

SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE

Fakulta informatiky a informačných technológií

Optimalizácia konfiguračných parametrov predikčných metód

BAKALÁRSKA PRÁCA

SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE

Fakulta informatiky a informačných technológií

Optimalizácia konfiguračných parametrov predikčných metód

BAKALÁRSKA PRÁCA

Študijný program:

Číslo študijného odboru:

Názov študijného odboru:

Školiace pracovisko:

Vedúci záverečnej práce:

Bratislava 2016

Matúš Cuper

Obsah

1	Analýza predičných algoritmov	3
1.0.1	Lineárna regresia	3
1.0.2	Support vector regression	3
1.0.3	Random forrest	3
1.0.4	Rozhodovacie stromy	3
2	Analýza optimalizačných algoritmov	4
3	Špecifikácia	5
3.0.1	Meranie chyby predikcie	5
3.1	Naivné prístupy	5
3.1.1	5
3.2	Klasické prístupy	5
3.2.1	Stupňovitý regresný model	5
3.2.2	Neurónové siete	5
3.2.3	Rozhodovací strom	5
3.2.4	AR	5
3.2.5	MA	5
3.2.6	ARMA	5
3.2.7	ARIMA	5
3.3	Prírastkové učenie	5
3.3.1	Prírastkové SVM	5
3.3.2	Extrémne strojové učenie	5
3.3.3	Prírastková ARIMA	5
3.4	Učenie súborov klasifikátorov	5
3.4.1	Bagging	6
3.4.2	Boosting	6
3.4.3	Adaboost	6
3.4.4	Stacked generalization	6
3.4.5	Mixture of experts	6
4	Záver	8

Úvod

Pri meraní odberu elektriky vznikajú dva typy tzv. Concept drift. **Concept drift** je zmena správania veličiny, ktorú sa snažíme predpovedať. Model sa tak stáva postupne nepresný a je potrebné aby sa tejto zmene prispôbil. Prvým typom je trvalá alebo dočasná zmena spôsobená ekonomickými alebo ekologickými faktormi. Druhým typom je sezónna zmena, spôsobená zmenami ročných období a množstvom denného svetla. Sezónnu zmenu môžeme pozorovať na úrovni dní, týždňov alebo rokov. Kvôli tomu je nutné v každom modeli rozdeľovať tieto dva typy concept drift[1].

Kapitola 1

Analýza predičných algoritmov

Na základe množstva predikčných

Lineárna regresia

Multiple regression model

Viacnásobná regresia sa pokúša modelovať vzťah medzi dvoma alebo viacerými nezávislými premennými a závislou premennou vhodnou lineárnou rovnicou pre pozorované dáta. Výsledný model je vyjadrený ako funkcia viacerých nezávislých premenných a predpoveďou nie je priamka ako je to pri lineárnej regresii[1]. Nezávislé premenné môžu predstavovať meteorologické vplyvy, ekonomický rast, ceny elektriky či kruzy mien[2].

Support vector regression

Random forrest

Rozhodovacie stromy

Regresný rozhodovací strom

Kapitola 2

Analýza optimalizačných algoritmov

Kapitola 3

Špecifikácia

Meranie chyby predikcie

Naivné prístupy

Klasické prístupy

Stupňovitý regresný model

Neurónové siete

Rozhodovací strom

AR

MA

ARMA

ARIMA

Prírastkové učenie

Prírastkové SVM

Extrémne strojové učenie

Prírastková ARIMA

Učenie súborov klasifikátorov

Dva prístupy: homogénne a heterogénne učenie ale aj ich kombinácia **Homogénne učenie súborov klasifikátorov** pozostáva z modelov rovnakého typu, ktoré sa učia na rôznych podmnožinách datasetu. **Heterogénne učenie súborov klasifikátorov** aplikuje rôzne typy modelov nad rovnakými dátovými množinami[1].

Bagging**Boosting****Adaboost****Stacked generalization****Mixture of experts**

Regression model je funkcia (f), ktorou sa snažíme odhadnúť závislú premennú (Y), pomocou nezávislej premennej (X) a neznámych parametrov (Beta), (Beta) odhadujeme na základe dát, (f) je známa na základe vzťahov medzi (Y) a (X)[1]. najpoužívanejšia štatistická metóda, modeluje vzťah závislej premennej (spotreba elektriky) a nezávislej premennej, čiže ostatných faktorov (počasie, deň v týždni, odoberateľ), spotreba tak môže byť rozdelená na štandardný smer (trend) a trend lineárne závislý od niektorých faktorov vplývajúcich na spotrebu[2].

Multiple regression model najpopulárnejšia metóda, často používa na predpoveď spotreby ovplyvnenej množstvom faktorov (meteorologické, cena elektriky, ekonomický nárast)[2].

Autoregressive model môže modelovať profil záťaže za predpokladu, že záťaž je lineárnou kombináciou predchádzajúcich záťaží[2].

Autoregressive Moving-Average model model reprezentuje súčasnú hodnotu časového rádu lineárne na základe jeho hodnôt a hodnôt bieleho šumu v predchádzajúcich periódach[2].

Ensemble model používa sa na jednoduchú predikciu, h je počet meraní, ktoré sú denne dostupné, v deň t sa vykoná h predikcií podľa váženého priemeru m modelmi, nasledujúci deň sa vypočíta chyba predpovede, na základe ktorej sa znova prepočítajú váhy a každý model sa aktualizuje[1].

Seasonal naïve method poslednú zmeranú hodnotu použijem ako predpoveď alebo pre high seasonal data použijem už nameranú hodnotu z rovnakého obdobia (napr. rok dozadu)[1].

Naïve average long-term method je založené na predpoklade non-seasonal patterns, predpokladá, že časové rady sú lokálne stabilné s pomaly meniacim sa priemerom, predpovedaná hodnota je priemerom viacerých hodnôt[1].

Naïve In median long-term method je alternatíva k predchádzajúcej metóde, priemer nie je schopný reagovať na rapídne výkyvy a abnormality, lepšia možnosť je preto spraviť median z posledných n časových radov[1].

Stochastic Time Series metódy časových radov sú založené na predpoklade, že dáta majú vnútornú štruktúru, ako napr. autokoreláciu, trend či sezónnu variáciu, najprv sa precízne zostaví vzor zodpovedajúci dostupným dátam a potom sa predpovie hodnota[2].

Support Vector Machine based Techniques je metóda analyzujúca dáta a rozpoznávajúca vzory, používaná na roztriedenie a regresnú analýzu, kombinuje zovšeobecnené riadenie s technikou ??????[2].

Concept drift je správanie premennej, ktorú sa snažíš predikovať sa môže časom meniť, čím sa postupne stáva model menej a menej presný[1].

Online algorithm spracováva vstup sériovo kúsok po kúsku, vstupné dáta nie sú dostupné na začiatku výpočtu musí spracovať vstup v jednej iterácii bez žiadnej podrobnej znalosti budúcich vstupov viac dát, časové obmedzenia, môže sa časom meniť

Offline algorithm rieši problém od začiatku so všetkými vstupnými dátami vopred je daná celá séria vstupov

Support Vector Machine je ML algoritmus používaný ako na klasifikáciu tak na regresiu support vector sú koordináty jednotlivých meraní napr. muž a žena a ich merané veličny reprezentované na osy, ktoré sú hraničnými elementami rôznych skupín maximalizuje rozmädie medzi support vektormi jednej kategórie a support vektormi druhej kategórie, rozhodovacia funkcia je definovaná podmnožinou testovacej vzorky (jednotlivé support vektory) v 2D sú kategórie oddelené čiarou vo viacrozmerných dimenziách rovinou

Incremental SVM základom je pridávanie nový bod má najskôr pridelenú váhu 0, ak toto pridelenie nie je optimálnym riešením, teda bod sa môže stať support vectorom, váhy ostatných vektorov a rozhodovací prah musia byť aktualizované kvôli získaniu optimálneho riešenia nad novou množinou support vektorov

Linear SVM lineárna kombinácia elementov (features, črty) značí, že sa jedná aj o lineárny klasifikátor napr ak $(w_1 * x_1 + w_2 * x_2) \geq C$ potom element patrí do skupiny A, hodnotami x_1 a x_2 je element definovaný, tak ako je bod definovaný x a y súradnicou w je váha a C rozhodovací prah, čiže ak nejaký ohodnotený element neprekročí hranicu spadá do jednej skupiny, ak prekročí spadá do druhej

Kernel trick problém nie je lineárne separovateľný, originálny nelineárny priestor je premietnutý do viacrozmerného priestoru pomocou nejakej nelineárnej transformácie s očakávaním, že to problém už bude riešiteľný

Extreme learning machine je novovznikajúca technika učenia poskytujúca efektívne a zjednotené riešenie na všeobecné dopredné siete ako neurónové siete, RBF siete alebo kernelové učenie

Kapitola 4

Záver

Tu bude záver

Literatúra

- [1] G. Grmanová, P. Laurinec, V. Rozinajová, A. Bou Ezzeddine, M. Lucká, P. Lacko, P. Vrablecová, and P. Návrát, “Incremental Ensemble Learning for Electricity Load Forecasting,” *Acta Polytechnica Hungarica*, vol. 13, no. 2, 2016.
- [2] A. Kumar Singh, S. Khatoon, M. Muazzam, and D. K. Chaturvedi, “An Overview of Electricity Demand Forecasting Techniques,” *NAJDI JOURNAL*, vol. 3, no. 3, 2013.