhady rôznych úrovní predikcií;
- Použi jednoduchú reprezentáciu trendu, až na stavy kde je silná evidencia v prospech komplexnejších kriviek;
- Naváhuj aktuálnejšie dáta viac ako staršie dáta v prípade, že je chyba merania malá, horizont predpovede je krátky a opakované merania na rôznych vzorkách sú stabilné;
- Buď konzervatívny v príprave dát, analýze a interpretácii, ak je prítomná veľká miera neistoty (osobnej alebo/aj štatistickej);
- Použi doménové poznatky na vopred špecifikované úpravy extrapolácie.
- Použi deskriptívne štatistické metódy ako pomôcku pri výbere metódy extrapolácie;
- Aktualizuj odhady parametrov predikčného modelu často;
- Používaj cykly iba vtedy, ak sú dôkazy o budúcom načasovaní a amplitúde trendu veľmi presné.

Zohľadni štatistickú neistotu (pravdepodobnosť a chybovosť):

- Použi empirické odhady vychádzajúce z out-of-sample testov,
- Pre pomerovo škálovateľné dáta, odhadni predikčné intervaly pomocou log transformácie reálnych a predikovaných hodnôt;
- Použi bezpečnostné faktory pre ostatné typy dátových sérií.

Použi extrapoláčnne techniky:

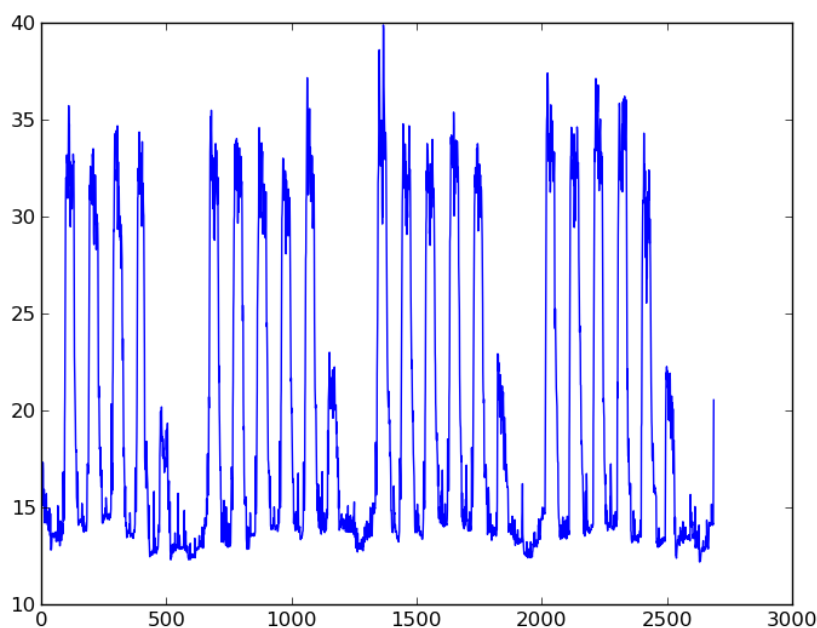
- Keď je potrebné robiť veľa predpovedí alebo opakované predpovede;

- Keď si môžeš dovoliť ignorovať väčšinu kontextuálnych premenných;
- Keď je situácia stabilná v čase;
- Keď použitie iných komplexnejších metód je nemožné;
- Keď potrebuješ merania nižšej miery efektivity predikcií (napr. identifikovanie predikčného potenciálu dát alebo stability správania sa javu).

6. Extrapolácia časových radov v kontexte odhadu odchýlky elektrickej spotreby

Ako sme videli v časti 3. Výhody a prínosy vs. nevýhody a riziká extrapoláčnych techník, štatistické techniky extrapolácie sú veľmi špecifické vo svojej aplikácii. Pre potreby predikcie v kontexte odhadu výroby a spotreby elektrickej energie na Slovensku je však extrapolácia veľmi použiteľným nástrojom. Striktne definované pravidlá výroby a spotreby, stabilné a očakávateľné vývoje trendov vo vzťahu ku klimatickým zmenám (napr. teplota), možnosť agregovania dát, pravidelnosť ich zberu, množstvo zbieraných dát, ich kvalitné namapovanie na rozlohu a relatívne malý počet premenných umožňuje využiť extrapoláciu v plnej sile najmä pri krátkych, frekventovane opakovaných predikciách na minút, hodín.

Ak by sme sa rozhodli extrapolovať dáta ktoré máme k dispozícii zo smartmetrov, t.j. 65 odberných miest zbieraných v časovom úseku od 1.1.2014 do 3.4.2014, na úrovni dní a týždňov, presnosť predikcie značne klesne. Na druhej strane, vďaka štatistickej kontrole extrapoláčného procesu môžeme vidieť, že sa nám stále darí nájsť v dátach trendy. Ako je jasne vidieť na nasledujúcom grafe, časové rady vykazujú aj na tejto úrovni pomerne periodické správanie:



Periódy sú hlavne týždenné a denné. Pokiaľ by sme chceli extrapolovať dáta automaticky na istú dobu dopredu pomocou automatu, ktorý by používal triviálnu metódu: “predikuj to isté, čo

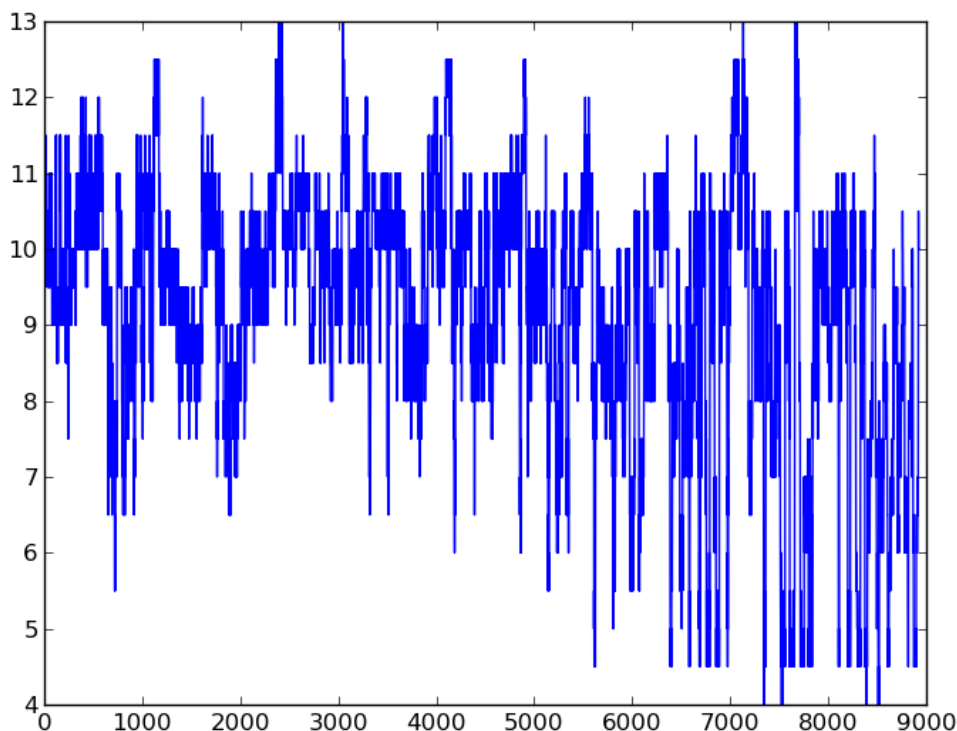
bolo x dní dozadu”, tak by sme dostali výsledky, ktoré nám určia hranicu, chybovosti, pod ktorú sa nikdy nedostaneme.

Do dostaneme nasledovné relatívne chyby predikcie:

Interval na predikciu dopredu [dní]	Relatívna chyba (cez všetky rady)	Relatívna chyba (na najlepšom rade)	Relatívna chyba (na najhoršom rade)
1	222%	5.83%	700%
7	75%	5.80%	189%
14	86%	5.78%	227%
21	90%	5.70%	258%
28	95%	5.80%	270%

Tieto chyby môžu pôsobiť ako veľké čísla, ale treba si uvedomiť, že automat predikuje spotrebu jedného odmerného miesta v krátkom časovom úseku. To znamená, že pracuje s pomerne veľkým výskytom rôznych náhodných faktorov, na ktoré, ako sme si povedali, sú extrapoláčne techniky veľmi citlivé. Ďalej si je treba uvedomiť, že väčšina predikcií pomocou extrapolácie sa doteraz zameriavala na hodinové úseky a väčšinou predikovala len lokálnu časť siete. My sa týmto testom snažíme určiť, kde má využitie extrapolácie limit. Preto, sa z tohto základného výsledku odrážame a posúvame výkonnosť systému smerom k väčšej presnosti tak, aby mohol byť jeho beh stále manažovaný automatom.

Pokiaľ zlepšime kvalitu vzorky, t.j. spotrebu predikujeme z viacerých dátových bodov (smartmetering) alebo ak chceme predikovať spotrebu na týždeň dopredu, tak pridáme okrem dát o súčasnej spotrebe aj spotrebu pred týždňa a dvoch týždňov, dostaneme lepšie výsledky. Ak použijeme nielen súčasný 15 minútový interval, ale aj intervaly okolo neho a tieto hodnoty vhodne skombinujeme, výsledok sa zase zlepší. Ešte lepšia presnosť by sa dala dosiahnuť, keby sme odberné miesta mohli segmentovať do viacerých skupín (napr. domácnosti, veľké fabriky), keďže sa dá predpokladať, že u týchto segmentov existuje iné správanie v odbere energie. Predpokladáme, že významný posun k presnosti predikcie pridá zakomponovanie sezónnosti (sezónneho správania spotrebiteľov a výrobcov). Takto sa dá chyba pri týždňovej predikcii dopredu znížiť na 55%. To je veľmi dobrý výsledok na metódu všeobecne považovanú za použiteľnú najmä na odhad minútových a hodinových intervalov. Ešte lepší spôsob je natrénovať lineárnu regresiu, ktorá na vstupe dostane posledný týždeň časového radu a predikuje spotrebu na týždeň dopredu. Ukazuje sa, že je možné dosiahnuť chybu okolo 50%. Zaujímavým faktom je, že u niekoľkých radov ale periódu nevidno (viď nasledujúci graf). Dôvodov môže byť viac, od neštandardného správania sa odberateľa, po nedostatky vo vzorke dát, s ktorou pracujeme.



7. Umelé rožširovanie vzorky

Pri nedostatku dát sme v niektorých situáciach schopní použiť metódy na umelé rozšírenie testovacej vzorky. Tieto metódy však treba používať s veľkou opatrnosťou, nakoľko s roširením trendu často v nepriamom pomere rozširujeme aj chybu predikcie.

Jednou z metód je pridávanie nových časových radov. Rozširovanie vzorky je štandardná technika pri tvorbe automatických expertných systémov rozpoznávaní obrazu, keď ľahko vyrábame nové vzorky malou modifikáciou známych vzoriek (posunutie, otočenie, natiahnutie). Technika sa používa hlavne na zlepšenie robustnosti algoritmov, teda na vylepšenie schopnosti systému predikovať - identifikovať trendy, nie na vylepšenie deskriptívnych štatistík. Tento prístup sa dá presunúť aj do energetiky, kde vyrobíme nové časové rady malým pozmenením predchádzajúcich (pomocou manipulácie šumu, prenasobením konštantou, atď.) a potom môžeme zlepšiť robustnosť algoritmov na predikciu a segmentáciu.

Druhou možnosťou je už spomenutá simulácia, pridávanie umelých budúcich meraní. Presnosť týchto meraní nebude veľká, no dala by sa zlepšiť použitím analogických (napr. teplotných) dát. Na týchto dátach nemá zmysel predikovať na viac ako týždeň dopredu a majú slúžiť na orientačnú analýzu.

Poslednou užitočnou technikou rozšírenia vzorky je pridávanie nových atribútov. Nové atribúty môžu predstavovať štatistické konštrukty alebo agregáty veličín. Dajú sa dopočítať

dodatočné štatistiky pre každý časový rad ako: plávajúce priemery, hodinové/denné dáta, štandardné odchýlky, časové priemery.

8. Záver

Naším cieľom v tomto reporte bolo urobiť prehľad problematiky použitia extrapoláčnych metód na vzorkách dát, ktoré máme k dispozícii a vyjadriť sa k ich použiteľnosti pri tvorbe nového predikčného systému v kontexte smartmeteringu.

Technický popis tohto procesu je príliš komplexný pre potreby výstupného reportu, ktorý odovzdávame v rámci projektu, preto report obsahuje hlavné závery, odporúčania a špecifikácie, ktoré sme extrahovali počas našej práce.

Extrapoláčne techniky sú veľmi užitočné a jednoduché nástroje na popisovanie vývoja trendov v minútových, hodinových a denných časových úsekoch. Majú potenciál byť relevantným zdrojom verifikačných dát pri predikciách na dlhšie časové úseky – týždňov a mesiacov. Najlepší výkon dosahujú pri jemne rastúcich a klesajúcich hodnotách trendov a pri optimalizácii identifikovania trendov z periodicky zbieraných dát.

Doménovo-špecifické poznatky o energetike, dátovej štruktúre a sezónnych fluktuáciách významne zlepšujú výkon extrapoláčnych nástrojov. Automatizované identifikovanie trendov pomocou extrapolácie je veľmi efektívne v kontexte zvládania dátového objemu vtedy, ak sú správne nastavené systémy tvorby dátových agregátov a periodického vyťahovania testovacích vzoriek z toku surových dát, bežiacého v reálnom čase.