

TECHNICKÁ UNIVERZITA V KOŠICIACH
FAKULTA ELEKTROTECHNIKY A INFORMATIKY

**EVOLUČNÁ OPTIMALIZÁCIA PARAMETROV METÓDY
PREDIKCIE ZALOŽENEJ NA STRUNOVÝCH INVARIANTOCH**

Diplomová práca

TECHNICKÁ UNIVERZITA V KOŠICIACH
FAKULTA ELEKTROTECHNIKY A INFORMATIKY

**EVOLUČNÁ OPTIMALIZÁCIA PARAMETROV METÓDY
PREDIKCIE ZALOŽENEJ NA STRUNOVÝCH
INVARIANTOCH**
Diplomová práca

Študijný program:	Hospodárska informatika
Študijný odbor:	9. 2. 10. Hospodárska informatika
Školiace pracovisko:	Katedra kybernetiky a umelej inteligencie
Školiteľ:	Ing. Marek Bundzel, PhD.

Abstrakt v SJ

Práca je venovaná návrhu a analýze evolučného algoritmu pre optimalizáciu parametrov predikčnej metódy založenej na strunových invariantoch (PMBSI), porovnaní PMBSI s existujúcimi štatistickými metódami a metódami založenými na umelej inteligencii. Cieľom práce je opísať moderné metódy predikcie časového radu pomocou *Support vector machines* a Neurónových sietí a navrhnúť evolučný algoritmus pre novú metódu PMBSI a implementovať v programovacom jazyku Python. Výsledkom práce je efektívna optimalizácia parametrov navrhnutým algoritmom a porovnanie PMBSI s modernými metódami na dostatočnom počte dát.

Kľúčové slová

predikčný model založený na strunových invariantoch, neurónové siete, evolučné algoritmy

Abstract

The thesis is dedicated to development of an evolution algorithm for optimization of parameters for prediction model based on string invariants (PMBSI) and comparison with existing statistical methods and methods based on artificial intelligence. The goal of this diploma thesis is to describe modern methods for time series forecasting like Support vector machines and Neural networks and develop evolution algorithm for new forecasting method PMBSI and implement it in Python programming language. The result of this thesis is effective optimization of parameters for PMBSI and comparison with modern methods for time series forecasting on adequate data set.

Key words

Prediction Model Based on String Invariants, PMBSI, Neural Networks, NN, Support Vector Machines, SVM, Evolution Algorithms, EA

TECHNICKÁ UNIVERZITA V KOŠICIACH

FAKULTA ELEKTROTECHNIKY A INFORMATIKY

Katedra kybernetiky a umelej inteligencie

ZADANIE DIPLOMOVEJ PRÁCE

Študijný odbor: **9.2.10 Hospodárska informatika**

Študijný program: **Hospodárska informatika**

Názov práce:

**Evolučná optimalizácia parametrov metódy predikcie založenej na
strunových invariantoch.**

Evolutionary optimization of the parameters of the prediction method based on
string invariants

Študent: **Bc. Matúš German**

Školiteľ: **Ing. Marek Bundzel, PhD.**

Školiace pracovisko: **Katedra kybernetiky a umelej inteligencie**

Konzultant práce:

Pracovisko konzultanta:

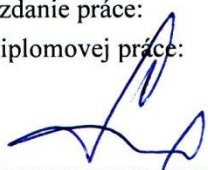
Pokyny na vypracovanie diplomovej práce:

1. Vypracovať prehľad problematiky predikcie vývoja časových radov so zameraním na moderné metódy.
2. Popísať metódu predikcie založenej na strunových invariantoch.
3. Navrhnuť a implementovať evolučnú optimalizáciu parametrov metódy predikcie založenej na strunových invariantoch.
4. Navrhnuť a realizovať experimentálne overenie funkčnosti optimalizovaného predikčného systému.
5. Vypracovať dokumentáciu podľa pokynov vedúceho DP.

Jazyk, v ktorom sa práca vypracuje: slovenský

Termín pre odovzdanie práce: 02.05.2014

Dátum zadania diplomovej práce: 31.10.2013


.....
prof. Ing. Peter Sinčák, CSc.
vedúci garantujúceho pracoviska




.....
prof. Ing. Liberios Vokorokos, PhD.
dekan fakulty

Čestné vyhlásenie

Vyhlasujem, že som celú bakalársku prácu vypracoval samostatne s použitím uvedenej odbornej literatúry.

Košice, 30. apríl 2014

.....

vlastnoručný podpis

Pod'akovanie

Touto cestou by som sa chcel poďakovať môjmu vedúcemu a zároveň konzultantovi diplomovej práce

Ing. Marekovi Bundzelovi, PhD.

za ochotu, všetok svoj čas a námahu, ktorú vynaložil, aby ma viedol pri písaní diplomovej práce.

Predhovor

Poznanie budúcnosti láka ľudí, či už za účelom predchádzaniu ovplyvniteľných nežiaducich udalostí, alebo úspora nákladov poznaním technologického vývoja či pre zisk z predpovede pohybov na finančných trhoch. V súčasnosti sa veľká pozornosť venuje predikčným metódam založeným na umelej inteligencii ako *Neural Networks* (NN), *Support Vector Machines* (SVM), *Evolution Algorithms* (EA), *Particle Swarm Optimization* (PSO) apod. Je to hlavne vďaka súčasnému stavu výpočtovej techniky s relatívne rýchlym výpočtom. Matematicky presne opísať chaotický časový rad, napríklad časový rad z finančných trhov, je zložité, možno nemožné.

Nový prístup z teórie strún s názvom *Prediction Model Based on String Invariants* (PMBSI) - predikčný model založený na strunových invariantoch je jeden z prvých pokusov aplikovať teóriu strún [18] v oblasti predikcie časových radov. Zámerom mojej práce je navrhnúť evolučnú optimalizáciu pre efektívne nájdenie parametrov pre PMBSI.

Obsah

Zoznam obrázkov	9
Zoznam tabuliek	10
Zoznam symbolov a skratiek	11
Slovník termínov	12
Úvod	13
1 Formulácia úlohy	15
2 Analýza	16
3 Predikcia časových radov modernými metódami	17
3.1 Časový rad	17
3.2 Predikcia pomocou neurónových sietí	17
3.3 Predikcia pomocou <i>Support Vector Machines</i>	20
4 Metóda predikcie časového radu založená na strunových invariantoch	23
4.1 Definovanie strunových máp	23
4.2 Strunové invarianty	24
4.3 Tvorba predikcie	26
5 Návrh evolučného algoritmu pre optimalizáciu parametrov pre model predikcie časového radu založeného na strunových invariantoch	28
5.1 Evolučné algoritmy, prvky a princípy	29
5.2 Reprezentácia jedincov	30
5.3 Selekcia jedincov	30
5.4 Genetické operátory	31
5.5 Náhrada jedincov	33
5.6 Vývojový diagram navrhnutého evolučného algoritmu	34
5.7 Vývojové prostredie pre návrh evolučného algoritmu	36
6 Experimenty	37
6.1 Úvodné poznámky k experimentom	37
6.2 Experiment na umelých dátach	38
6.3 Experiment na finančných dátach	40
6.4 Porovnanie s výsledkami zo súťaže „2008 Time Series Forecasting Competition for Computational Intelligence“	43
6.5 Zhodnotenie experimentov	47

Záver	49
Zoznam použitej literatúry	50
Prílohy.....	52
Príloha A: Kompletné výsledky z experimentu zo súťaže	53

Zoznam obrázkov

Obrázok 1 Matematický trojvrstvový model neurónovej siete pre predikciu	18
Obrázok 2 Výsledky návratnosti pri teste NN, lineárnej regresie a tradičných stratégií z [10]	19
Obrázok 3 Transformácia z lineárne neseparovateľnej úlohy na lineárne separovateľnú [13]	20
Obrázok 4 Model SVM nelineárnej regresie (SVR) [12]	21
Obrázok 5 MAE chyba s 5 krokovou predikciou z experimentu z článku. Červená bodka predstavuje globálne minimum nájdených riešení. [16]	28
Obrázok 6 Alternatívy pre voľbu reálnej hodnoty pri krížení [23]	32
Obrázok 7 Ilustrácia aritmetického krížení [23]	32
Obrázok 8 Vývojový diagram navrhnutého evolučného algoritmu	34
Obrázok 9 Vývojové prostredie PyScripter	36
Obrázok 9 Graf evolúcie trénovacej a testovacej MAE na sínusoide pre 2 krokovú predikciu	39
Obrázok 10 Porovnanie 3 krokovej predikcie PMBSI s EA a článkom [16]	40
Obrázok 11 10 kroková priama predikcia finančných dát pomocou PMBSI	42
Obrázok 12 Grafy evolúcie trénovacej a kontrolnej MAE na finančných dátach	43
Obrázok 13 Príklad časových radov zo súťaže [25]	44
Obrázok 14 56-kroková predikcia pomocou PMBSI zo súťaže s najnižším SMAPE 19,75	46
Obrázok 15 56-kroková predikcia pomocou PMBSI zo súťaže s najhorším SMAPE 76,98	46
Obrázok 16 56-kroková predikcia pomocou PMBSI zo súťaže so SMAPE najbližšej priemeru 37,75	47

Zoznam tabuliek

Tabuľka 1 Výsledky testov porovnania ARIMA a NN z [8].....	19
Tabuľka 2 Výsledky testov porovnania BPNN a SVM z [14]	21
Tabuľka 3 Výsledky experimentu na umelých dátach.....	38
Tabuľka 4 Optimálne parametre PMBSI pre sínusoidu	39
Tabuľka 5 Výsledky experimentu na finančných dátach.....	41
Tabuľka 6 Optimálne parametre PMBSI pre finančné dáta	42
Tabuľka 7 Porovnanie PMBSI s výsledkami súťaže "2008 Time Series Forecasting Competition for Computational Intelligence"	45

Zoznam symbolov a skratiek

PMBSI	P rediction M odel B ased on String I nvariants – predikčný model založený na strunových invariantoch
NN	N eural N etwork – neurónová sieť
BPNN	B ack- P ropagation N eural N etwork – neurónová sieť so spätným šírením chyby
EA	E volution A lgorithm – evolučný algoritmus
SVM	S upport V ector M achines
SRM	S tructural R isk M inimization – minimalizácia štrukturálneho rizika
SVR	S upport V ector R egression – regresia pomocou SVM
ARIMA	A uto- R egressive I ntegrated M oving A verage – autoregresívny integrovaný klzavý priemer
CBR	C ase B ased R easoning
MAE	M ean A bsolute E rror – stredná absolútna chyba
SMAPE	S ymetric M ean A bsolute P ercentage E rror – symetrická stredná absolútna percentuálna chyba
RMSE	R oot M ean S quared E rror – stredná kvadratická chyba
NMSE	N ormalized M ean S quare E rror – normalizovaná stredná kvadratická chyba
l_s	dĺžka struny PMBSI
l_{pr}	počet krokov dopredu ktoré chceme predpovedať PMBSI
Q	parameter pre umocnenie významných udalostí PMBSI
η_1, η_2	homotopné premenné pre parametrizáciu PMBSI

Slovník termínov

Časový rad je postupnosť dát, ktoré sú jednoznačne usporiadané z hľadiska času od minulosti k prítomnosti, respektíve budúcnosti.

Predikcia je predpoveď zmeny v čase, alebo extrapolácia z aktuálnych podmienok do budúcnosti.

Invariant znamená nemenný, znamená to že niečo sa transformáciou nemení, ako napríklad úprava rovníc do iného referenčného tvaru.

Lineárna regresia predstavuje aproximáciu daných hodnôt pomocou polynómu prvého rádu metódou najmenších štvorcov.

Buy and hold je investičná stratégia založená na tom, že v dlhodobom meradle finančné trhy rastú bez ohľadu na výkyvy.

T-Bill index je index založený na dennej výnosovej krivke U.S. Treasury bills

Neuron predstavuje základnú stavebnú jednotku neurónových sietí.

Evolúcia je vzchádzanie alebo utváranie sa jedných stavov z druhých

Populácia je množina jedincov

Selekcia slúži na výber rodičov z aktuálnej populácie pre vytvorenie novej populácie

Kríženie je kombinácia rodičov rôznymi metódami pre vytvorenie potomka

Mutácia náhodná zmena s určitými obmedzeniami, slúži pre vyhnutie sa zacykleniu v lokálnych extrémoch.

Python programovací jazyk

Úvod

Vo svete je narastajúci trend potreby predpovedania finančných časových radov, predikcie budúcich cien mien, akcií, indexov atď. Umelá inteligencia sa stáva dôležitou súčasťou pri rozhodovaní investorov. Avšak ako vysvetľuje Yaser v [2], finančné časové rady sú vo svojej podstate veľmi zašumené a nestacionárne. Šum je spôsobený nemožnosťou získania kompletných informácií zo správania sa finančných trhov. Nestacionárnosť je spôsobená rozdelením finančných časových radov ktoré sa časom mení. Preto predikcia finančných časových radov je považovaná za jednu z najnáročnejších úloh predikcie.

Najpopulárnejšie metódy predikcie finančných časových radov pomocou umelej inteligencie sú NN pretože dokážu zachytiť jemné finančné vzťahy medzi dátami napriek tomu, že pôvodne tieto vzťahy sú neznáme alebo ťažko vyjadriteľné. Sú neparametrické a dokážu zmapovať nelineárne funkcie bez predošlých predpokladov na dáta. [3] Predikciou trendových časových radov sa zaoberá Zhang a Qi v [8] a [9]. V oblasti predikcie finančných časových radov je najznámejšia *Back-Propagation Neural Network* (BPNN) – neurónová sieť so spätným šírením chyby vďaka jednoduchej architektúre v porovnaní s výkonom. BPNN ale vyžaduje veľký počet parametrov a vzniká problém možného preučenia. Preučenie je problém ktorý vedie k slabému zovšeobecneniu, čiže NN bude preoptimalizovaná pre trénovacie dáta a výkon vysoko klesá na dátach mimo týchto trénovacích dát. Typickým riešením je používanie krížovej validácie. Boli vyvinuté rôzne postupy pre zovšeobecnenie NN v [5] a [6].

SVM sú podobné neurónovým sieťam, založené na štatistických učiacich teóriách, analyzovaní dát a spoznávaní vzorov. [11] Finančnou predikciou pomocou SVM sa zaoberá Tay a Cao v [14] kde aj porovnávajú SVM model s BPNN.

Nový prístup z teórie strún s názvom *Prediction Model Based on String Invariants* (PMBSI) - predikčný model založený na strunových invariantoch je na základe prístupu z [17]. PMBSI reprezentuje jeden z prvých pokusov aplikovať teóriu strún v oblasti predikcie časových radov. Teória strún bola vyvinutá počas minulých 25 rokov a stala sa veľmi populárna [18]. V článku [16] je analyzovaný PMBSI a experimentálne porovnaný s SVM. Taktiež bolo analýzou odporúčané navrhnuť optimalizáciu parametrov evolučným algoritmom.

V tejto práci sa zaoberám návrhom evolučného algoritmu pre efektívnu optimalizáciu parametrov pre PMBSI a porovnávam výsledky s článkom [16]. Porovnam PMBSI s inými metódami založenými na metódach umelej inteligencie a štatistickými metódami.

1 Formulácia úlohy

Mojou úlohou a cieľmi diplomovej práce sú:

1. Teoretická časť

- popísať problematiku predikcie časových radov
- vypracovať prehľad moderných metód predikcie časových radov
- popísať metódu predikcie časového radu založenú na strunových invariantoch
- popísať evolučné algoritmy, prvky a princípy

2. Praktická časť

- navrhnúť evolučný algoritmus pre optimalizáciu parametrov pre metódu predikcie založenej na strunových invariantoch
- navrhnúť a realizovať experimentálne overenie funkčnosti optimalizovaného predikčného systému použitím programovacieho jazyka Python
- porovnať výsledky s výsledkami z článku [16]
- porovnať navrhnutý model s inými metódami predikcie založenými na metódach umelej inteligencie a štatistickými metódami
- Vypracovať dokumentáciu

2 Analýza

Pre analýzu a popísanie predikcie časových radov modernými metódami som musel naštudovať problematiku NN, SVM, evolučných algoritmov. Musel som pochopiť princíp predikcie časového radu založený na strunových invariantoch.

K vypracovaniu diplomovej práce mi pomohli predmety ako Programovanie, Počítače a Algoritmizácia, kde som sa naučil programovať v objektovo orientovanom jazyku, metódy algoritmizácie, a správne používať základné zložky algoritmov. Ďalej mi pomohli predmet Numerická matematika, pravdepodobnosť a matematická štatistika, Aplikovaná štatistika, Ekonometria kde som sa naučil správne pracovať s časovými radmi, a vyhodnocovať výsledky.

Na programovú realizáciu evolučného algoritmu som si vybral programovací jazyk Python pretože predikčný model PMBSI dodaný vedúcim DP je vytvorený v tomto jazyku a ako vývojové prostredie som zvolil program PyScripter.

3 Predikcia časových radov modernými metódami

V tejto kapitole sa budem venovať opisu moderných metód predikcie časových radov, ktoré modelujú časové rady iba pomocou informácií obsiahnutých v ich vlastných minulých hodnotách. Predikcia časových radov je neustále sa rozvíjajúce odvetvie s nespočetným množstvom metód. V živote sú procesy často nelineárne, ktoré môžu byť zachytené, predpovedané nelineárnymi modelmi založenými na prístupoch, z ktorých najmodernejšie modely sú založené na NN a SVM.

3.1 Časový rad

Pod časovým radom rozumieme postupnosť dát, ktoré sú jednoznačne usporiadané z hľadiska času od minulosti k prítomnosti, respektíve budúcnosti. Matematicky ako vektor $x(t), t=0,1,2,\dots$ kde t predstavuje uplynutý čas. Merania v časovom rade sú usporiadané v chronologickom poradí.[1]

Časový rad je ovplyvnený štyrmi komponentmi:

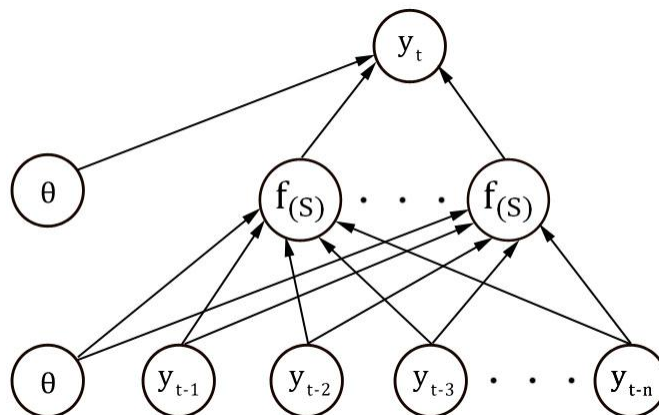
- Trend – tendencia časového radu stúpať, klesať alebo stagnovať počas dlhšej časovej periódy.
- Cyklus – predstavuje v časovom rade strednodobé zmeny, spôsobené okolnosťami ktoré sa opakujú v cykle.
- Sezóna – fluktuácie spôsobené ročnými obdobiami.
- Nepravidelnosť – náhodné variácie, spôsobené nepredvídateľnými vplyvmi, ktoré nie sú pravidelné a neopakujú sa v určitom vzore.

3.2 Predikcia pomocou neurónových sietí

„Neurónová sieť je masívny paralelný procesor pozostávajúci z jednoduchých procesných jednotiek, ktorý má sklon k uchovávaní experimentálnych poznatkov a ku ich sprístupňovaniu na ďalšie použitie. Podobá sa na ľudský mozog v dvoch ohľadoch:

- 1. Poznatky sú získané neurónovou sieťou procesom učenia.*
- 2. Sila medzineurónových spojení vyjadrená veľkosťou synaptických váh slúži na ukladanie znalostí.“[3]*

Neural Networks (NN), neurónové siete sú skupina pružných nelineárnych modelov, ktoré môžu adaptívne objavovať vzory z dát. Teoreticky bolo preukázané, že vzhľadom k počtu nelineárnych neurónov, môžu sa NN učiť zo skúseností a odhadnúť akékoľvek zložité funkčné vzťahy s vysokou presnosťou. Existuje rad úspešných aplikácií založených na rozpoznávaní vzorov a predpovedaní časových radov. Aj keď bolo navrhnutých veľa typov NN, pre predikciu časových radov je najznámejšia dopredná NN. Obrázok 1 znázorňuje typický 3 vrstvový model pre predikciu.[8]



Obrázok 1 Matematický trojvrstvový model neurónovej siete pre predikciu

Vstupné neuróny predstavujú predchádzajúce dáta a výstupný neurón predstavuje skúmanie budúcej hodnoty. Skrytá vrstva obsahuje neuróny s nelineárnymi aktivačnými funkciami pre spracovanie informácií získané zo vstupných neurónov. Model môže byť znázornený ako v [8]:

$$y_t = \alpha_0 + \sum_{j=1}^n \alpha_j f \left(\sum_{i=1}^m \beta_{ij} y_{t-i} + \beta_{0j} \right) + \varepsilon_t, \quad (3.1)$$

kde m je počet vstupných neurónov, n je počet neurónov v skrytej vrstve, f je aktivačná funkcia, α_j je vektor váh od skrytých neurónov k výstupným, β_{ij} sú váhy od vstupných neurónov k skrytým.

Neurónové siete vedľa veľmi slabo modelovať sezónne a trendové dáta. Z dát musí byť odstránená sezónnosť čo bolo testované v [7] a zároveň potvrdené Qi a Zhangom v [8] a preukázané že najefektívnejšie je z dát odstrániť nie len sezónnosť ale aj vyfiltrovať periodické zložky. Tento model NN pre predikciu časového radu porovnávali Qi a Zhang

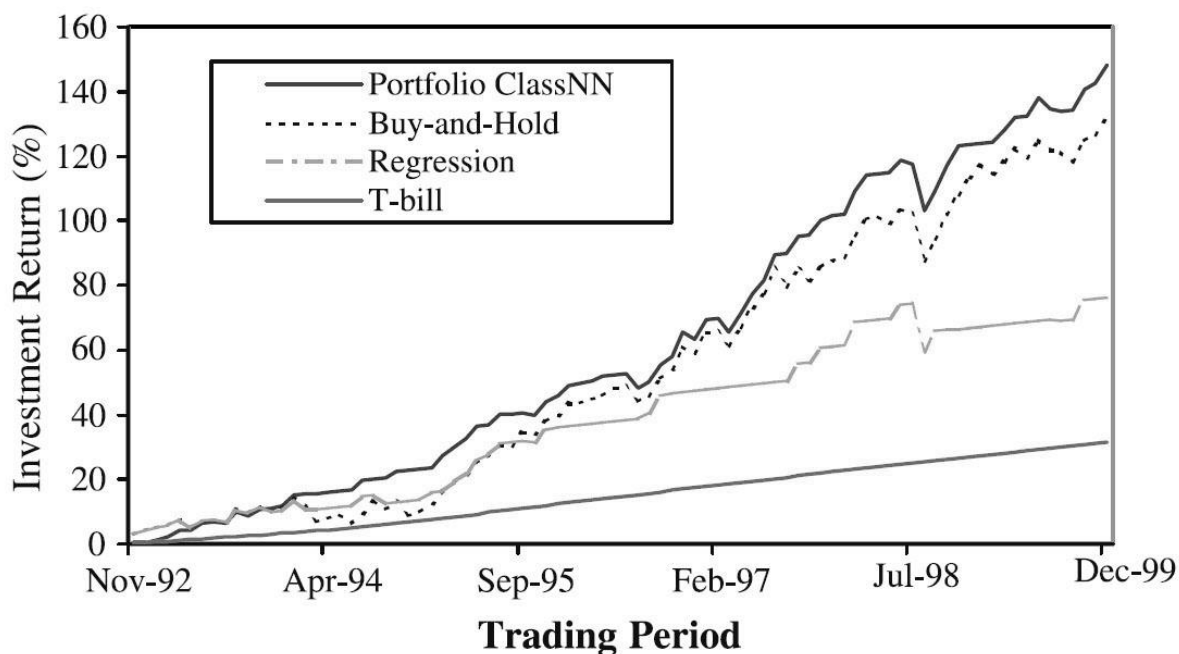
s ARIMA modelom na umelých a rôznorodých reálnych dátach z USA ako maloobchodné tržby, priemyselná sériová výroba, súkromné vlastníctvo bytových jednotiek atd. Tabuľka 1 zobrazuje výsledky testov z [8] kde je jasne vidno že neurónové siete prekonali model ARIMA.

Data	Data type ^a	Model	Training			Validation			Testing		
			RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
Iptotal	O	ANN	0.81	0.62	1.08	1.22	0.97	0.66	1.64	1.41	1.01
	DT	ANN	0.81	0.62	1.09	1.24	1.03	0.70	1.62	1.32	0.94
	DS	ANN	0.54	0.39	0.68	0.73	0.63	0.43	1.14	1.03	0.74
	DSDT	ANN	0.47	0.35	0.62	0.76	0.62	0.43	0.85	0.71	0.50
	O	ARIMA							8.94	8.05	5.79

^a O = original, DT = detrended, DS = deseasonalized, DSDT = deseasonalized and detrended.

Tabuľka 1 Výsledky testov porovnania ARIMA a NN z [8]

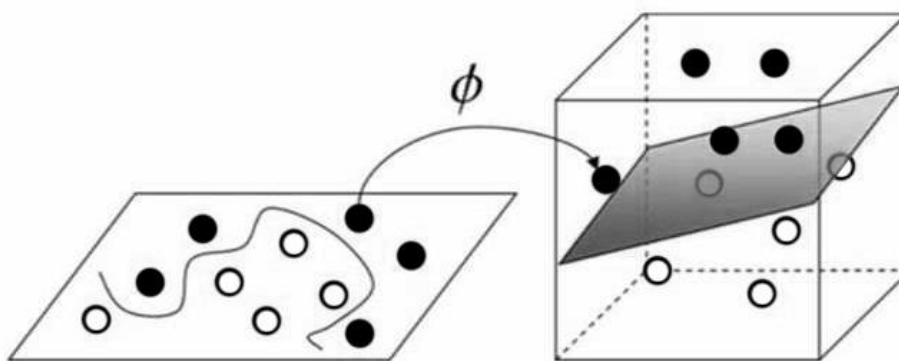
Enke a Thawornwong sa venovali v [10] porovnaniu NN s modelom lineárnej regresie a tradičnými stratégiami ako Buy-and-Hold a T-Bill index. Obrázok 2 zobrazuje návratnosť testovaných obchodných stratégií vytvorených na základe predikčných modelov na vybraných akciových trhoch za obdobie november 1992 až december 1999. NN prekonali všetky metódy pri rovnakej investičnej neistote.



Obrázok 2 Výsledky návratnosti pri teste NN, lineárnej regresie a tradičných stratégií z [10]

3.3 Predikcia pomocou *Support Vector Machines*

Support Vector Machines (SVM) sú metódy strojového učenia, sú veľmi blízke neurónovým sieťam. SVM boli vyvinuté Vapnikom a kolektívom v [11]. Základným využitím SVM je binárna klasifikácia, ale rýchlo sa našli aplikácie v ostatných odvetviach, ako aproximácia funkcie či problém predikcie časového radu. SVM hľadá nadrovinu, ktorá v priestore príznakov optimálne rozdeľuje trénovacie dáta. Optimálna nadrovina je taká, že body ležia v opačných polopriestoroch a hodnota minima vzdialenosti bodov od roviny je čo najväčšia. Na popis nadroviny stačia najbližšie body, tieto body sa nazývajú podporné vektory (*support vectors*). Nadrovina je lineárna funkcia v priestore príznakov. Dôležitou súčasťou je jadrová transformácia (*kernel transformation*) priestoru príznakov dát do priestoru transformovaných príznakov, do vyššej dimenzie. Táto jadrová transformácia umožňuje previesť pôvodnú lineárne neseparovateľnú úlohu na úlohu lineárne separovateľnú, na ktorú je možné aplikovať optimalizačný algoritmus pre nájdenie rozdeľujúcej nadroviny. [11]

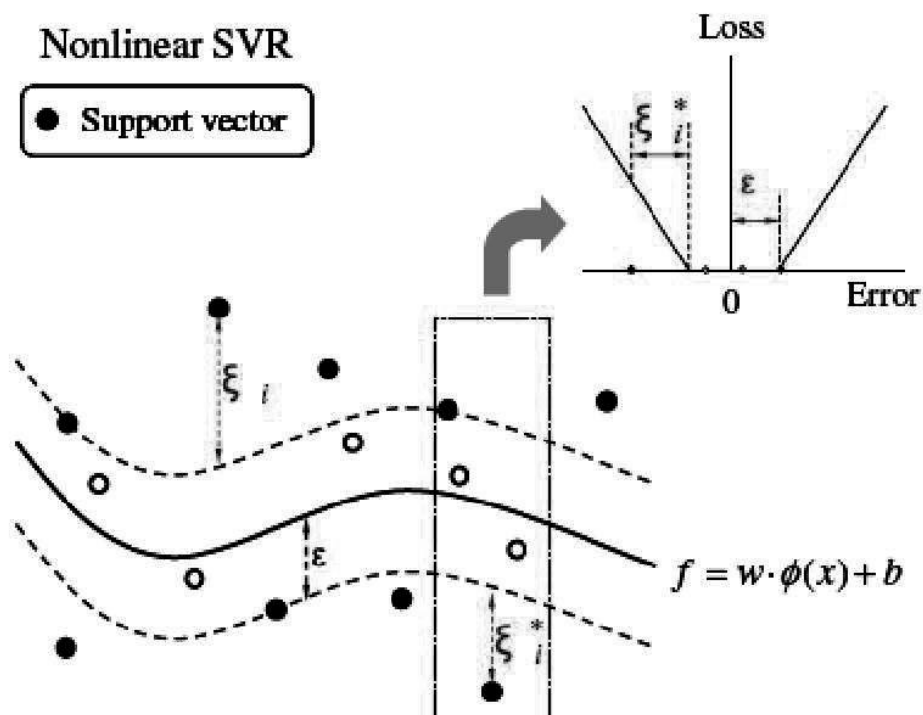


Obrázok 3 Transformácia z lineárne neseparovateľnej úlohy na lineárne separovateľnú [13]

Model SVM pre predikciu *Support vector regression* (SVR) možno vyjadriť ako:

$$y = f_{(x)} = w\phi(x) + b,$$

kde y je výstup modelu a $\phi(x)$ je viacrozmerný priestor, ktorý je nelineárne mapovaný zo vstupu x . Koeficienty w a b sú odhadnuté minimalizáciou. Obrázok 4 znázorňuje štruktúru takého modelu. Parametre, ktoré určujú nelineárne SVM, sú konštanty C (Cost), polomer rozpätia ϵ a parameter x . [12]



Obrázok 4 Model SVM nelineárnej regresie (SVR) [12]

Uplatnenie SVM metódy v úlohách regresnej analýzy teda spočíva vo vyhľadávaní optimálneho pásma k regresnej funkcii a na definovaní stratovej funkcie. Prostredníctvom optimálneho pásma sa reguluje veľkosť chyby aproximácie. Preto sa toto pásmo nazýva chyba aproximácie pozorovaných dát prostredníctvom regresnej funkcie a môžeme ho chápať ako mieru chýb regresie. Zmyslom zavedenia stratovej funkcie je stanovenie miery presnosti aproximácie, ktorou sa určuje priebeh zanedbávania malých chýb v rámci vymedzeného pásma. Správne určenie regresnej funkcie a nastavenie SVM, závisí od nájdenia optimálnych parametrov.

Tay a Cao v [14] porovnávali metódu predikcie SVM s BPNN. Ukázali že SVM prekonala neurónovú sieť, čo je vidno v tabuľke 2.

Futures	SVMs				BP			
	NMSE	MAE	DS	WDS	NMSE	MAE	DS	WDS
CME-SP	0.9365	0.2361	58.29	0.7380	1.0485	0.2556	55.27	0.8273
CBOT-US	1.2970	0.3425	45.22	0.9039	1.3424	0.3609	44.72	0.9882
CBOT-BO	1.1349	0.2989	46.73	0.9563	1.1877	0.3149	41.70	1.1093
EUREX-BUND	1.3337	0.3502	43.21	1.0542	1.3478	0.3520	41.20	1.0977
MATIF-CAC40	1.2083	0.4105	45.22	0.9827	1.2379	0.4224	42.22	1.0960

Tabuľka 2 Výsledky testov porovnania BPNN a SVM z [14]

Taktiež Kim v [15] porovnával SVM s BPNN a CBR na denných cenách Kórejských akcií. Výsledky ukázali že SVM prekonala BPNN aj CBR.

SVM systémy vykazujú lepšie výsledky pravdepodobne kvôli SRM princípu vedúcemu k lepšiemu zovšeobecneniu.

4 Metóda predikcie časového radu založená na strunových invariantoch

Nový predikčný model založený na strunových invariantoch, anglicky *Prediction Model Based on String Invariants* (PMBSI) vysvetlený a použitý na experimentálnu predikciu umelých a finančných dát v [16].

4.1 Definovanie strunových máp

Pri použití štandardnej metodiky odstránenia trendu bol prevedený pôvodný časový rad $p_{(\tau)}$ následovne:

$$\frac{p_{(\tau+h)} - p_{(\tau)}}{p_{(\tau+h)}}, \quad (4.1)$$

kde h označuje oneskorenie medzi $p_{(\tau)}$ a $p_{(\tau+h)}$, τ je index elementu časového radu.

Použitím teórie strún definujeme otvorenú strunovú mapu s jedným koncovým bodom:

$$P^{(1)}_{(\tau,h)} = \frac{p_{(\tau+h)} - p_{(\tau)}}{p_{(\tau+h)}}, \quad h \in \langle 0, l_s \rangle \quad (4.2)$$

kde index (1) predstavuje počet koncových bodov a l_s je dĺžka struny, l_s je celé kladné číslo a h je premenná obmedzená veľkosťou struny l_s . Prirodzeným dôsledkom transformácie (4.2) je splnenie hraničnej podmienky

$$P^{(1)}_{(\tau,0)} = 0, \quad (4.3)$$

ktorá platí pre každé τ . Pre zvýraznenie efektu významných udalostí použijeme umocnený deformovaný Q model

$$P^{(1)}_{(\tau,h)} = \left(1 - \left[\frac{p_{(\tau)}}{p_{(\tau+h)}} \right]^Q \right), \quad Q > 0 \quad (4.4)$$

Prvý koncový bod struny definuje pôvod a odráža lineárny trend v $p(\cdot)$ na rozhraní l_s . Prítomnosť dlhodobého trendu je čiastočne upravená $P^{(2)}_{(\tau,h)}$ kde $h=l_s$. Otvorená struna s dvoma koncovými bodmi sa zavádza cez nelineárnu mapu, ktorá kombinuje informácie o trende z p v dvoch po sebe idúcich segmentoch

$$P^{(2)}_{(\tau,h)} = \left(1 - \left[\frac{p(\tau)}{p(\tau+h)}\right]^Q\right) \left(1 - \left[\frac{p(\tau+h)}{p(\tau+l_s)}\right]^Q\right), \quad h \in \langle 0, l_s \rangle \quad (4.5)$$

Mape priradíme okrajové hodnoty, Dirichletové hodnoty

$$P^{(2)}_{(\tau,0)} = P^{(2)}_{(\tau,l_s)} = 0, \quad (4.6)$$

ktorá platí pre každé τ .

Znamienko $P^{(2)}_{(\tau+h)}$ obzvlášť zahŕňa informácie o rozdieloch správania sa $p(\cdot)$ v troch bodoch $(\tau, \tau+h, \tau+l_s)$. $P^{(2)}_{(\tau+h)} < 0$ predstavuje trend rozdielných znamienok a naopak $P^{(2)}_{(\tau+h)} > 0$ predstavuje zhodné znamienka.

4.2 Strunové invarianty

Invariant znamená nemenný, znamená to že niečo sa transformáciou nemení, ako napríklad úprava rovníc do iného referenčného tvaru. Táto myšlienka je aplikovaná na predikciu časového radu hľadaním invariantov v dátach a ich využitím na predikciu budúcich hodnôt. Podobný výskum zameraný na nájdenie invariantných stavov na finančnom trhu je opísaný v [19].

l_{pr} predstavuje počet krokov pre predikciu od hodnoty τ_0 . Λ je pomocné kladné celé číslo a musí byť splnená podmienka

$$\Lambda = l_s - l_{pr}, \quad l_s > l_{pr} \quad (4.7)$$

Sila nelineárnych dát strunových máp časových radov musí byť využitá pre vytvorenie modelu pre predikciu obdobne ako v [20], [21], [22]. Model obsahuje zmiešané struny s 2 koncovými bodmi kde jedna struna je neustále deformovaná na druhú.

Najprv potrebujeme definovať prístup vytvorenia strunových invariantov. Korelačná funkcia je štatistická korelácia medzi náhodnými premennými v dvoch rôznych bodoch, v našom prípade struny v časových radoch. Pre jednoduchosť je použitá iba struna s jedným koncovým bodom. Rovnica (4.4) s parametrom $Q=1$. Bežná korelačná funkcia je definovaná

$$C_{(\tau, l_0)} = \langle P^1_{(\tau, l_0)} P^1_{(\tau+1, l_0)} \rangle \quad (4.8)$$

Invariant vo forme korelačnej funkcie

$$C_{(\tau, l_0)} = \sum_{h=l_0}^{h=l} W_{(h)} \left(1 - \frac{p_{(\tau-h)}}{p_{(\tau-1-h)}} \right) \left(1 - \frac{p_{(\tau-1-h)}}{p_{(\tau-2-h)}} \right), \quad (4.9)$$

s váhou $W_{(h)}$, čo je vybraný bimodálny parameter

$$W_{(h)} = \begin{cases} 1 - W_0, & h \leq l_s/2, \\ W_0, & h > l_s/2, \end{cases} \quad (4.10)$$

kde W_0 je parameter váhy, Nasledujúca formulácia nie je jediná ani ideálna

$$W_0 = \frac{1}{\sum_{h'=0}^{l_s} e^{-h'/\Lambda}}. \quad (4.11)$$

Rodina invariantov je zapísaná použitím parametrizácie

$$\begin{aligned} C_{(\tau, \Lambda)} &= (1 - \eta_1)(1 - \eta_2) \sum_{h=0}^{\Lambda} W_{(h)} \\ &\times \left(1 - \left[\frac{p_{(\tau)}}{p_{(\tau+h)}} \right]^Q \right) \left(1 - \left[\frac{p_{(\tau+h)}}{p_{(\tau+l_s)}} \right]^Q \right) \\ &+ \eta_1(1 - \eta_2) \sum_{h=0}^{\Lambda} W_{(h)} \left(1 - \left[\frac{p_{(\tau)}}{p_{(\tau+h)}} \right]^Q \right) \\ &+ \eta_2 \sum_{h=0}^{\Lambda} W_{(h)} \left(1 - \left[\frac{p_{(\tau+h)}}{p_{(\tau+l_s)}} \right]^Q \right), \end{aligned} \quad (4.12)$$

kde $\eta_1 \in (-1; 1)$, $\eta_2 \in (-1; 1)$ sú homotopné premenné, Q je parameter reálneho čísla.

4.3 Tvorba predikcie

Predpokladáme podmienku invariance medzi uzatváracou strunou v čase τ a v ďalšom kroku $\tau + 1$ v časovom rade vo forme

$$C_{(\tau, l_0)} = C_{(\tau+1, l_0)}. \quad (4.13)$$

Presný výraz pre jednokrokovú predikciu $p_{(\tau+1)}$ dostaneme vyhodnotením jednokrového korelačného invariantu rovnice (4.12) s počiatočnou podmienkou $l_0 = 0$.

$$\begin{aligned} W_{(0)} \left(1 - \frac{p_{(\tau)}}{p_{(\tau-1)}} \right) \left(1 - \frac{p_{(\tau-1)}}{p_{(\tau-2)}} \right) = \\ W_{(0)} \left(1 - \frac{p_{(\tau+1)}}{p_{(\tau)}} \right) \left(1 - \frac{p_{(\tau)}}{p_{(\tau-1)}} \right) + \\ W_{(1)} \left(1 - \frac{p_{(\tau)}}{p_{(\tau-1)}} \right) \left(1 - \frac{p_{(\tau-1)}}{p_{(\tau-2)}} \right), \end{aligned} \quad (4.14)$$

ktorá môže byť upravená do podoby

$$C_{(\tau, 0)} = W_{(0)} \left(1 - \frac{p_{(\tau+1)}}{p_{(\tau)}} \right) \left(1 - \frac{p_{(\tau)}}{p_{(\tau-1)}} \right) + C_{(\tau+1, 1)} \quad (4.15)$$

a

$$\left(1 - \frac{p_{(\tau+1)}}{p_{(\tau)}} \right) = \frac{C_{(\tau, 0)} - C_{(\tau+1, 1)}}{W_{(0)} \left(1 - \frac{p_{(\tau)}}{p_{(\tau-1)}} \right)}. \quad (4.16)$$

Nakoniec postupnou úpravou dostávame jednokrokovú predikciu.

$$p_{(\tau+1)} = p_{(\tau)} \left(1 + \frac{C_{(\tau+1, 1)} - C_{(\tau, 0)}}{W_{(0)} \left(1 - \frac{p_{(\tau)}}{p_{(\tau-1)}} \right)} \right), \quad (4.17)$$

platné pre $p_{(\tau)} \neq p_{(\tau-1)}$. To sú základné definície pre jednokrokovú predikciu korelačných invariantov. Podobným spôsobom je zostavená nasledujúca predikcia $p_{(\tau_0 + l_{pr})}$ pre model so

zmiešanými strunami 1 a 2 koncovými bodmi s $Q > 0$. $p_{(\tau_0+l_{pr})}$ je vyjadrená pomocnými premennými:

$$A_1(\Lambda, \tau) = (1 - \eta_1)(1 - \eta_2) \sum_{h=0}^{\Lambda} W_{(h)} \left(1 - \left[\frac{p_{(\tau)}}{p_{(\tau+h)}} \right]^Q \right), \quad (4.18)$$

$$A_2(\Lambda, \tau) = -(1 - \eta_1)(1 - \eta_2) \sum_{h=0}^{\Lambda} W_{(h)} \left(1 - \left[\frac{p_{(\tau)}}{p_{(\tau+h)}} \right]^Q \right) p_{(\tau+h)}^Q, \quad (4.19)$$

$$A_3(\Lambda, \tau) = \eta_1(1 - \eta_2) \sum_{h=0}^{\Lambda} W_{(h)} \left(1 - \left[\frac{p_{(\tau)}}{p_{(\tau+h)}} \right]^Q \right), \quad (4.20)$$

$$A_4(\Lambda, \tau) = \eta_2 \sum_{h=0}^{\Lambda} W_{(h)}, \quad (4.21)$$

$$A_5(\Lambda, \tau) = -\eta_2 \sum_{h=0}^{\Lambda} W_{(h)} p_{(\tau+h)}^Q. \quad (4.22)$$

Teda očakávaná predikcia predstavuje

$$p_{(\tau_0+l_{pr})} = \left[\frac{A_2(\Lambda, \tau') + A_5(\Lambda, \tau')}{C_{(\tau_0-l_s)} - A_1(\Lambda, \tau') - A_3(\Lambda, \tau') - A_5(\Lambda, \tau')} \right]^{\frac{1}{Q}}, \quad (4.23)$$

kde $\tau' = \tau_0 + l_{pr} - l_s$, ($\tau' = \tau_0 - \Lambda$). Odvodenie je založené na invariancii

$$C_{(\tau, l_s - l_{pr})} = C_{(\tau - l_{pr}, l_s - l_{pr})}, \quad (4.24)$$

a model bude efektívny ak

$$C_{(\tau_0, \Lambda)} \simeq C_{(\tau_0+l_{pr}, \Lambda)}. \quad (4.25)$$

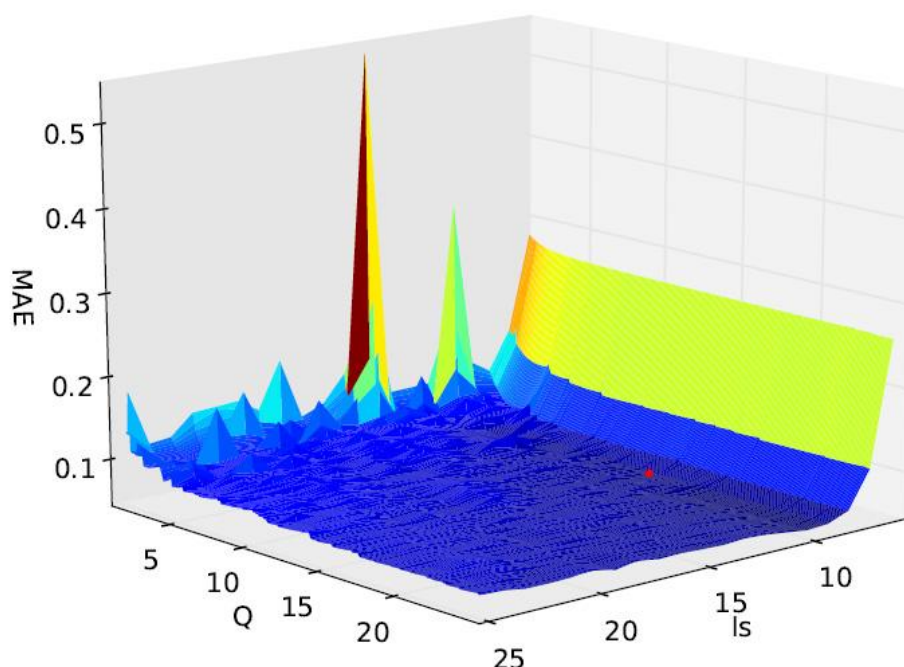
Model má voľné parametre l_s , η_1 , η_2 , Q . Tieto musia byť správne nastavené optimalizáciou. PMBSI vyžaduje časový rad s kladnými hodnotami. Pre záporné hodnoty predikcia nie je definovaná. Napriek tomu PMBSI občas vráti nedefinovanú hodnotu (NaN). Tento problém bol vyriešený nahradením NaN predikcie poslednou známou hodnotou pre $l_{pr}=1$ (naivná predikcia).

5 Návrh evolučného algoritmu pre optimalizáciu parametrov pre model predikcie časového radu založeného na strunových invariantoch

V tejto kapitole popíšem evolučné algoritmy a navrhнем evolučný algoritmus pre optimalizáciu parametrov pre metódu predikcie časového radu založenej na strunových invariantoch.

V našom prípade potrebujeme do modelu PMBSI vybrať 4 premenné l_s , η_1 , η_2 , Q . Pričom η_1 , η_2 , Q môžu byť reálne čísla z definovaného rozsahu, preto náš problém má nekonečne veľa možných riešení. Pri hodnotení vhodnosti jednotlivých prehľadávaných riešení použijeme chybu merania, MAE – strednú absolútnu chybu definovanú rovnicou (6.1).

Obrázok 5 zobrazuje MAE chybu s 5 krokovou predikciou z experimentu na finančných dátach z článku [16]. Experiment testoval predikciu PMBSI s kombináciou hodnôt parametrov l_s a Q , pričom η_1 , $\eta_2 = 0$.



Obrázok 5 MAE chyba s 5 krokovou predikciou z experimentu z článku. Červená bodka predstavuje globálne minimum nájdených riešení. [16]

Optimálne nastavenia boli nájdené testovaním všetkých kombinácií Q , l_s z definovaného rozsahu vzoriek. Na obrázku je vidno že najlepšie riešenie nie je ľahké nájsť keďže v priestore prehľadávania existuje veľa lokálnych miním. Výkon PMBSI je približne rovnaký pre veľký rozsah riešení. Pre nájdenie najlepších hodnôt premenných l_s , η_1 , η_2 , Q sa evolučný algoritmus zdá byť najvhodnejší, kvôli paralelnému prehľadávaniu priestoru riešení a schopnosti prekonávať lokálne minimá.

5.1 Evolučné algoritmy, prvky a princípy

Evolučné algoritmy sú špeciálnym prípadom prehľadávacích algoritmov, konkrétne populačným, teda naraz skúma viacero kandidátov na riešenie. Každý skúmaný kandidát sa označuje ako jedinec, ktorého vnútorná štruktúra definuje jeho polohu v priestore kandidátov. Jedince, ktoré sú algoritmom skúmané v určitom čase tvoria populáciu. Na začiatku činnosti algoritmu je vytvorená prvotná populácia jedincov, ktorá postupom času prechádza vývojom v rámci evolučného cyklu. Čas t je meraný v diskrétnych jednotkách – generáciách. Každá generácia zodpovedá jednému prechodu evolučným cyklom – teda v každom prechode je pozornosť algoritmu sústredená na tú istú generáciu populácie jedincov, označovanú ako populácia v čase t alebo populácia t -tej generácie. Po vytvorení generácie náhodným spôsobom alebo tak, aby pokrývali priestor prehľadávania sa kontroluje podmienka ukončenia evolučného cyklu. Podmienkou môže byť, že hľadané riešenie už bolo nájdené, uplynutie maximálneho povoleného počtu generácií, zvýšenie stupňa konvergenzie populácie nad zvolený prah. Ak podmienka nebola splnená realizuje sa ďalší prechod evolučným cyklom, vytvorí sa nová generácia jedincov použitím predošlej generácie. Najprv sú selekciou identifikované jedince z aktuálnej populácie (rodičia), ktorých vlastnosti sú považované za dostatočne atraktívne pre to, aby boli prenesené na nových jedincov. Rodičia môžu byť vybraný raz, viackrát alebo vôbec. Následne pomocou genetických operátorov (kríženie, mutácia) sú generované nové jedince z rodičov (potomkovia). Určí sa vhodnosť potomkov a na záver sa nahradia rodičia, všetci alebo časť, a vytvorí sa nová populácia, ostatné jedince upadajú do zabudnutia. [23]

Evolučný algoritmus pozostáva z viac častí, ktoré treba analyzovať a navrhovať pri definovaní algoritmu. Pri návrhu je potrebné analyzovať tieto komponenty algoritmu:

- Reprezentácia jedincov
- Populácia

- Selekcia jedincov
- Genetické operátory
- Náhrada jedincov

5.2 Reprezentácia jedincov

Jedince musia byť vhodným spôsobom reprezentované, čiže kódované, pre účelnú manipuláciu algoritmu s jedincami. Neexistuje všeobecný návod, ako zvoliť správnu reprezentáciu. Pomocou použitej reprezentácie definujeme priestor prehľadávania a tým zároveň stanovujeme:

- Veľkosť prehľadávaného priestoru
- Vzťah prehľadávaného priestoru k priestoru kandidátov riešení
- Požiadavky na genetické operátory

Čím je väčší priestor prehľadávania, tým je potrebné väčšie úsilie na prehľadanie tohto priestoru. Vzťah medzi prehľadávaným priestorom a priestorom kandidátov je sprostredkovaný použitým spôsobom dekódovania. Konečný návrh reprezentácie je ovplyvnený zložitou dekódovania a vzťahom medzi reprezentáciami riešení a samotnými riešeniami. [23]

Pre PMBSI je potrebné optimalizovať 4 parametre reálnych čísel, tak aj schéma pre kódovanie jedincov je reálna s pevnou dĺžkou a pevným usporiadaním zložiek, čiže jedinec je reprezentovaný ako vektor reálnych čísel. Konkrétne parametre z intervalov: $l_s \in [l_{pr} + 1; \maxStringLength]$, $\eta_1 \in (-1; 1)$, $\eta_2 \in (-1; 1)$, $Q \in (0; 25)$, kde spodná hranica pre l_s je definovaná kvôli splneniu rovnice 4.7 a horná hranica je pri každom experimente definovaná zvlášť, η_1 a η_2 sú ohraničené na základe parametrizácie z rovnice 4.12. Dolná hranica Q je definovaná rovnicou 4.4 a hornú som ponechal z experimentov z článku [16].

5.3 Selekcia jedincov

Selekcia alebo vzorkovanie populácie slúži na výber rodičov z aktuálnej populácie jedincov, pomocou ktorých bude vytvorená nová populácia potomkov. Výber jedincov sa realizuje na základe šancí na reprodukciu. Jedince, ktorých šanca na reprodukciu je vysoká, môžu byť medzi rodičov zaradení viackrát, teda jedinec s väčšou šancou na reprodukciu sa stáva častejšie rodičom ako jedinec s menšou šancou. Pre určenie šancí na reprodukciu sa používa vhodnosť jedincov. Existujú 2 triedy selekcie [23]:

- Vzorkovanie s explicitnými pravdepodobnosťami
- Vzorkovanie s implicitnými pravdepodobnosťami
 - Turnaje
 - Orezanie
 - Náhodný výber

Pre evolučný algoritmus pre PMBSI som v tejto práci zvolil vzorkovanie s implicitnými pravdepodobnosťami, konkrétne q-árny turnaj. Jedince spolu súťažia a o víťazovi rozhoduje vhodnosť tých jedincov, ktorí sa zúčastňujú turnaja. Výhodou turnajov oproti orezaniu je, že sa nevyžaduje aby jedince boli zotriedené podľa vhodnosti. Odpadá fáza triedenia jedincov a znižuje sa časová náročnosť, čo bol tiež dôvod voľby tohto druhu selekcie. Princíp použitej metódy:

Z populácie sa náhodne vyberie definovaný počet jedincov. Z nich sa vyberú dvaja s najlepšou vhodnosťou a stávajú sa rodičmi. Z týchto rodičov sa prostredníctvom genetických operátorov vytvorí nový potomok. Táto selekcia prebehne toľko krát aby bol počet potomkov rovný počtu rodičov.

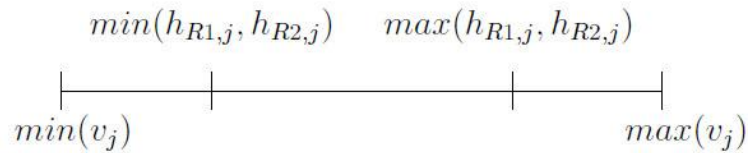
5.4 Genetické operátory

Delenie genetických operátorov je založené na počte rodičov potrebných pre vygenerovanie nových jedincov. Z tohto hľadiska rozlišujeme 3 triedy operátorov [24]:

- Asexuálne operátory – potomok má iba jedného rodiča
- Sexuálne operátory – potomok má dvoch rodičov
- Panmiktické operátory – potomok má viac ako dvoch rodičov

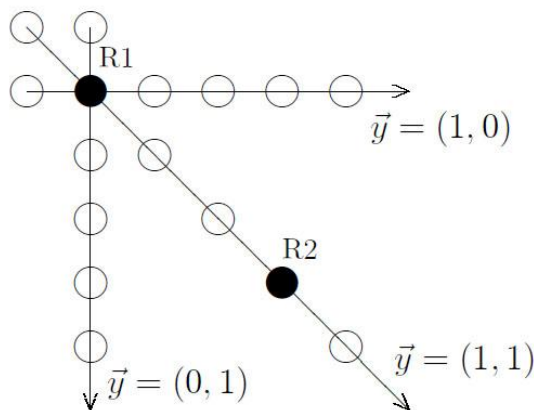
Asexuálne operátory sa zvyknú nazývať mutačný operátor resp. mutácia a sexuálny operátor ako krížiaci operátor resp. kríženie. Tieto dva typy sa používajú v praktických aplikáciách najčastejšie. Rovnako v evolučnom algoritme pre nájdenie najvhodnejších parametrov pre PMBSI som použil tieto operátory. Dvojica s najlepšou vhodnosťou zvolená selekciou je spracovaná krížiacim operátorom a následne výsledok kríženia je spracovaný mutačným operátorom.

Obrázok 6 zobrazuje kríženie pre reálne kódovanie. **Chyba! Nenalezen zdroj odkazů..** Na j -tej pozícii každého jedinca je prípustná iba hodnota z intervalu $\langle \min(v_j), \max(v_j) \rangle$ a prvý rodič má na tejto pozícii hodnotu $h_{R1,j}$, a druhý rodič $h_{R2,j}$. [23]



Obrázok 6 Alternatívy pre voľbu reálnej hodnoty pri krížení [23]

Použil som aritmetické kríženie. Obidvaja rodičia sú chápaní ako body v mnohorozmernom priestore. Potomok je určený ako nový bod v tomto priestore, pričom pre jeho určenie sa používajú aritmetické operácie s hodnotami rodičov. Obrázok 7 znázorňuje tento typ kríženia, pričom prázdne krúžky znázorňujú možné polohy potomkov.



Obrázok 7 Ilustrácia aritmetického kríženia [23]

Najjednoduchším typom aritmetického kríženia je lineárna kombinácia hodnôt rodičov. Potomok sa určuje podľa vzťahu

$$h'_j = x h_{R1j} + (1 - x) h_{R2j} \quad (5.1)$$

kde x je náhodné číslo z intervalu $\langle 0,1 \rangle$. Tento typ kríženia dokáže produkovať potomkov na úsečke, ohraničenej bodmi reprezentujúcimi oboch rodičov.

Po krížení prebieha mutácia. Mutácia sa vykonáva podobným spôsobom ako kríženie. Potomok sa mutuje podľa vzťahu

$$h'_j = h_j + x(\max(v_j) - \min(v_j))$$

kde x je náhodné číslo z intervalu $<-1,1>$. Hodnota jedinca sa zvýši/zníži maximálne na okraj intervalu dovolený pre parameter. Veľkosť mutácie sa postupne rovnomerne znižuje k 0, čiže generácia bude konvergovať k jednému jedincovi s najlepšou vhodnosťou. Ak sa v ďalšej generácii nájde jedinec s lepšou vhodnosťou ako najlepší v predošlej, mutácia sa začne odznova znižovať z maximálnej možnej mutácie definovanej intervalom parametra. Mutácia slúži nato, aby sme zabránili stagnácii v lokálnom extréme.

5.5 Náhrada jedincov

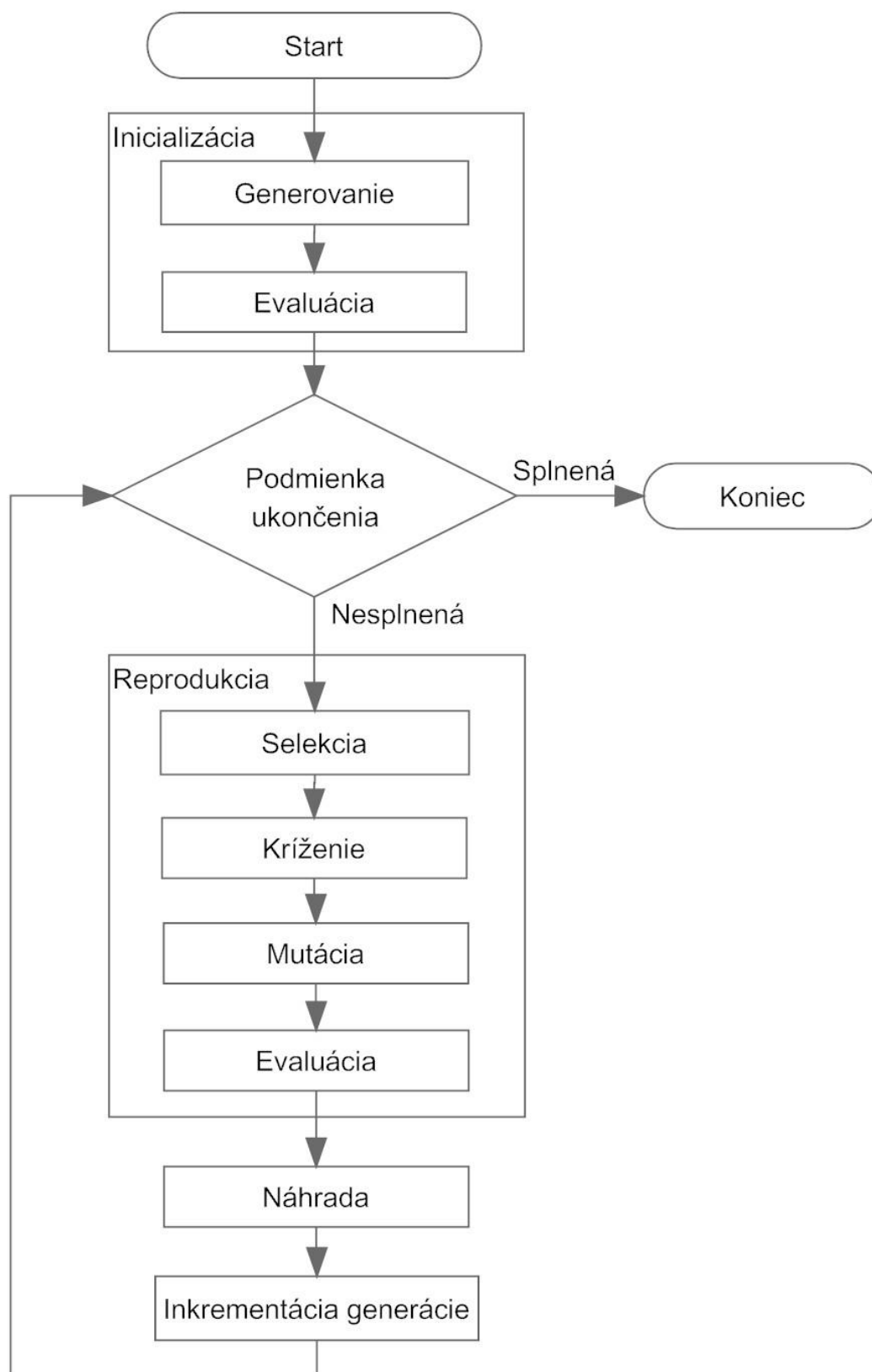
Pri náhrade dochádza k formovaniu novej generácie jedincov z jedincov, ktorí sú k dispozícii. Sú to jedince, ktoré vytvárajú pôvodnú populáciu v t-tej generácii a nové jedince vytvorené genetickými operáciami, čiže rodičia a potomkovia. Z týchto dvoch skupín sa vhodným spôsobom vyberajú jedince, ktoré budú reprezentované v novej populácii. Pri náhrade môže byť každý z jedincov vybraný maximálne raz. [23]

Spôsoby náhrady:

- Jednogeneračná – v novej populácii sa použijú iba novo vytvorení potomkovia, teda potomkovia kompletne nahradia rodičov
- Generácia z potomkov aj rodičov
 - Oddelenou súťažou – je dané, koľko miest v novej generácii obsadia rodičia a koľko potomkovia
 - Spoločnou súťažou – všetci jedinci, rodičia aj potomkovia, súťažia medzi sebou

Pri náhrade v navrhnutom evolučnom algoritme som použil zachovanie 1% rodičov s najlepšou vhodnosťou, čiže najlepšieho jedinca s najnižšou MAE.

5.6 Vývojový diagram navrhnutého evolučného algoritmu



Obrázok 8 Vývojový diagram navrhnutého evolučného algoritmu

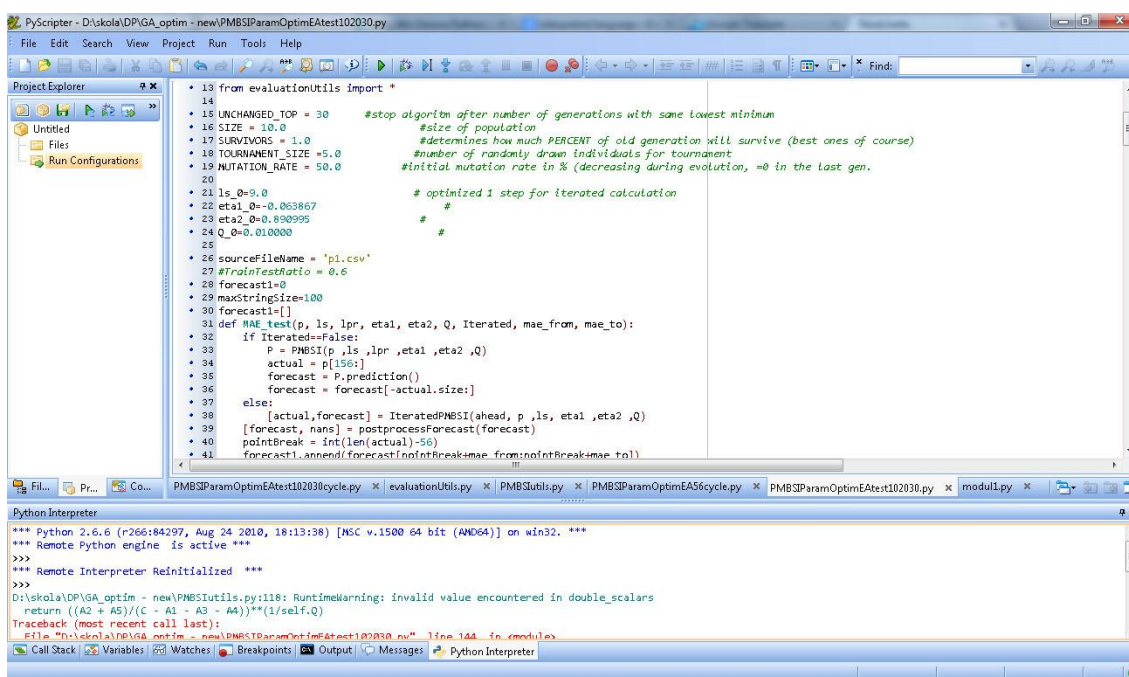
Popis jednotlivých blokov vývojového diagramu:

- *Generovanie* – vytvorenie nulte generácie $t=0$, parametre jedincov sú zvolené náhodne z intervalov.
- *Evaluácia* – vyhodnotenie populácie, výpočet vhodnosti, MAE z tréovacích dát pre každého jedinca
- *Podmienka ukončenia* – evolučný algoritmus sa ukončí, ak za definovaný počet generácií sa nenájde jedinec s lepšou vhodnosťou, teda nižším MAE, než v predošlej generácii,
- *Selekcia* – výber 2 jedincov turnajom z terajšej generácie
- *Kríženie* – aritmetické kríženie zvolených jedincov zo selekcie
- *Mutácia* – mutácia jedincov
- *Náhrada* – náhrada starej populácie novou, tvorenou 1% najlepších rodičov a 99% najlepších potomkov, vlastne nahradenie celej populácie až na najvhodnejšieho jedinca
- *Inkrementácia generácie* – zvýšenie generácie $t=t+1$

5.7 Vývojové prostredie pre návrh evolučného algoritmu

Na programovú realizáciu evolučného algoritmu som si vybral programovací jazyk Python pretože je stabilný, výkonný a potrebuje relatívne malé množstvo textu pre prácu s časovým radom a hlavne predikčný model PMBSI dodaný vedúcim DP je vytvorený v tomto jazyku. Python je tiež odporúčaný ako štartovací programovací jazyk pre jeho intuitívne používanie a potrebe používania odsadenia čo učí robiť program prehľadnejším.[26]

Ako vývojové prostredie som si zvolil open source program PyScripter, zobrazený na obrázku 9, pre prehľadné zobrazenie a jednoduché odstránenie chýb. Bližšie špecifikácie prostredia a samotného navrhnutého kódu sú v prílohe Systémová a používateľská príručka.



Obrázok 9 Vývojové prostredie PyScripter

6 Experimenty

V tejto kapitole porovnáam výsledky predikcie s výsledkami dosiahnutými v článku [16] na umelých dátach, konkrétne sínusoide rozdelenej na 51 pravidelne rozložených bodov a denných finančných dátach v rozsahu 1295 dní. Následne vykonám predikciu 111 časových radov zo súťaže „2008 Time Series Forecasting Competition for Computational Intelligence“ pre porovnanie s navrhnutým modelom. Všetkých 111 časových radov obsahujú 775 hodnôt, z ktorých posledných 56 je potrebné predpovedať.

6.1 Úvodné poznámky k experimentom

Pri experimentoch na umelých dátach a finančných dátach bol výkon PMBSI porovnaný s SVM a naivnou predikciou. Sú použité 2 chyby merania, MAE – stredná absolútna chyba a SMAPE – symetrická stredná absolútna percentuálna chyba, definované:

$$MAE = \frac{1}{n} \sum_{t=1}^n |A_t - F_t|, \quad (6.1)$$

$$SMAPE = \frac{100}{n} \sum_{t=1}^n 0,5 \frac{|A_t - F_t|}{|A_t| + |F_t|}, \quad (6.2)$$

kde n je počet vzoriek, A_t je skutočná hodnota a F_t je predpovedaná hodnota.

Výsledky v súťaži „2008 Time Series Forecasting Competition for Computational Intelligence“ boli porovnávané na základe SMAPE vzorca (6.2).

Nastavenia evolučného algoritmu som zisťoval metódou pokus omyl, teda hľadal som nastavenia ktoré dokážu nájsť parametre s čo najlepšou vhodnosťou, ale aby evolúcia netrvala zbytočne dlho. Zvolené nastavenia boli vhodné pre všetky experimenty. Parametre som nastavil nasledovne:

1. počet generácií, po ktorých ukončiť algoritmus, ak sa nenájde jedinec s lepším MAE, ako bolo dosiahnuté v predošlej generácii, som nastavil na 50,
2. veľkosť populácie na 20,
3. zachovanie rodičov z predošlej generácie som zvolil 1 %,
4. veľkosť turnaja pre kríženie nastavil na 5 jedincov,

5. veľkosť mutácie krížených jedincov som nastavil na max 50%, čo sa znižuje počtom generácií, pre ktoré sa nenašlo nižšie MAE, až na 0%. Po nájdení nižšieho MAE lepšej vhodnosti sa mutácia vracia na max 50%.

6.2 Experiment na umelých dátach

Sínusoida bola rozdelená na 2 časti a to tak, že pozitívna polovica periódy bola použitá na tréning a vyhodnotenie. Negatívna polovica bola použitá pre kontrolu. Pre metódu PMBSI boli dáta posunuté nad nulu pričítaním konštanty, ktorá bola potom odčítaná od predikcie. Výsledky PMBSI predikcie s parametrami l_s , η_1 , η_2 , Q získanými evolučnou optimalizáciou sú porovnané s SVM s lineárnym jadrom a naivnou predikciou na 1,2 a 3 krokovej predikcii z článku [16]. Tabuľka 3 zobrazuje výsledky experimentu a porovnanie s výsledkami z článku. Najlepšie výsledky sú zvýraznené. Evolučná optimalizácia dokázala efektívne a rýchlo nájsť vhodné parametre pre model PMBSI v porovnaní s článkom.

Metóda	l_{pr}	MAE tréning		MAE test		SMAPE test	
		článok	EA optim	článok	EA optim	článok	EA optim
PMBSI	1	0.000973	0.000278	0.002968	0.002197	8.838798	8.656631
	2	0.006947	0.001416	0.034032	0.013792	14.745538	11.065498
	3	0.015995	0.004247	0.161837	0.061837	54.303315	25.692156
Iterovaný PMBSI	1	-	-	-	-	-	-
	2	0.003436	0.001057	0.011583	0.009102	10.879313	10.101545
	3	0.008015	0.002455	0.028096	0.023102	14.047025	12.635537
SVM	1	0.011831		0.007723		10.060302	
	2	0.012350		0.007703		10.711573	
	3	0.012412		0.007322		11.551324	
Naivná predikcia	1	-		0.077947		25.345352	
	2	-		0.147725		34.918149	
	3	-		0.207250		41.972591	

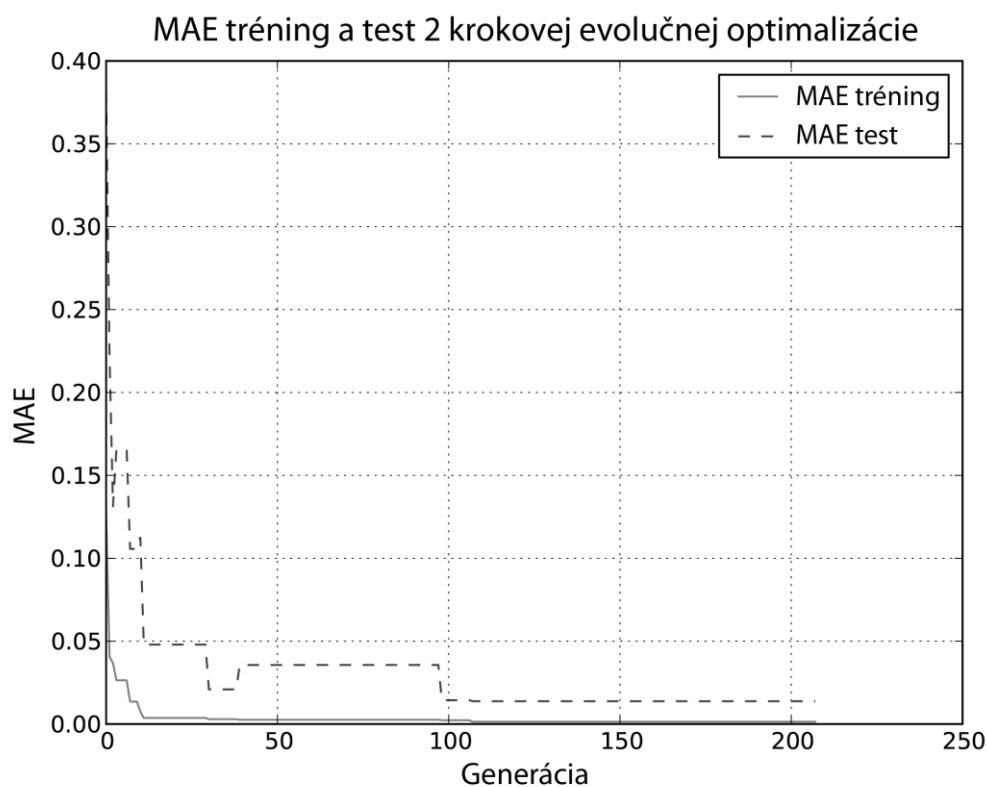
Tabuľka 3 Výsledky experimentu na umelých dátach

V tabuľke 4 sú zaznamenané parametre l_s , η_1 , η_2 , Q získané evolučnou optimalizáciou s časom trvania optimalizácie a potrebného počtu generácií. V tabuľke je vidno, že parametre Q a η_1 pre 1,2 a 3 krokovú predikciu sú rovnaké.

l_{pr}	L_s	Q	η_1	η_2	Generácia	Čas(s)
1	2.0	0.01	0.96	-0.2418061866	155	57.2
2	3.0	0.01	0.96	-0.1523626591	208	81.1
3	5.0	0.01	0.96	-0.3820427543	482	189.4

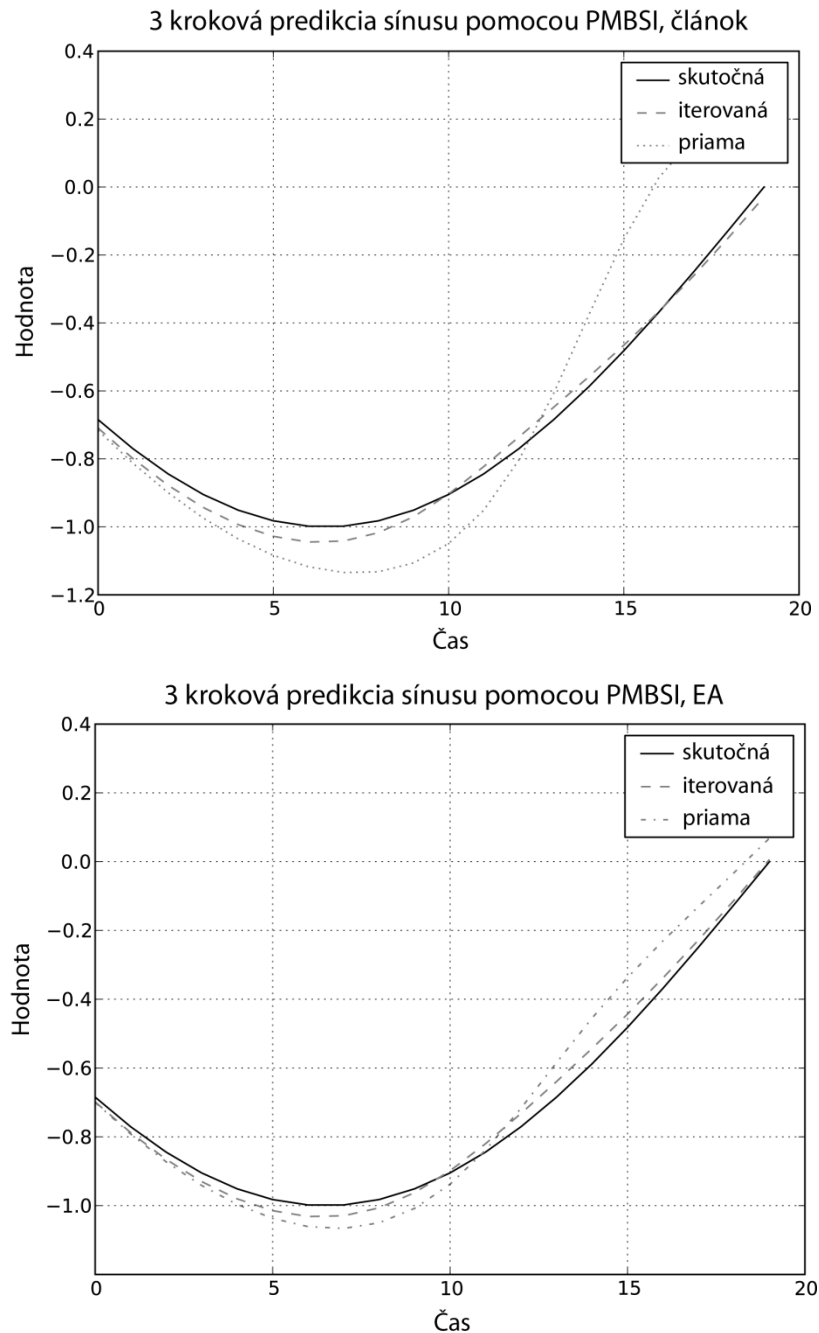
Tabuľka 4 Optimálne parametre PMBSI pre sínusoidu

Obrázok 10 zobrazuje postupnú evolúciu trénovacej a kontrolnej MAE pre 2 krokovú predikciu. Evolúcia sa chovala podobne pre 1,2 a 3 krokovú predikciu, začala na vyšších hodnotách a postupne bolo nájdené nižšie lokálne minimum.



Obrázok 10 Graf evolúcie trénovacej a testovacej MAE na sínusoide pre 2 krokovú predikciu

Obrázok 11 zobrazuje 3-krokovú predikciu sínusoidy pomocou priamej aj iterovanej PMBSI s parametrami optimalizovanými evolučným algoritmom a PMBSI experimentálnej predikcie z článku [16].



Obrázok 11 Porovnanie 3 krokovej predikcie PMBSI s EA a článkom [16]

6.3 Experiment na finančných dátach

Finančné dáta boli rozdelené na tréningovú a validačnú časť v pomere 6:4. Oproti sínusoide sú finančné dáta vysoko nelineárne a chaotické. Výsledky PMBSI predikcie s parametrami l_s , η_1 , η_2 , Q získanými evolučnou optimalizáciou sú porovnané s SVM

s Gaussovou RBF kernel funkciou a naivnou predikciou na 1 až 10-krokovej predikcii z článku [16]. V tabuľke 5 sú zaznamenané výsledky experimentu a porovnané s výsledkami z článku. Najlepšie výsledky sú zvýraznené. Evolučná optimalizácia dokázala efektívne a rýchlo nájsť vhodné parametre pre model PMBSI v porovnaní s článkom až na jednokrokovú predikciu, kde sa algoritmu nepodarilo nájsť nižšie lokálne minimum, no napriek tomu nedokázala prekonať model SVM.

Method	lpr	MAE tréning		MAE test		SMAPE test	
		článok	EA optim	článok	EA optim	článok	EA optim
PMBSI	1	0.023227	0.024094	0.023595	0.023799	7.380742	7.505595
	2	0.037483	0.034083	0.036335	0.033463	11.378275	10.442204
	4	0.048140	0.045598	0.046381	0.044731	14.876330	14.341023
	6	0.054556	0.051771	0.049755	0.052516	16.094349	17.196778
	8	0.057658	0.056242	0.056097	0.058517	18.546008	19.243273
	10	0.060192	0.058841	0.058216	0.054138	18.752986	18.004554
Iterovaný PMBSI	1	-		-		-	
	2	0.032706	0.034302	0.031940	0.033170	9.953547	10.375175
	4	0.043134	0.047085	0.042414	0.045690	13.250729	14.130408
	6	0.049916	0.056509	0.047784	0.054769	15.102693	16.930280
	8	0.055326	0.065350	0.051355	0.062976	16.306971	19.394236
	10	0.057802	0.072621	0.052353	0.070780	16.552731	21.428264
SVM	1	0.021383		0.025546		8.046289	
	2	0.027721		0.031878		10.046793	
	4	0.036721		0.039702		12.578553	
	6	0.041984		0.044450		14.157343	
	8	0.044525		0.047175		15.036534	
	10	0.046166		0.050236		15.898355	
Naivná predikcia	1			0.023273		7.287591	
	2			0.031486		9.822408	
	4			0.041811		13.078883	
	6			0.047238		14.958371	
	8			0.050788		16.148619	
	10			0.051923		16.428804	

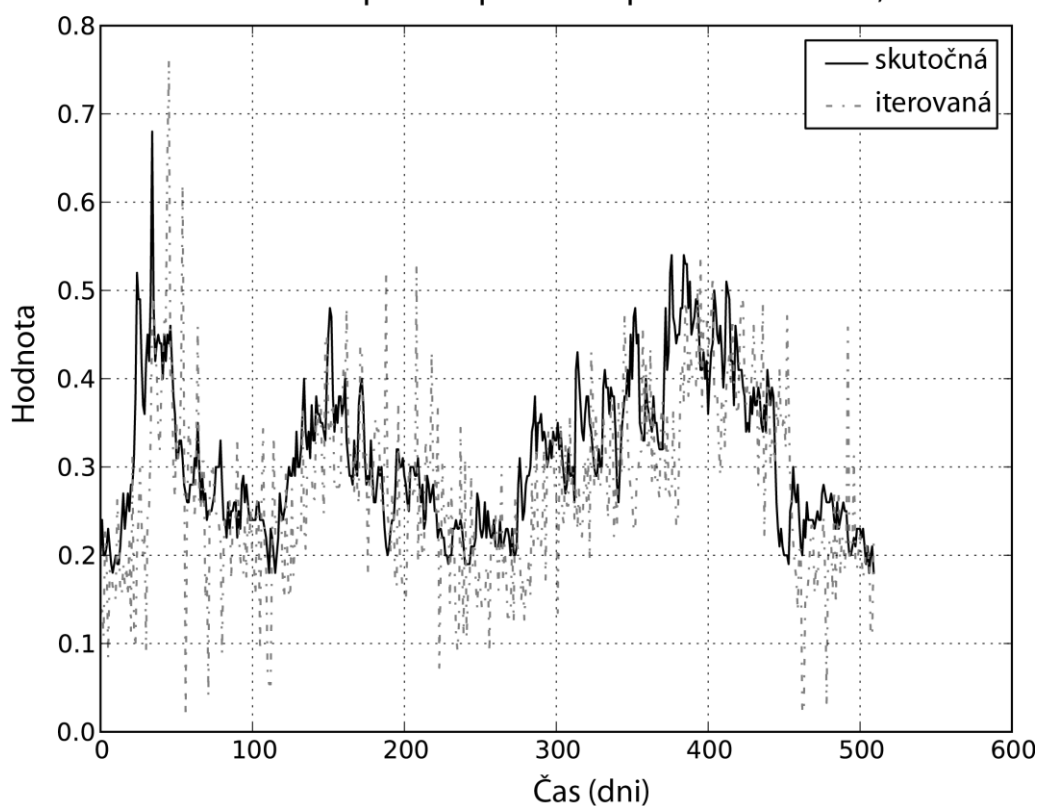
Tabuľka 5 Výsledky experimentu na finančných dátach

V tabuľke 6 sú zaznamenané parametre l_s , η_1 , η_2 , Q získané evolučnou optimalizáciou s časom trvania optimalizácie a potrebného počtu generácií..

l_{pr}	L_s	Q	η_1	η_2	NaN(%)	gen	Čas(s)
1	20	0.01	0.96	0.93398	1.6484	281	11002.7
2	24	0.3833823	0.837201	0.96	1.8913	40	373.7
4	19	17.095816	0.837166	0.96	11.3994	107	865.9
6	20	24.551452	0.551884	0.96	15.6028	80	605.9
8	20	23.786910	0.241268	0.536606	27.3875	113	648.8
10	12	21.696502	0.368874	0.010192	22.1524	128	381.3

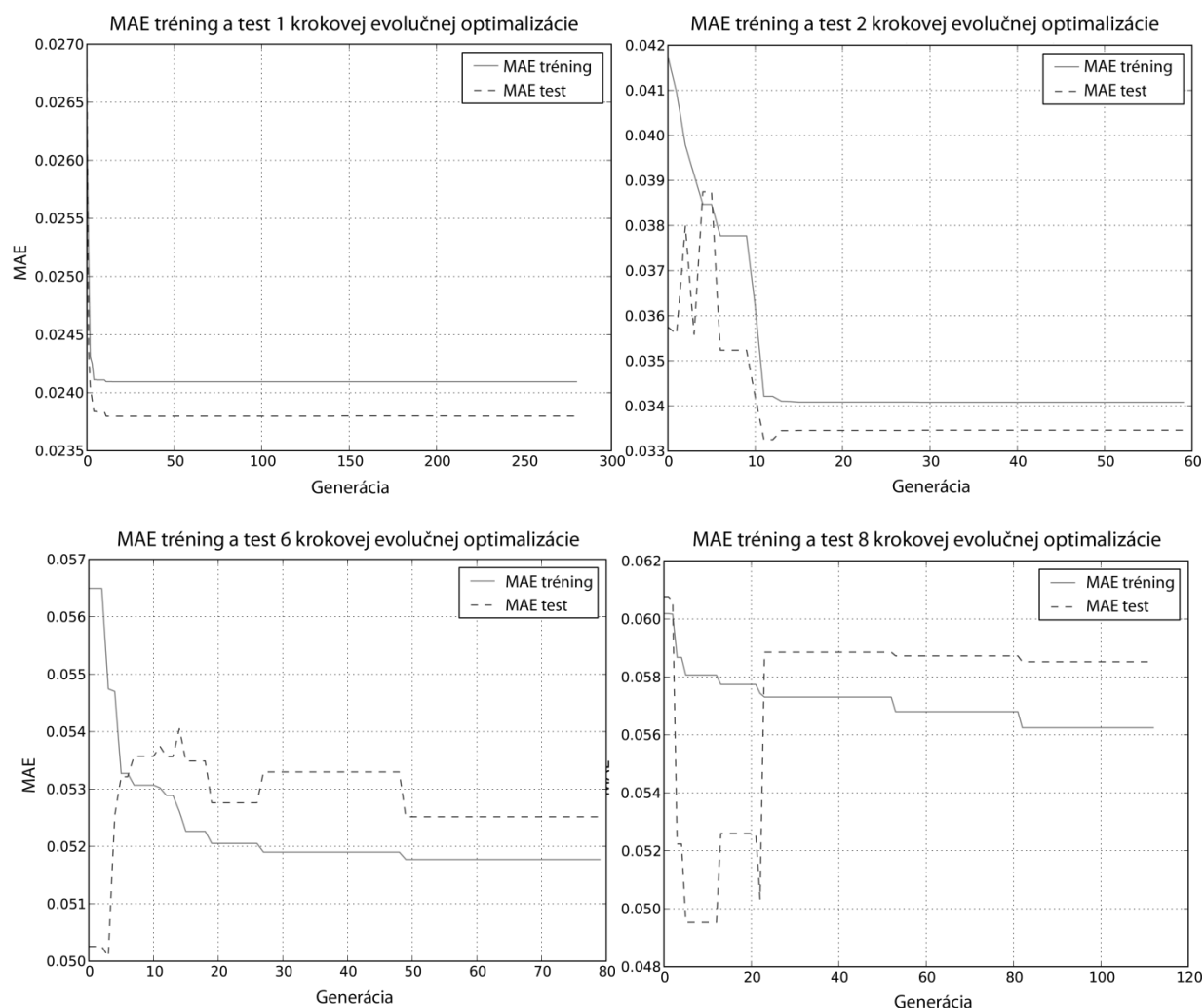
Tabuľka 6 Optimálne parametre PMBSI pre finančné dáta

10 kroková priama predikcia pomocou PMBSI, EA



Obrázok 12 10 kroková priama predikcia finančných dát pomocou PMBSI

Obrázok 13 zobrazuje postupnú evolúciu trénovacej a kontrolnej MAE pre 1, 2, 6 a 8-krokovú predikciu. Je vidno že predikcia na viac krokov dopredu dosahuje postupne pomalšie znižovanie MAE s menšími rozdielmi.



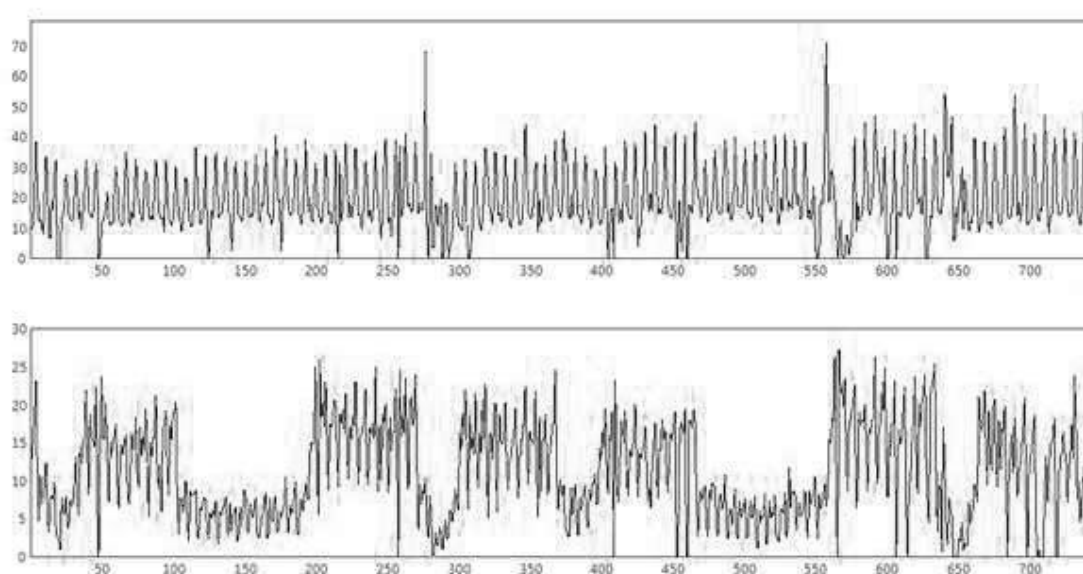
Obrázok 13 Grafy evolúcie trénovacej a kontrolnej MAE na finančných dátach

6.4 Porovnanie s výsledkami zo súťaže „2008 Time Series Forecasting Competition for Computational Intelligence“

Pre porovnanie a získanie prehľadu výkonnosti modelu som sa rozhodol porovnať PMBSI model s výsledkami so súťaže „2008 Time Series Forecasting Competition for Computational Intelligence“. Súťaže sa zúčastnilo 8 štatistických metód a 19 metód umelej inteligencie. V súťaži bolo potrebné predpovedať v 111 časových radoch 56 hodnôt pomocou 735 známych hodnôt.

Dáta predstavujú denný výber peňazí z rôznych bankomatov z rozličných náhodne zvolených častí Anglicka. Všetky časové rady začínajú 18. 3. 1996 a končia 22. 3. 1998, teda poskytujú 2 roky denných výberov. Bankomaty fungujú ako predajne a ponúkajú zákazníkom peniaze. Dáta môžu obsahovať vzory časového radu ako sezónnosť, deň

v týždni, týždeň v mesiaci, mesiac v roku, kalendárne dni ako Vianoce, Veľká noc, lokálne trendy. Dáta obsahujú aj štrukturálne zlomy, odľahlé hodnoty, nuly a chýbajúce hodnoty, atď. Obrázok 14 zobrazuje grafy priebehu 2 časových radov zo súťaže.



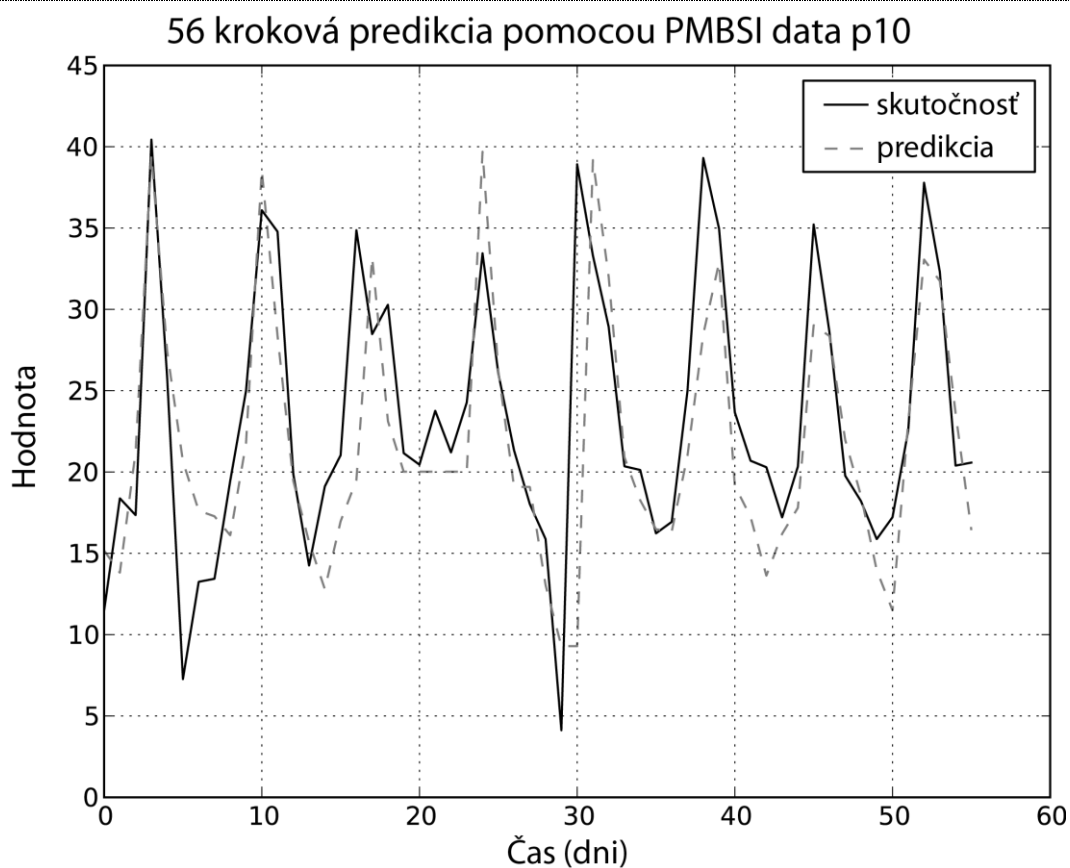
Obrázok 14 Príklad časových radov zo súťaže [25]

V každom časovom rade bola testovacia časť 56 hodnôt vyhodnotená SMAPE podľa vzorca 6.2 a poradie bolo stanovené priemerom SMAPE všetkých časových radov. Nasledujúca tabuľka 7 zobrazuje porovnanie výsledkov zo súťaže s výsledkami PMBSI modelu s 56 krokovou predikciou.

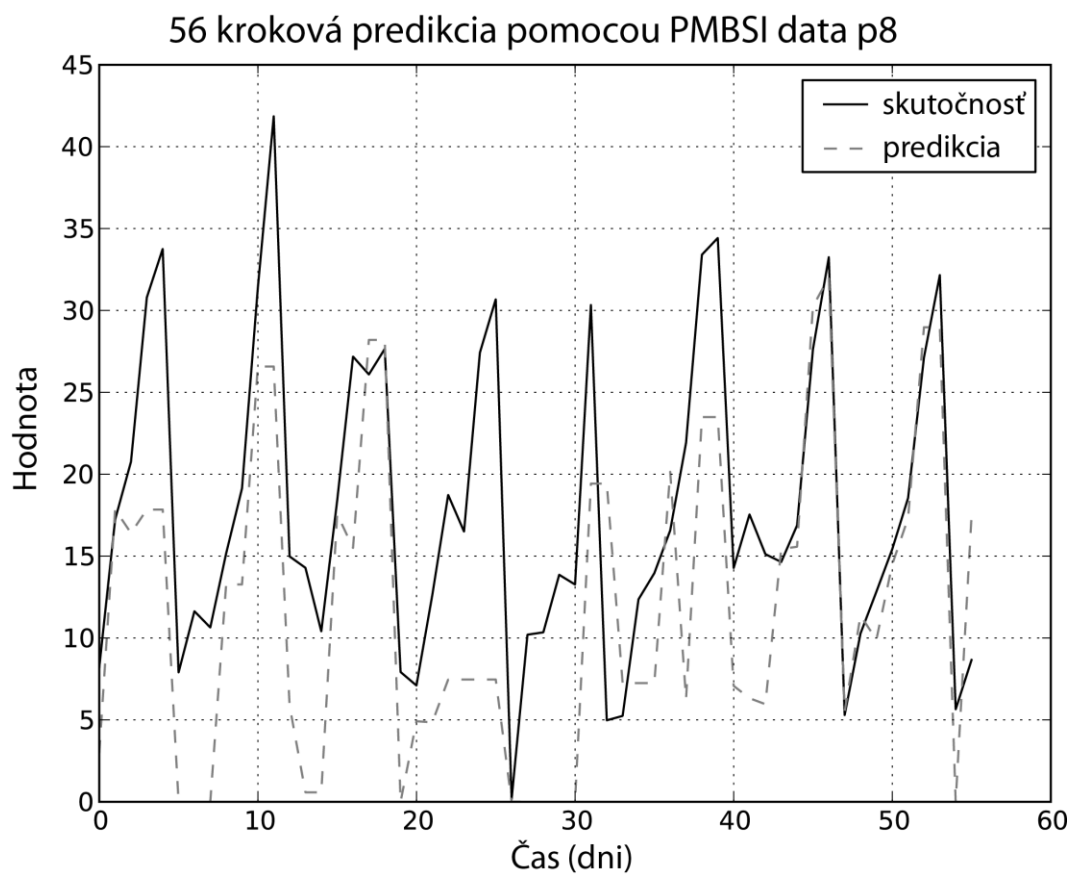
Priemer SMAPE	Poradie celkové	Poradie metód umelej inteligencie	Poradie štatistických metód	Súťažiaci
19,9	1		1	Wildi
20,4	2	1		Andrawis
20,5	3	2		Vogel
20,6	4	3		D'yakonov
21,1	5		2	Noncheva
21,7	6	4		Rauch
21,8	7	5		Luna
21,9	8		3	Lagoo
22,1	9	6		Wichard
22,3	10	7		Gao
23,7	11	8		Puma-Villanueva
24,1	12		4	Autobox(Reilly)
24,5	13		5	Lewicke
24,8	14		6	Brentnall
25,3	15	9		Dang
25,3	16	10		Pasero
25,3	17	11		Adeodato
26,8	18	12		neuvverejnene
27,3	19	13		neuvverejnene
28,1	20	14		Tung
28,8	21		7	Naive Seasonal
33,1	22	15		neuvverejnene
36,3	23	16		neuvverejnene
38,8				PMBSI
41,3	24	17		neuvverejnene
45,4	25	18		neuvverejnene
48,4	26		8	naive Level
53,5	27	19		neuvverejnene

Tabuľka 7 Porovnanie PMBSI s výsledkami súťaže "2008 Time Series Forecasting Competition for Computational Intelligence"

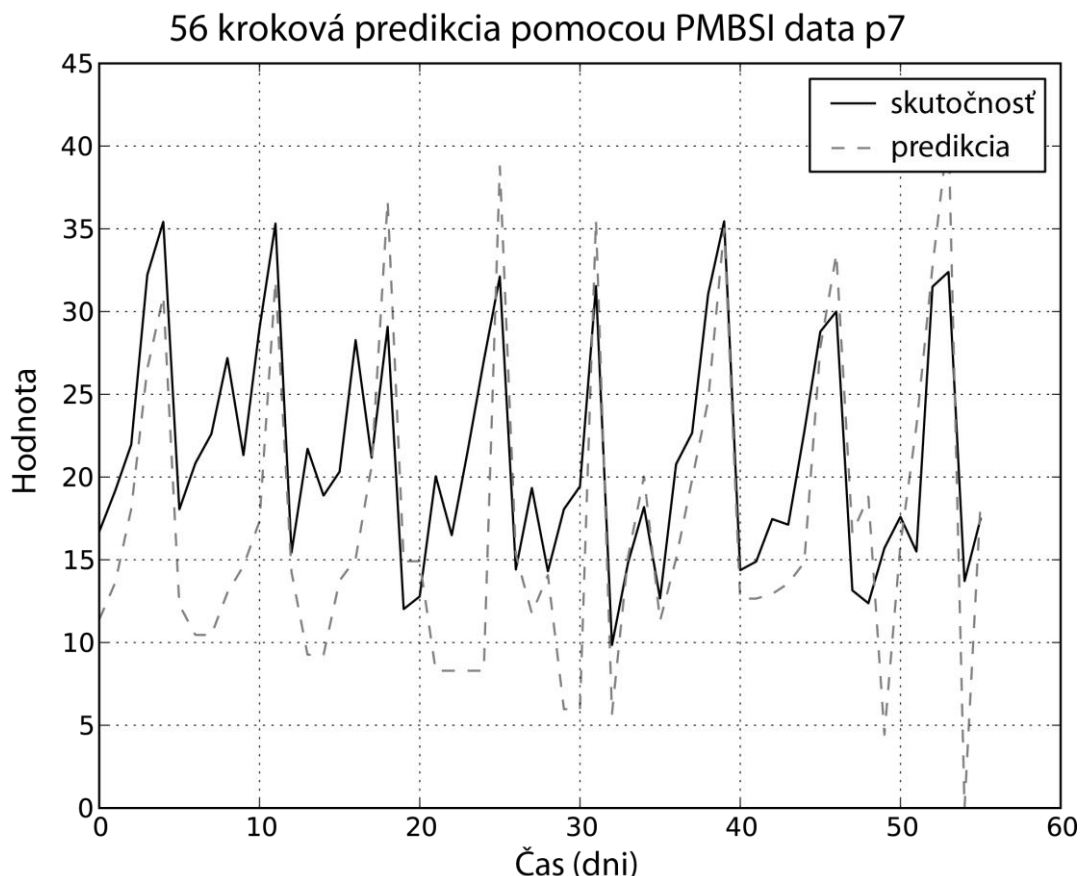
Kompletná tabuľka výsledkov predikcie všetkých 111 časových radov je v prílohe C. Obrázok 15, Obrázok 16, Obrázok 17 zobrazujú predikcie vybraných časových radov zo súťaže so SMAPE najlepším 19,75 (p10), najhorším 76,98 (p8) a najbližším priemeru 38,75 (p7).



Obrázok 15 56-kroková predikcia pomocou PMBSI zo súťaže s najnižším SMAPE 19,75



Obrázok 16 56-kroková predikcia pomocou PMBSI zo súťaže s najhorším SMAPE 76,98



Obrázok 17 56-kroková predikcia pomocou PMBSI zo súťaže so SMAPE najbližšej priemeru 37,75

6.5 Zhodnotenie experimentov

Experimenty na umelých dátach a finančných dátach boli vykonané kvôli otestovaniu a kontrole evolučného algoritmu pre efektívnu a časovo nenáročnú optimalizáciu vstupných parametrov l_s , η_1 , η_2 , Q pre predikciu časového radu pomocou PMBSI a porovnané s článkom [16]. Na základe vykonaných experimentov je možné povedať, že evolučný algoritmus pri implementácii do PMBSI je schopný nájsť vhodné parametre pre predikciu časových radov a to v krátkom čase.

Umelé dáta PMBSI s parametrami nájdenými evolučným algoritmom výrazne prekonal výsledky z článku pre 1,2 aj 3-krokovú jednoduchú aj iterovanú predikciu. PMBSI dosahovalo lepšie výsledky ako SVM aj pre 1 a 2-krokovú iterovanú predikciu, ako je vidno v tabuľke 3. Pri 3-krokovej predikcii PMBSI už SVM nedokázala prekonať. Pri finančných dátach dokázal evolučný algoritmus taktiež rýchlo nájsť lepšie parametre, okrem 1-krokovej predikcie, kde sa nepodarilo nájsť nižšie MAE. Napriek tomu ani SVM ani PMBSI nedokázali lepšie predpovedať časový rad ako Naivná predikcia 1 a 2 kroky dopredu. Pre viac krokov dopredu už SVM dosahovala najlepšie výsledky.

Porovnanie s výsledkami zo súťaže „*2008 Time Series Forecasting Competition for Computational Intelligence*“ bolo vykonané z dôvodu porovnania PMBSI s inými metódami predikcie časových radov, štatistickými alebo založených na princípoch umelej inteligencie. Priemerné SMAPE pre PMBSI v súťaži bolo 38,8 a priemer všetkých SMAPE zo súťaže je 27,9. PMBSI prekonala v súťaži iba 4 metódy, z čoho je možné usúdiť, že PMBSI v terajšej forme nie je veľmi vhodnou metódou na predikciu časových radov a sú potrebné ďalšie úpravy.

Záver

Na základe formulácie uvedených úloh diplomovej práce boli vypracované nasledovné časti:

V tretej kapitole som sa venoval opisu a porovnaniu moderných metód predikcie časových radov, konkrétne pomocou neurónových sietí a SVMi, ich vlastnostiach a implementáciách z publikácií [8], [10].

Štvrtá kapitola opisuje Predikčný model založený na strunových invariantoch. Sú v nej definované strunové mapy, strunové invarianty a je odvodený matematický model predikcie.

V piatej kapitole som navrhol evolučný algoritmus pre efektívne nájdenie vhodných parametrov pre PMBSI, a realizoval návrh v zvolenom programovacom jazyku a vývojovom prostredí.

V šiestej kapitole sa nachádzajú experimenty na ktorých som overil funkčnosť evolučného algoritmu vloženého do PMBSI. Po porovnaní s článkom [16] a potvrdení efektívneho fungovania evolučného algoritmu pre nájdenie parametrov pre PMBSI som model porovnal s ostatnými metódami pre predikciu časových radov použitých v súťaži „2008 Time Series Forecasting Competition for Computational Intelligence“. Nakoniec boli výsledky experimentov zhodnotené.

Pre vylepšenie výkonu metódy PMBSI je potrebné zamerať sa do budúcnosti na vývoj a testovanie parametru váhy z rovnice 4.10, 4.11, zapojenia tejto váhy do evolučného algoritmu a experimentálne overiť na ďalších dátach. Keďže PMBSI v niektorých testoch dokázala prekonať SVM bolo by vhodné pokúsiť sa vytvoriť obchodný systém na základe predikcie PMBSI modelu pre zhodnotenie výkonnosti.

Zoznam použitej literatúry

- [1] COCHRANE, H., John: Time Series for Macroeconomics and Finance, Graduate School of Business, University of Chicago, spring 1997.
- [2] YASER, Sam – ATIYA Af: Introduction to financial forecasting. Applied Intelligence 6:205-13. 1996.
- [3] HAYKIN S.: Neural networks: A comprehensive foundation (2nd ed.), Prentice Hall, NJ, USA, 1999.
- [4] SINČÁK, Peter – ANDREJKOVÁ, Gabriela: Neurónové siete Inžiniersky prístup (1. diel). Košice: TUKE 1996.
- [5] PAN, Zh – WAN, Xd: Wavelet-based density estimator model for forecasting. Journal of Computational Intelligence in Finance 6:6-13. 1998.
- [6] DORSEY, R. – SEXTON, R.: The use of parsimonious neural networks for forecasting financial time series. Journal of Computational Intelligence in Finance 6:24-30. 1998.
- [7] NELSON, M – HILL, T. – REMUS, T. – O'Connor. M.: Time series forecasting using NNs: Should the data be deseasonalized first? Journal of Forecasting 18 (1999) s359-367.
- [8] ZHANG, G., Peter – QI, Min: Neural network forecasting for seasonal and trend time series. European Journal of Operational Research 160 (2005) s501-514.
- [9] QI, Min – ZHANG, G., Peter: Trend Time-Series Modeling and Forecasting With Neural Networks. IEEE Transactions On Neural Networks, Vol. 19, No. 5, 2008.
- [10] ENKE, David – THAWORNWONG, Suraphan: The use of data mining and neural network for forecasting stock market returns, Expert Systems with Applications 29 (2005) 927-940.
- [11] VAPNIK, Vladimir, N.: The nature of statistical learning theory (1995). New York: Springer, ISBN: 0-387-94559-8.
- [12] SMOLA, J., A. – SCHOLKOPF, B.: A tutorial on Support Vector Regresion, NeoroCOLT2 Technical Report Series, NC2-TR-27150, 1998.
- [13] Institute of Microbial Technology [online]. [2005] [cit. 2014-02-27]. Algorithm of Rice Blast Prediction. Dostupné na internete: <<http://www.imtech.res.in/raghava/rbpred/algorithm.html>>.
- [14] TAY, Francis, E.H. – CAO, Lijuan: Application of support vector machines in financial time series forecasting. The International Journal of Management Science, Omega 29, 309-317, 2001.
- [15] KIM, Kyoung-jae: Financial time series forecasting using support vector machines: Neurocomputing 55 (2003) 307-319.

-
- [16] BUNDZEL, Marek – KASANICKY, Tomáš – PINCAK, Richard: Experimental Analysis Of The Prediction Model Based On String Invariants: Computing and Informatics, Vol. 32, No. 6, (2013).
- [17] HORVÁTH, D. - PINCAK, R.: From The Currency Rate Quotations onto Strings and Brane World Scenarios, Physica A: Statistical Mechanics and its Applications, Volume 391, Issue 21, pp. 51725188.
- [18] POLCHINSKI, J.: String Theory, Cambridge University Press, (1998).
- [19] C., Münnix, Michael, Shimada, Takashi, Rudi, Schäfer, Seligman, Francois Leyvraz Thomas H., Guhr, Thomas and Stanley, H. E., (2012), Identifying States of a Financial Market, Quantitative Finance Papers, arXiv.org.
- [20] CHRISTIANSEN – JED, D.: The Journal of Prediction Market 1, 17. (2007).
- [21] CHANG, Ch. Ch. – HSIEH, Ch. Y. – LIN, Y. Ch.: Applied Economics Letters iFirst (2011).
- [22] WOLFERS, J. – ZITZEWITZ, E.: Prediction Markets, Journal of Economic Perspectives 18, 107. (2004).
- [23] MACH, Marián: Evolučné algoritmy: Prvky a princípy. Košice: TU-FEI, 2009. ISBN 978-80-8086-123-0.
- [24] BÄCK, T.: Evolutionary Algorithms in Theory and Practice. Oxford University Press, New York, (1996), ISBN 0-19-509971-0, s. 314
- [25] NN5 Forecasting Competition for Artificial Neural Networks & Computational Intelligence [online] [2008] [cit. 2014-3-27]. Dostupné na internete: <http://www.neural-forecasting-competition.com/NN5/datasets.htm>
- [26] ZELLE, John, M.: Python as a First Language. Department of Mathematics, Computer Science, and Physics, Wartburg College, (2010).

Prílohy

- Príloha A: Kompletne výsledky 111 časových radov z experimentu zo súťaže „2008 *Time Series Forecasting Competition for Computational Intelligence*“
- Príloha B: CD médium – diplomová práca v elektronickej podobe, prílohy v elektronickej podobe, zdrojový kód navrhnutého evolučného algoritmu
- Príloha C: Systémová a používateľská príručka

Príloha A: Kompletné výsledky z experimentu zo súťaže

Časový rad	Generácia	MAE tréning	MAE test	SMAPE test	ls	η_1	η_2	Q	Nans %
1	146	7.375437	7.730503	26.84357	57	0.031362	0.003958	12.87788	10.91445
2	68	5.186019	5.200247	29.6436	57	0.022325	0.011177	24.91	17.84661
3	106	6.192545	6.649781	28.96462	57	0.002847	0.018198	8.799085	7.669617
4	118	5.754562	6.392964	39.00201	57	-0.00882	0.004241	10.78137	6.342183
5	46	7.301445	7.132578	42.15279	57	-0.5792	-0.00133	24.91	14.74926
6	119	6.882448	5.958763	41.0464	57	-0.16234	-0.00714	23.64889	17.25664
7	112	5.514884	6.154977	38.74674	57	-0.03449	0.011269	15.48568	12.83186
8	75	8.012493	7.369843	76.97893	57	-0.00655	-0.0331	21.41572	15.63422
9	123	3.689761	2.064363	26.29321	57	-0.40571	0.02288	13.93974	23.30383
10	99	4.244709	4.045375	19.75011	57	-0.00051	-0.00587	10.85266	6.19469
11	69	6.046342	7.863267	41.55061	57	-0.09423	0.016676	10.03829	11.35693
12	86	4.603612	5.259658	42.04366	57	-0.38222	-0.00214	18.25333	18.43658
13	111	6.021333	8.098166	55.87695	58	-0.09684	-0.75571	5.153597	18.02068
14	73	5.045144	5.712904	38.24532	57	-0.52468	0.020667	24.91	20.35398
15	122	4.862897	5.412074	36.47102	57	-0.11118	-0.01928	12.66363	8.40708
16	122	7.705448	6.292655	33.60609	57	-0.0404	-0.00803	11.96978	7.964602
17	79	8.663695	8.368228	50.69139	57	-0.8651	-0.00072	23.5859	19.76401
18	132	8.822663	6.796251	36.15951	57	0.006365	0.010316	5.830678	8.702065
19	161	8.86888	9.182171	44.76348	57	-0.59943	0.053706	22.97975	18.73156
20	105	4.567564	4.948823	34.18707	57	0.020786	0.007367	14.78051	11.06195
21	145	4.213956	4.92399	32.08795	57	0.001246	-0.0155	8.090139	7.964602
22	135	4.108404	3.885211	24.81036	57	0.014112	0.014456	9.91295	9.144543
23	43	3.642021	3.425882	44.18399	57	-0.64934	-0.01747	18.84458	15.19174
24	105	11.22773	10.88144	50.19879	57	-0.24347	0.028118	14.02022	14.30678
25	81	6.086759	7.149013	41.97884	57	-0.88695	0.08891	23.56477	25.66372
26	59	4.01332	4.791122	29.52571	57	-0.00814	0.027028	8.861227	9.882006
27	108	2.667566	2.250305	34.52127	57	-0.52881	0.05841	14.10492	24.77876
28	94	4.111187	4.646327	33.10356	57	0.036235	0.001085	10.5792	9.144543
29	102	6.105972	7.2437	37.18377	57	-0.03653	0.005034	12.94906	10.76696
30	98	7.196853	4.67819	36.13333	57	-0.18028	-0.03579	8.506945	14.30678
31	48	5.606091	4.467711	32.00453	57	0.095207	0.007519	24.91	20.20649
32	96	5.451503	5.811208	35.02483	57	-0.00288	0.007476	8.068115	8.702065
33	100	7.605146	8.008639	47.0497	57	-0.08997	0.00923	13.07399	16.0767
34	117	8.57095	6.82212	22.22103	57	0.002097	0.021914	7.082756	8.702065
35	125	5.00987	6.162581	40.46364	99	-0.5855	0.96	0.646233	0
36	96	4.529991	4.400135	51.68015	57	-0.38874	-0.59333	15.19325	20.94395
37	127	6.385449	5.392694	42.48766	57	-0.00274	0.023974	8.281031	16.0767
38	158	7.169702	7.518774	35.9648	57	-0.59572	-0.00626	12.81102	13.42183
39	46	4.990873	5.943896	40.91953	57	-0.99	-0.00016	24.91	20.35398
40	82	4.280566	4.278385	27.45566	57	-0.03214	0.016461	11.57536	11.79941

Časový rad	Generácia	MAE tréning	MAE test	SMAPE test	ls	η_1	η_2	Q	Nans %
41	101	5.637464	6.72869	47.85684	57	-0.54552	0.015025	14.2278	15.33923
42	48	6.533284	5.782649	40.97747	57	-0.67516	-0.00142	24.91	16.37168
43	99	5.996021	7.165659	37.54232	57	-0.04756	0.010857	21.70445	16.0767
44	38	4.355646	4.637338	30.12088	57	0.004826	0.00261	13.92938	7.964602
45	77	4.777904	4.879611	34.27282	57	-0.05184	-0.00276	15.94969	16.0767
46	53	4.770652	5.174429	33.90979	57	-0.04849	0.014706	16.74849	16.66667
47	70	7.391234	7.62262	25.04013	57	0.006338	0.006634	18.96447	11.06195
48	82	7.938376	8.24272	45.00524	57	0.007821	0.018375	6.899275	8.99705
49	186	5.380234	4.654435	36.72395	57	-0.01863	0.001045	6.645553	8.112094
50	101	6.014516	7.92889	42.13617	57	-0.02694	0.016755	17.35209	14.0118
51	123	4.624493	6.276438	45.16595	57	0.000476	-0.02222	11.10403	11.20944
52	89	9.910648	9.769681	37.68222	57	-0.03826	-0.04222	16.24491	12.38938
53	62	2.967953	3.63067	53.05033	57	-0.01826	0.004987	15.46749	10.47198
54	95	5.390445	7.11765	51.07468	57	-0.06264	0.064898	12.15024	14.74926
55	67	2.171842	4.43343	31.52594	57	0.004336	0.008613	12.2743	7.522124
56	108	4.807656	4.478194	49.31897	57	-0.04965	0.011772	15.40195	14.74926
57	89	4.769251	5.233417	33.49253	57	0.003748	0.017015	10.48366	8.702065
58	235	6.350137	6.638797	38.27299	57	-0.00134	0.015524	13.82484	13.86431
59	149	3.712146	4.338698	49.12885	57	-0.92647	-0.01884	15.68064	19.9115
60	111	5.724113	5.705621	43.09099	57	-0.24346	0.016207	16.67121	19.76401
61	120	5.719962	5.236569	38.2319	57	-0.04816	-0.03779	6.956161	9.882006
62	78	6.555009	6.28225	43.48983	57	-0.02033	0.017232	18.4006	15.33923
63	70	4.077367	3.845946	30.77912	57	-0.00976	-0.00793	14.81798	9.439528
64	116	5.218	5.001459	50.356	57	-0.18133	-0.31392	13.14857	14.0118
65	144	7.815439	7.601577	34.94537	57	-0.77457	-0.00244	22.10222	19.61652
66	124	6.519301	6.685981	32.73238	57	-0.02915	-0.01218	7.917305	8.112094
67	99	8.305579	6.842778	28.39341	57	-0.21511	0.024529	18.24731	23.45133
68	84	10.1215	11.33673	29.60228	57	0.06882	0.017212	19.74706	15.63422
69	70	6.70096	6.628938	56.51242	57	-0.99	-0.0026	24.91	20.50147
70	139	4.97894	4.112747	27.80371	57	0.006089	0.011593	13.13496	8.849558
71	146	3.299789	3.225822	28.18047	57	0.097796	0.007879	16.43465	15.9292
72	79	4.67202	6.004605	42.7549	57	-0.01094	0.01517	23.48737	18.58407
73	121	6.767804	4.801458	35.22323	57	-0.05787	-0.02877	18.14149	17.10914
74	83	3.901022	4.568021	45.27024	57	-0.05232	-0.00079	12.1215	10.91445
75	137	4.989534	4.732956	36.45218	98	-0.65773	0.917267	0.740188	0
76	81	3.867347	3.792337	25.85051	57	-0.00553	0.005895	6.197431	5.457227
77	69	4.569528	5.736134	37.61976	57	-0.99	0.04906	24.91	25.51622
78	66	4.948117	5.191968	28.19222	57	-0.362	0.022794	14.1817	19.02655
79	61	4.908678	4.473572	30.657	57	0.00257	0.006942	18.66787	13.12684
80	86	7.535046	8.251957	43.38197	57	0.184961	0.001267	13.64318	17.9941
81	151	2.90583	2.749569	40.24247	57	-0.03944	-0.00052	9.695469	11.50442
82	256	3.170057	2.707214	24.45916	57	0.026806	0.00901	7.758348	6.19469
83	48	8.351278	9.970101	37.72557	57	-0.05551	0.01172	10.94013	12.0944

Časový rad	Generácia	MAE tréning	MAE test	SMAPE test	ls	η_1	η_2	Q	Nans %
84	75	4.350572	5.030698	32.81891	57	-0.72003	0.009268	18.33235	20.64897
85	138	4.733165	5.174251	45.91004	57	-0.00382	-0.01007	14.25764	10.91445
86	155	4.484392	5.711378	60.49561	57	-0.10907	-0.0083	15.55313	15.33923
87	77	6.678017	5.464705	39.6258	57	-0.99	-0.00741	20.93167	16.66667
88	82	8.715262	7.519667	69.8864	58	0.96	0.009633	15.81662	35.00739
89	177	4.714873	2.674676	32.28916	57	-0.13079	0.018898	16.06799	25.66372
90	75	6.586901	5.202768	27.65093	57	-0.21902	-0.002	9.131219	14.74926
91	137	5.612373	5.726563	55.32904	57	-0.01541	0.032522	17.54341	17.69912
92	64	5.210354	6.410433	52.42714	57	-0.78473	-0.00446	21.74146	18.87906
93	60	7.583905	7.890545	32.23815	57	0.008642	0.024678	23.54415	18.73156
94	84	4.466916	4.312455	26.30337	57	0.010283	0.027994	10.37734	9.587021
95	54	8.127095	8.075579	44.12222	57	-0.01279	0.00523	14.84549	13.42183
96	113	6.245231	5.309154	29.97683	57	0.000101	0.007921	7.423844	8.40708
97	86	6.61692	5.23675	50.19475	57	0.230287	0.011125	14.43069	22.41888
98	133	5.755281	6.20725	38.40375	57	-0.02036	-0.00043	7.768992	8.40708
99	107	3.990614	4.534755	30.74387	57	-0.07509	0.02052	23.61976	18.28909
100	66	3.964489	3.531078	29.26295	57	0.005364	0.012051	14.44577	11.20944
101	80	7.58361	3.934889	21.87106	57	-0.33447	0.014377	17.67269	23.45133
102	57	8.080058	7.600795	51.3779	57	-0.15452	-0.04355	14.75684	11.9469
103	142	2.852395	3.858088	56.10039	58	0.328131	-0.19845	13.11891	33.08715
104	94	8.487458	6.150738	31.65861	57	-0.05759	-0.00178	11.23693	12.53687
105	70	4.866781	5.770578	39.01088	57	-0.02298	-0.00025	12.02394	12.38938
106	139	7.105349	5.397591	36.45721	57	-0.00426	-0.04037	10.25666	11.06195
107	81	5.231364	6.130163	50.97561	57	-0.99	-0.08306	19.21312	19.02655
108	120	5.093745	5.417889	42.5295	57	-0.27693	0.032924	19.07787	16.66667
109	154	6.837007	7.916478	37.43654	57	-0.15105	-0.01445	11.5934	8.849558
110	90	9.477505	13.70364	57.34391	57	0.499423	0.019286	24.91	28.46608
111	56	4.951055	3.775774	28.22633	57	-0.15448	-0.00279	23.32177	13.12684