

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

PREDIKCE KURSŮ PRO OBCHODOVÁNÍ NA AKCIOVÝCH TRZÍCH

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. ROMAN MIKULENČÁK

BRNO 2015



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

PREDIKCE KURSŮ PRO OBCHODOVÁNÍ NA AKCIOVÝCH TRZÍCH

PREDICTION OF PRICES IN STOCK EXCHANGE TRADING

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. ROMAN MIKULENČÁK

VEDOUcí PRÁCE

SUPERVISOR

doc. Dr. Ing. JAN ČERNOCKÝ

BRNO 2015

Abstrakt

Práce se zabývá automatickým obchodním systémem s využitím neuronových sítí a adaptivním trénováním. Použita je jak technická tak automatická fundamentální analýza, proto jsou jako vstupy do neuronové sítě použita jak historická data burzy tak i textová data ze zpráv. Práce také vysvětluje základy obchodování, technickou analýzu a odborné termíny. Obsahuje popis algoritmické podstaty, implementace programu a experiment vytvořený obchodním systémem. Vybraná strategie je srovnána s jinými přístupy.

Abstract

The work deals with an automatic trading system and adaptive training. Is used both technical and automatic fundamental analyses, therefore as inputs to the neural network is used historical data exchanges and text data from reports. It explains the basics of trading, technical analysis and technical terms. The work deals with technical and fundamental analysis. It contains a description of algorithmic nature, program implementation and experiment with developed trading system. The selected strategy is compared to other approaches.

Klíčová slova

Obchodování, trh cenných papírů, burza, strategie, svíčkový graf, svíčkové formace, predikce burzy, předpověď vývoje kurzu, neuronové sítě, fundamentální analýza

Keywords

Trading, stock market, stock exchange, strategy, candlestick chart, candle formation, stock market prediction, stock forecast, neural network, fundametal analysis

Citace

Roman Mikulenčák: Predikce kursů pro obchodování
na akciových trzích, diplomová práce, Brno, FIT VUT v Brně, 2015

Predikce kursů pro obchodování na akciových trzích

Prohlášení

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně pod vedením doc. Černockého

.....
Roman Mikulenčák
27. května 2015

Poděkování

Na tomto místě bych rád poděkoval vedoucímu diplomové práce doc. Dr. Ing. Janu Černockému, za cenné připomínky a odborné rady, kterými přispěl k vypracování této práce.

© Roman Mikulenčák, 2015.

Tato práce vznikla jako školní dílo na Vysokém učení technickém v Brně, Fakultě informačních technologií. Práce je chráněna autorským zákonem a její užití bez udělení oprávnění autorem je nezákonné, s výjimkou zákonem definovaných případů.

Obsah

1 Úvod	3
1.1 Popis obecného obchodování	4
1.1.1 Fundamentální analýza	4
1.1.2 Psychologická analýza	5
1.1.3 Technická analýza	5
1.1.4 Grafické znázornění pohybu ceny	7
1.2 Schéma navrženého systému	9
2 Neuronové sítě	10
2.1 Inspirace v biologii	10
2.2 Perceptron	10
2.3 Topologie	11
2.4 Trénování	11
2.4.1 Backpropagation	11
2.4.2 Křížová validace	12
2.4.3 Adaptivní trénování	12
2.5 Konkrétní použití	14
2.5.1 Důležitost času	14
2.5.2 Vstupy	15
2.5.3 Výstup	15
3 Technické vstupy	16
3.1 Svíčkové formace	16
3.2 Index relativní síly (RSI)	17
3.3 Commodity Channel Indicator (CCI)	18
3.4 Acceleration/Deceleration Technical Indicator (AC)	18
3.5 Moving Average Convergence/Divergence Indicator (MACD)	18
3.6 Normalizace technických vstupů	18
4 Fundamentální vstupy	20
4.1 Titulky zpráv	20
4.2 Forexní kalendář	20
4.3 Zpracování textových dat	20
4.4 Principal component analysis	22
4.4.1 Konkrétní použití	22

5	Obchodní model	24
5.1	Otevření obchodu	24
5.2	Ukončení obchodu	24
5.3	Ověření ziskovosti nového trénování	24
5.3.1	Násobná inicializace neuronové sítě	26
5.3.2	Funkce hodnotící váhy	26
6	Implementace	27
6.1	Struktura systému	27
6.2	Obchodní model	27
6.2.1	Programovací jazyk	27
6.2.2	Platforma	27
6.3	Knihovna neuronové sítě	29
6.4	Skript pro stažení dat titulků	29
6.5	Skript pro stažení dat forexního kalendáře	30
6.6	Program na vytvoření vektoru výskytu slov	30
6.7	Skript na vytvoření kovarianční matice	30
7	Data	31
7.1	Technická data	31
7.2	Fundamentální data	33
7.2.1	Titulky zpráv	33
7.2.2	Forexní kalendář	33
7.2.3	Nejčastější slova	34
8	Experimenty a výsledky	35
8.1	Srovnání fixní sítě s adaptivním dotrénováním	35
8.2	Ověření zisku nově natrénovaných vah	35
8.3	Násobná inicializace neuronové sítě	37
8.4	Srovnání strategií	37
8.5	Forward test na demo účtu	39
9	Závěr	41
9.1	Souhrn	41
9.2	Možné pokračování práce	41
A	Slovník pojmů	46
B	Obsah CD a návod k použití	49
B.1	AOS	49
B.2	BackPropDll	49
B.3	Data	50
B.3.1	Technická data	50
B.3.2	Správa dat	50
B.3.3	Fundamentální data	50
B.4	MetaTrader	50
B.5	Scripts	51
B.6	VectorExtract	51
B.7	Instalace a spuštění	51

Kapitola 1

Úvod

Už od dávných časů lidé chtějí co nejjednodušeji zbohatnout. Je pro to několik legálních způsobů. Například vsadit si sportku. Avšak jsou i mnohem duchaplnější způsoby. Jednomu z nich jsem se věnoval ve své diplomové práci i já. Obchodování na kapitálových trzích mě zajímalo již dříve a vypracoval jsem na toto téma svou bakalářskou práci [41]. Především mě fascinovala možnost vytvořit obchodní strategii, která by obchodovala automaticky a vyžadovala by jen malou údržbu. V dnešní době, kde velká část obchodů se uskutečňuje přes internet, není pro nikoho problémem takto obchodovat přes vybraného zprostředkovatele. Práce také může sloužit jako návod, jak začít obchodovat.

Ve své práci využívám data historického vývoje ceny trhu, nazývaná také technická analýza a také data textová, která se dají zařadit do analýzy fundamentální. Data jsou v rámci předzpracování pročištěna, agregována a normalizována. Poté jsou přivedeny na neuronovou síť, která je schopná se vypořádat s rozmanitostí vstupních dat. Po naučení neuronové sítě je možné získat odezvu, na základě které se bude obchodovat. Takto vytvořený obchodní systém jsem testoval a výsledky jsem porovnal s jinými obchodními systémy ve stejném časovém období. Porovnání můžeme vidět v kapitole 8.

Z možných trénovacích algoritmů jsem zvolil backpropagation, který konvergoval nejrychleji. Učení neuronové sítě jsem prováděl nejenom na začátku, ale také adaptivně v průběhu experimentu na nově dostupných datech.

Na co nejlepší propojení výstupů neuronové sítě a zisku celého obchodního systému jsem vytvořil techniku, která vybírá z několika natrénovaných neuronových sítí pouze tu, která si na minulém vývoji vedla nejlépe.

Pro kombinování technické i fundamentální analýzy jsem se rozhodl, protože pomocí pouze technické analýzy obchoduje velký počet obchodníků na burze, a proto je těžké konkurovat v obchodování, když máte pouze stejné informace jako všichni ostatní. Proto je zajímavé zkoumat jiné zdroje informací pro podporu rozhodování na burze než pouhý historický vývoj ceny [43].

Automatický obchodní systém jsem vytvořil v programu MetaTrader. Je to program, který je využíván bezpočtem světových brokerů a traderů. Je proto ihned možné ho spustit v praxi, na reálném účtu s reálnými penězi. Tato platforma zajišťuje průběh experimentů a počítá zisk se započítanými poplatky. Proto, aby strategie byla výnosná, musí zisky z uzavřených obchodů pokrýt ztráty ze špatných rozhodnutí, ale také poplatky za udržování obchodu, uzavření obchodu a jiné.

1.1 Popis obecného obchodování

Tato kapitola byla částečně převzata z bakalářské práce [41]. U obchodování je jasný cíl a to vydělat co nejvíce. Tedy získat co největší zisk. Pro obchodování na kapitálových trzích to platí především. Zisk je kladný rozdíl mezi prodejní a nákupní cenou. Pokud je rozdíl záporný, nazývá se ztráta. Ideální finanční investice navíc vykazuje co největší likviditu a co nejmenší riziko. To je nebezpečí, že obchodník utrpí ztrátu [17].

I obchodování na kapitálových trzích podléhá zdanění, a proto se od každého zisku odečítá daň. Kromě daně zisk snižují poplatky obchodníkovi, kterým podléhá každý obchod. Pouze po jejich odečtení získáváme čistý zisk. Proto již při otevření nového obchodu jsme ve ztrátě. Podle toho, kam se bude aktuální cena trhu ubírat, se můžeme dostat do ještě větší ztráty nebo do zisku, ale zisk z obchodu získáme až po uzavření obchodu a pouze pokud bude dostatečně velký, aby pokryl dříve zmiňované poplatky a daň [57].

Vstup a výstup obchodu

Otevírání nových obchodů lze provádět přes různé obchodní systémy, které poskytuje obchodník s cennými papíry (broker). Mezi ně patří webové prostředí, mobilní aplikace a programy. Obchod se přes platformu zadá a odešle na server brokera [54]. Ten si potom tento obchod udržuje. Při otevření nového obchodu specifikujeme objem, s jakým budeme obchodovat, případně maximální možnou ztrátu (stoploss) a minimální možný zisk (takeprofit). Tyto hodnoty jsou uloženy na serveru a ten, pokud cena trhu dosáhne stoploss nebo takeprofit, obchod uzavře. Kromě této možnosti uzavření obchodu jsou ještě další dvě možnosti: uzavření obchodu pokud je ztrátový a na účtu už nezůstávají prostředky pro pokrytí této ztráty nebo uzavření traderem. Hodnoty stoploss a takeprofit lze dále upravovat a tím dynamicky měnit maximální možnou ztrátu a zisk.

Obchodování na trhu cenných papírů lze rozdělit do tří základních přístupů:

1.1.1 Fundamentální analýza

Fundamentální analýza vychází z informací získaných z médií, finančních výkazů, hodnoty produktů, konkurence, významných zpráv, apod. Dá se také definovat jako snaha o nalezení vnitřní hodnoty aktiva, měny nebo jiného finančního instrumentu. Jedná se o cenu, která není předem známá a jejím odhadem se můžeme dopracovat k zisku [19]. Pro fundamentální analýzu platí, že ji nikdo nevyužívá jako jedinou metodu pro rozhodování při realizaci obchodu. I trader, který využívá výhradně fundamentální analýzu, se před vstupem do obchodu podívá na graf historie vývoje ceny [52]. Pro obchodování s využitím fundamentální analýzy je třeba hodně znalostí a průběžného zjišťování informací.

Samotná fundamentální analýza má mnoho možností pojetí. Záleží na tom, jak k ní budeme přistupovat a jaké informace máme možnost získat. Analyzovat můžeme globální makroekonomiku, tedy hledat státy a jejich vyhlášení a jaký mají dopad na měnu nebo se zaměřit na jednotlivé firmy. Pro řadu investorů analýza začíná např. tak, že hledají státy, u kterých se předpokládá růst. Následně se v těchto státech snaží identifikovat odvětví ekonomiky a v nich konkrétní růstové či poklesové firmy [53]. Dále se fundamentální analýze věnujeme v kapitole 4.

Většina prací, zabývající se automatickým obchodním systémem s dolováním textu, používá finanční zprávy. Finanční zprávy vykazují méně šumu v porovnání s obecnými zprávami. V této práci budu využívat pouze titulky zpráv. Titulky zpráv jsou občas použity protože "jdou rovnou k věci" a tedy obsahují méně šumu [43]. Proto byla použita tato

data i v této práci. Dále také využívám data forexního kalendáře, což jsou nejdůležitější ekonomická vyhlášení států a národních bank a už ze své podstaty ovlivňují trh velkou měrou [21].

V tabulce 1.1 můžeme vidět srovnání prací zabývajících se automatickou fundamentální analýzou.

1.1.2 Psychologická analýza

Psychologická analýza zkoumá sociální prostředí daného aktiva. Tento typ analýzy je směsicí fundamentální analýzy, citu a intuice. Zabývá se zkoumáním sociálního prostředí např. firmy. Praktický způsob může vypadat tak, že trader bude komunikovat se zaměstnanci a ptát se na jejich plány. Zjišťuje finanční situaci firmy a analyzuje její prostředí. Další možností je zjišťování názorů a pocitů ostatních investorů nebo partnerů firmy. Tento způsob obchodování vyžaduje velké zaměření na určité aktivum a jeho hluboké zkoumání. Pro její použití je zapotřebí jistá zkušenost s obchodováním [53].

Automatická psychologická analýza je založena na detekování emocí a sentimentu v textu pomocí sémantické analýzy, například pro účel odhadu poptávky na trhu po novém produktu a obecného povědomí o značce [25]. Existuje skupina zabývajících se získáváním sentimentu a její analýzy nazývané "dolování názoru" [16]. Tento přístup spočívá v identifikování pozitivních a negativních slov v textu a jejich zpracování za účelem klasifikace pozitivního a negativního emociálního postoje. Jako příklad takové analýzy je práce autorů [35], která prezentuje lexikální model pro hloubkovou analýzu sentimentu a dolování názoru [43].

1.1.3 Technická analýza

Technická analýza se zabývá pouze historickým vývojem cen finančních aktiv [52]. Principy technické analýzy vycházejí ze sledování vývoje cen trhů už po několik století. Nejstarší metodou vyvinutou japonskými tradery je používání svíčkových technik, které je dnes nejrozšířenější grafickou metodou. S rozvojem informační techniky je tento přístup stále oblíbenější, protože počítače ulehčily výpočty nutné k technické analýze a umožnily výpočty jinak nerealizovatelné a rychlejší reakce.

Technická analýza je založená na předpovídání budoucího vývoje na základě historického vývoje cen. Cenová data z minulého vývoje na určitém trhu jsou nejpoužívanějším typem informací, které techničtí tradeři používají. Někteří tradeři také do svých výpočtů zahrnují další informace jako objem obchodů, nebo open interest což jsou je počet otevřených obchodů v daný okamžik.

Hlavním cílem technických analytiků je odhad správného času pro nákupní nebo prodejní rozhodnutí. Většinou se zaměřují na krátkodobé obchody [42].

Technická analýza je předpovídání budoucnosti na základě minulých poznatků a tím se podobá třeba předpovědi počasí. Proto nemůžeme očekávat nikdy stoprocentní výsledky, ale můžeme odhadnout, co se pravděpodobně v budoucnu s cenami stane. Lze ji aplikovat na akcie, komodity, futures nebo na kterákoliv finanční aktiva, která se řídí zákony nabídky a poptávky. Lze ji využít v jakémkoli časovém rámci (minutovém, hodinovém, denním, měsíčním nebo i ročním) [52].

V dnešní době téměř každý trader využívá využívá technickou analýzu, čistou fundamentální analýzu již nevyužívá téměř nikdo [36]. Lze ji využít jednoduše například pouhým pohledem na vývoj ceny na trhu v minulosti. Lehce lze zjistit přibližný trend v kterém se trh pohybuje, maximum nebo minimum za určité období a na základě toho se rozhodnout, zda vstoupit do obchodu nebo raději ještě počkat.

Autor	Typ textu	Zdroj textu	Počet položek	Naplánovaný	Nestrukturovaný
Wuthrich et al. (1998)	General news	The Wall Street Journal, Financial Times, Reuters, Dow Jones, Bloomberg	Not given	No	Yes
Peramunetilleke and Wong (2012)	Financial news	HFDF93 via www.olsen.ch	40 headlines per hour	No	Yes
Pui Cheong Fung et al. (2003)	Company news	Reuters Market 3000 Extra	600,000	No	Yes
Werner and Myrray (2004)	Message postings	Yahoo! Finance, Raging Bull, Wall Street Journal	1.5 million messages	No	Yes
Mittermayer (2004)	Financial news	Not mentioned	6602	No	Yes
Das and Chen (2007)	Message postings	Message boards	145,110 messages	No	Yes
Soni et al. (2007)	Financial news	FT Intelligence (Financial Times online service)	3493	No	Yes
Zhai et al. (2007)	Market-sector news	Australian Financial Review	148 direct company news and 68 indirect ones	No	Yes
Rachlin et al. (2007)	Financial news	Forbes.com, today.reuters.com	Not mentioned	No	Yes
Tetlock et al. (2008)	Financial news	Wall Street Journal, Dow Jones News Service from Factiva news database.	350,000 stories	No	Yes
Mahajan et al. (2008)	Financial news	Not mentioned	700 news articles	No	Yes
Butler and Kešelj (2009)	Annual reports	Company websites	Not mentioned	Yes	Yes
Schumaker and Chen (2009)	Financial news	Yahoo Finance	2800	No	Yes
Li (2010)	Corporate lings	Management's Discussion and Analysis section of 10-K and 10-Q lings from SEC Edgar Web site	13 million forward-looking-statements in 140,000 10-Q and K lings	Yes (company annual report)	Yes
Huang, Liao, Yang, Chang, and Luo (2010) and Huang, Chuang, et al. (2010)	Financial news	Leading electronic newspapers in Taiwan	12,830 headlines	No	Yes Yes
Groth and Muntermann (2011)	Adhoc announcements	Corporate disclosures	423 disclosures	No	Yes
Schumaker et al. (2012)	Financial news	Yahoo! Finance	2802	No	Yes
Lugmayr and Gossen (2012)	Broker newsletters	Brokers	Not available	No	Yes
Yu, Duan, et al. (2013)	Daily conventional and social media	Blogs, forums, news and micro blogs (e.g. Twitter)	52,746 messages	No	Yes
Hagenau et al. (2013)	Corporate announcements and Financial news	DGAP, EuroAdhoc	10870 and 3478 respectively	No	Yes
Jin et al. (2013)	General news	Bloomberg	361,782	No	Yes
Chatrath et al. (2014)	Macroeconomic news	Bloomberg	Not mentioned	Yes	No
Bollen and Huina (2011)	Tweets	Twitter	9,853,498	No	Yes
Vu et al. (2012)	Tweets	Twitter	5,001,460	No	Yes

Tabulka 1.1: Přehled prací zabývajících textovým vstupem na finančních trzích. Převzato z [43].

Autor	Přístup
Anastasakis & Mort, 2009	neural networks
Ghazali, Hussain, & Liatsis, 2011	neural networks
Sermpinis, Laws, Karathanasopoulos, & Dunis, 2012	neural networks
Vanstone & Finnie, 2010	neural networks
Bahrepour, Akbarzadeh-T, Yaghoobi, & Naghibi-S, 2011	fuzzy logic
Huang, Chuang, Wu, & Lai, 2010	Support Vector regression
Premanode & Toumazou, 2013	Support Vector regression
Mabu, Hirasawa, Obayashi, & Kuremoto, 2013	rule-based genetic network programming

Tabulka 1.2: Přehled prací zabývající historií kurzů na finančních trzích. Převzato z [43]



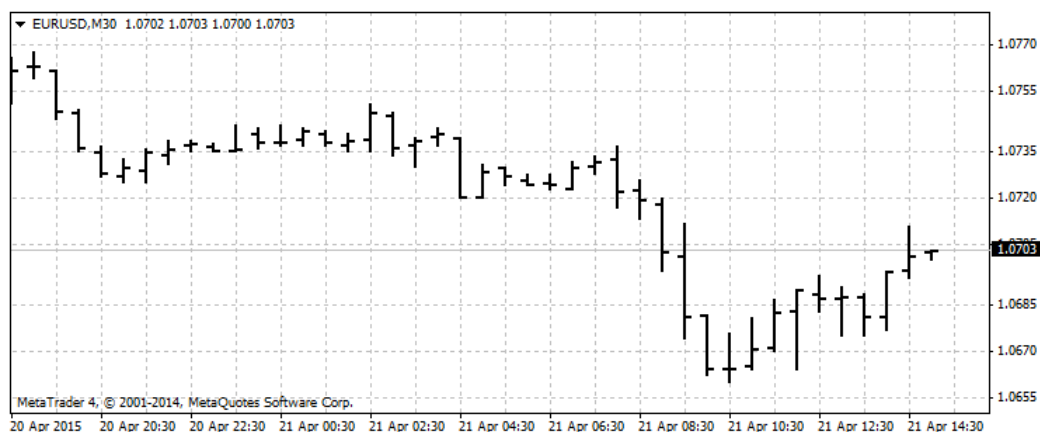
Obrázek 1.1: Čárový graf za období 20.4.2015 18:00 až 21.4.2015 14:30 trhu EURUSD. S periodou 30 minut. Získáno programem MetaTrader.

Automatickou technickou analýzou se zabývají práce uvedené v přehledu 1.2. Pro predikci používají různé typy strojového učení a vstupů [43].

1.1.4 Grafické znázornění pohybu ceny

V technické analýze se používají různé druhy grafů, které zobrazují pohyb ceny v průběhu daného období. Jde tedy o závislost ceny trhu na času. Nejrozšířenější jsou tyto tři:

- Čárové grafy (Obrázek 1.1) jsou pro svou jednoduchost přednostně využívány médii. Zobrazují cenu (obvykle průměrnou nebo zavírací) aktiv (např. komodit nebo měn) během předem stanoveného, např. tříměsíčního období. Jsou užitečné pro rychlé získání přehledu o střednědobém či dlouhodobém trendu aktiva [55].
- Sloupcové grafy (Obrázek 1.2), někdy nazývané též OHLC (open-high-low-close), jsou grafickým vyjádřením cenových úrovní za určité období. Jsou znázorněny jednou svislou čarou a dvěma vodorovnými krátkými čárkami z každé strany svislé čáry v místě, kde svislá čára představuje pohyb ceny za časovou jednotku (např. minutu, půl hodiny, hodinu, den, aj.). V počátečním okamžiku časové jednotky je uvedena vstupní



Obrázek 1.2: Sloupcový graf za období 20.4.2015 18:00 až 21.4.2015 14:30 trhu EURUSD. S periodou 30 minut. Získáno programem MetaTrader.

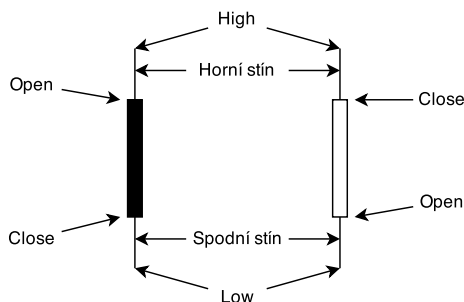


Obrázek 1.3: Svíčkový graf za období 20.4.2015 18:00 až 21.4.2015 14:30 trhu EURUSD. S periodou 30 minut. Získáno programem MetaTrader.

cena a vodorovná čárka se zobrazí na levé straně svislé čáry. Konečná cena se zobrazí stejným způsobem (ale na pravé straně svislé čáry) a může být vyšší nebo nižší než otevírací cena. Během zmíněného období se zobrazují maximální (High) a minimální (Low) ceny, na nichž závisí délka svislé čáry, protože Low se nachází v nejnižší části čáry (nejnižší cenová úroveň v dané časové jednotce) a High se nalézá v nejvyšší části čáry (nejvyšší cenová úroveň v dané časové jednotce) [55].

- Na první pohled složitější, ale nejvíce vypovídající grafy jsou svíčkové, které jsou zároveň pro tuto práci důležité. Jsou zastoupeny na všech elektronických platformách a investičních webech [53]. Svíčkové grafy (Obrázek 1.3) nám mohou prozradit mnohem více, než klasické OHLC grafy a v mnohých případech i mnohem spolehlivěji pomoci předvídat budoucí směr trhů [56].

Svíčkový graf se skládá z jednotlivých svíček s knoty na obou koncích [23], které ukazují pohyb ceny za určitou časovou jednotku. Na obrázku 1.4 si popíšeme svíčky. Vpravo na obrázku je svíčka s bílým tělem poukazující na vzestup ceny trhu, černá pak na pokles. Střední část svíčky neboli tělo se nazývá real body. Černé tělo (black



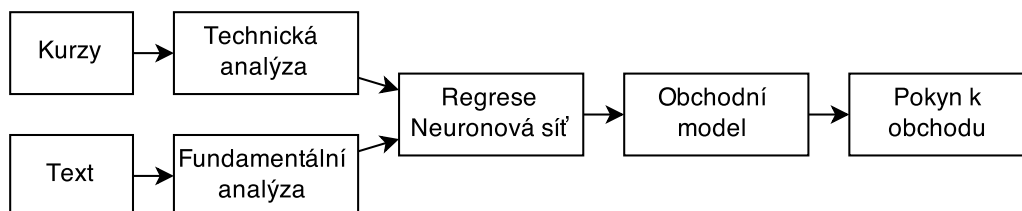
Obrázek 1.4: Popis jednotlivých svíček.

real body) reprezentuje svíčku, jejíž uzavírací hodnota (close) byla níže, než otevírací hodnota (open). To tedy znamená, že svíčka s černým tělem nám reprezentuje tak zvanou bearish (medvědí) periodu. Bílé tělo svíčky (white real body) nám naopak reprezentuje situaci, kdy bylo close výše než open, tudíž se jedná o bullish (býčí) periodu na grafu. Periodou samozřejmě myslíme časový úsek, který svíčka zobrazuje. Na denních grafech je to jeden den, na 5minutových grafech 5 minut, apod. Tenká vertikální čára nad tělem nebo pod tělem se nazývá horní stín a spodní stín (upper/lower shadow) a reprezentuje nám extrémní high a low cen dané periody. Z takto graficky znázorněné ceny je jednoduší vyčíst, zda cena trhu stoupla nebo klesla za danou periodu [56].

1.2 Schéma navrženého systému

Ze studia literatury a předchozí práce vyplývá, že bude vhodné technickou i fundamentální analýzu zkombinovat [52, 43].

Celý systém (schéma 1.5) automatického rozhodování funguje na technické (kapitola 3) a fundamentální analýze (kapitola 4). Za účelem regrese jsem se rozhodl využít neuronové sítě (kapitola 2). Podle odezvy neuronové sítě se rozhoduje obchodní model (kapitola 2.5.1), který poté zadává pokyny k obchodování.



Obrázek 1.5: Základní schéma systému.

Kapitola 2

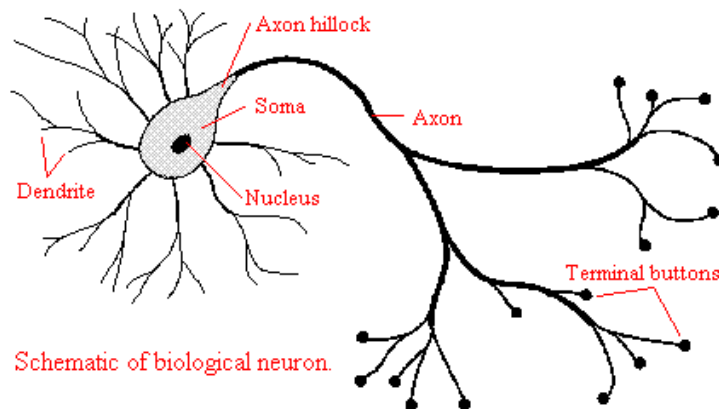
Neuronové sítě

2.1 Inspirace v biologii

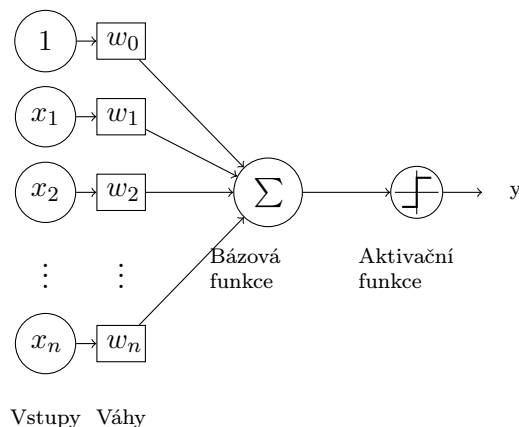
Mozek je jeden z nejkompexnějších systémů vůbec. Činnost mozku je inspirací pro umělou inteligenci už jejího vzniku. V mozku je přibližně 50-100 miliard neuronů a přibližně 10^{15} synaptických spojení (propojení neuronů). Neuron (obrázek 2.1) je vysoce specializovaná buňka, schopná přijmout, vést, zpracovat a odpovědět na speciální signály. Ke své činnosti používá elektro-chemické signály. Dendrity reprezentují místa vstupu signálů do těla neuronu. Tělo buňky (soma) zpracovává signály dané okolními neurony. Takto stanovený vnitřní potenciál vede k excitaci (vybuzení) neuronu. Axonové vlákno přenáší signál daný stupněm excitace k synapsím [24, 48, 58].

2.2 Perceptron

Umělý neuron se nazývá perceptron. Schéma perceptronu můžeme vidět na obrázku 2.2. Podobně jako biologický neuron má perceptron vstupy x_i a výstup y . Vnitřní hodnota u se spočítá na základě bázové funkce, jako suma součinů vah w_i a vstupů x_i . První vstup x_0



Obrázek 2.1: Biologický neuron. Převzato z [24].



Obrázek 2.2: Umělý perceptron. Převzato z [13].

je pevně připojen na hodnotu 1 a je násoben váhou w_0 a nazývá se bias:

$$u = \sum_{i=0}^n w_i x_i \quad (2.1)$$

Výstup je spočítán aplikací sigmoidní funkce na vnitřní hodnotu u . A spočítá se podle vztahu, kde λ je konstanta určující sklon funkce:

$$y = \frac{1}{1 + e^{-\lambda u}} \quad (2.2)$$

Proces nastavování vah se nazývá učení. Pokud jsou váhy správně nastavené, dokáže perceptron rozlišit lineárně separovatelné třídy [24, 48, 58].

2.3 Topologie

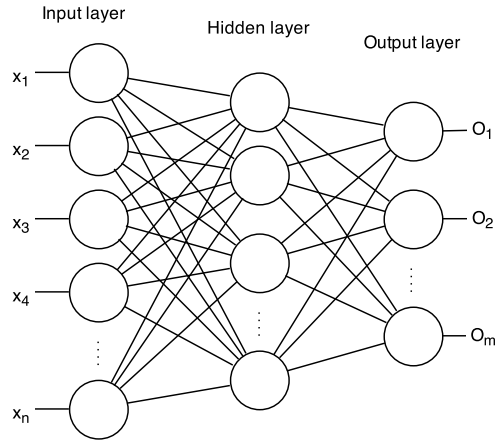
Jednotlivé perceptrony se propojují do komplexnějších sítí, které dokáží řešit problémy i lineárně neseparovatelné. Existují různé typy umělých neuronových sítí: Hopfieldova síť, Rekurzivní neuronová síť, Kohonenovy samoorganizující se mapy a jiné [48, 58]. Z těch jsem vybral plně propojenou dopřednou síť. Topologii dopředné neuronové sítě můžeme vidět na obrázku 2.3.

2.4 Trénování

Učení neuronových je možné provádět více způsoby. Například evoluční strategií, nebo horolezeckým algoritmem. Ale nejpoužívanější je algoritmus backpropagation, který také využívám ve své práci.

2.4.1 Backpropagation

Vstup algoritmu backpropagation 1 je množina vzorků E . Každému vzorku odpovídá množina vstupů $I(e)$ a množina požadovaných výstupů $O(e)$, což jsou při regresi požadované cílové hodnoty 2.5.3. Učení probíhá v cyklech. V každém cyklu probíhají postupně pro každý vzorek nastavení vstupů, vypočítání odezvy, zjištění chyby výstupu a zpětná propagace chyby.



Obrázek 2.3: Topologie dopředné neuronové sítě. Převzato z [14].

Počáteční hodnotu konstanty η jsem testoval na hodnoty v rozmezí 0,001 až 0,1. Z nichž nejlepší poměr rychlosti konvergence vůči přesnosti vykazovala hodnota 0,01. Cyklus učení probíhá dokud se krosvalidační chyba se snižuje (sekce 2.4.2).

2.4.2 Křížová validace

Pro předčasné ukončení trénování jsem využil techniku křížové validace. Množina vzorků je rozdělena na dvě: trénovací a validační část. Trénovací část množiny je vstupem do algoritmu backpropagation. Na konci každého trénovacího cyklu je spočítána celková chyba E_t u všech trénovacích vzorků:

$$E_t = \sum_{e \in E} (O(e) - T(e))^2 \quad (2.3)$$

a také chyba u validačních vzorků E_v z množiny validačních vzorků V :

$$E_v = \sum_{e \in V} (O(e) - T(e))^2 \quad (2.4)$$

Hypoteticky by měly mít chyby průběh daný obrázkem 2.4 [51]. Aby síť nebyla přetrénovaná, jakmile se krosvalidační chyba přestane snižovat, je učící konstanta η snižována vždy na polovinu. Také váhy, které byly nastaveny v poslední cyklu jsou zahozeny a přepsány zálohou vah z předchozího kroku [39].

2.4.3 Adaptivní trénování

Při testování strategie jsem vyzoroval, že i dobře natrénovaná neuronová síť funguje pouze krátkou dobu. Po delší době téměř v každém testovacím vzorku strategie už nedokáže s takto nastavenými parametry vykazovat zisk. Díky tomu jsem došel k závěru, že bude nutné trénovat neuronovou síť v průběhu experimentu na nově dostupných datech.

Trénování neuronové sítě je poměrně výpočetně náročné a proto jsem zvolil týdenní interval, po kterém se vždy neuronová síť znovu trénuje s novými daty.

Algoritmus 1: Backpropagation učící algoritmus. Převzato z [27].

Input: Set of examples E

foreach *example* e *in a set of examples* E **do**

- | Calculate $O(e)$ for $I(e)$ with Feedforward ;
- | Call function **CalculateOutputDeltas**($O(e), T(e)$);
- | Call function **CalculateInternalDeltas**;
- | Call function **UpdateWeights**;

end

Feedforward: foreach *layer* i **do**

- | **foreach** *neuron* j *in this layer* **do**
- | | Calculate $\sum_{k=0}^{\text{input layer}} w_{ijk} O_{ik}$;
- | **end**

end

CalculateOutputDeltas($O(e), T(e)$):

Get output values $O(e)$ from the output layer neurons;

foreach *individual output value* $O(e)_i$ **do**

- | Calculate error ϵ as $O(e)_i - T(e)_i$;
- | Calculate $\delta_{O(e)_i} = \partial f(O(e)_i) \times \epsilon$;
- | Add $\delta_{O(e)_i}$ to set of deltas Λ

end

CalculateInternalDeltas:

Let Λ_{i+1} be the next layer's set of deltas;

foreach *from non-output to input layer* i **do**

- | **foreach** *neuron* j *in this layer* **do**
- | | Initialize error ϵ as 0.0;
- | | **foreach** *neuron* k *of the next layer* **do**
- | | | Calculate ϵ as $\epsilon + \Lambda_{i+1,k} w_{ijk}$;
- | | **end**
- | | $\Lambda_{i,j} = \partial f(\epsilon \times \text{neuron } j\text{'s output})$;
- | **end**

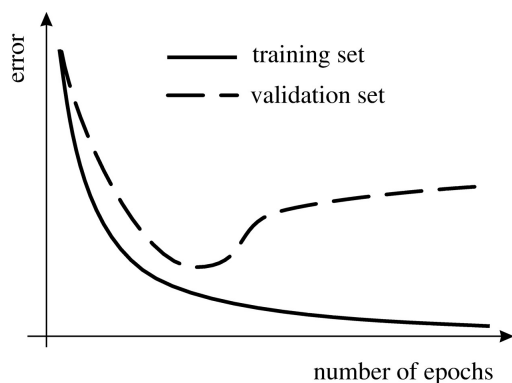
end

UpdateWeights:

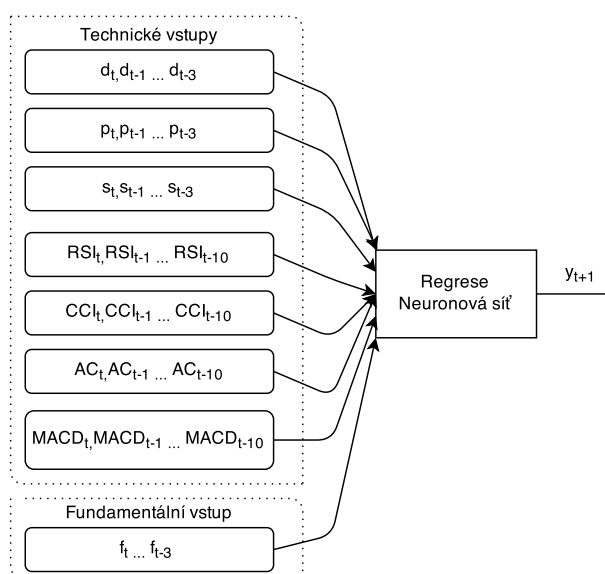
foreach *layer* i **do**

- | **foreach** *neuron* j *in this layer* **do**
- | | **foreach** *neuron* k *of the next layer* **do**
- | | | Calculate Δw_{ijk} as $\Lambda_{i,j} \times \text{neuron } j\text{'s output}$;
- | | | $w_{ijk} \leftarrow \eta \times \Delta w_{ijk}$;
- | | **end**
- | **end**

end



Obrázek 2.4: Graf závislosti chyby na trénovacích cyklech. Převzato z [26].



Obrázek 2.5: Schéma vstupů a výstupu neuronové sítě.

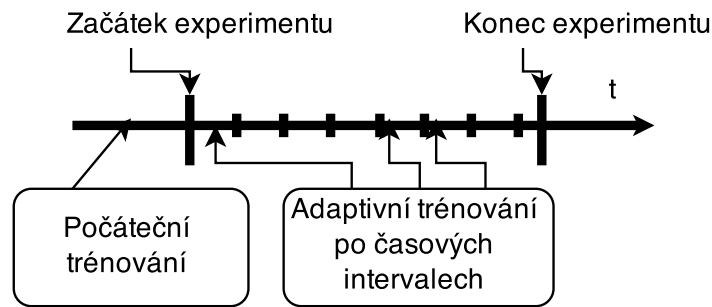
2.5 Konkrétní použití

Základní schéma vstupů a výstupů neuronové sítě můžeme vidět na obrázku 2.5. Neuronová síť se běžně používá pro klasifikaci nebo regresi (predikci). V práci je neuronová síť použita jako regrese, předpovídá budoucí hodnotu y_{t+1} (sekce 2.5.3).

Výstup (sekce 2.5.3) neuronové sítě je jediné číslo. Pro trénování (sekce 2.4) neuronové sítě jsem použil algoritmus Backpropagation a trénování probíhá jak na začátku tak i v průběhu na nově dostupných datech (sekc 2.4.3).

2.5.1 Důležitost času

Na obrázku 2.6 můžeme vidět časovou osu experimentu. Prediktor se natrénuje na minulém vývoji trhu a pak rozpoznává v reálném čase. Všechny hodnoty jsou závislé na čase, a proto jsem zavedl proměnou t pro aktuální čas. Pro historii je používána $t - i$.



Obrázek 2.6: Časová osa experimentu.

2.5.2 Vstupy

Vstupy jsou kombinací technické analýzy i analýzy fundamentální. Mezi technické parametry patří svíčkové formace, indexy Index relativní síly (RSI), Acceleration/Deceleration Technical Indicator (AC) a Commodity Channel Indicator (CCI) a Moving Average Convergence/Divergence Indicator (MACD) k fundamentálním vstupním datům se řadí titulky zpráv a forexní kalendář. Počty vstupů můžeme vidět na diagramu 2.5 a jsou 3 vstupy pro svíčkové formace a fundamentální vstupy a 10 pro ostatní. Technické a fundamentální vstupy jsou popsány v samostatných kapitolách 3 resp. 4.

2.5.3 Výstup

Toto číslo určuje o kolik a v kterém směru se trh pohnul v časovém rozmezí $t = 0$ až $t = +5$. Udává tedy relativní pohyb budoucí ceny od ceny aktuální. A vypočítá se na základě následujícího vzorce:

$$y_t = \left(\max_{High_t \dots High_{t+5}} - Open_t \right) - \left(Open_t - \min_{Low_t \dots Low_{t+5}} \right) \quad (2.5)$$

Tedy se zjišťuje pro 5 budoucích hodnot a pro aktuální čas tato hodnota není známá, pouze jako odezva naučené neuronové sítě. Pokud se trh pohne v o stejnou velikost v obou směrech není vhodné vstupovat do obchodu, což je zaručeno hodnotou 0. Pokud se trh pohne do kladného směru, maximum převáží minimum a výsledek bude nenulový. Pokud se trh pohne dostatečně hodně, tedy jestli přesáhne práh, pak dojde ke vstupu do obchodu.

Kapitola 3

Technické vstupy

Všechny technické vstupy jsou normalizovány (sekce 3.6). Jako vstupy neuronové sítě jsem se rozhodl použít následující:

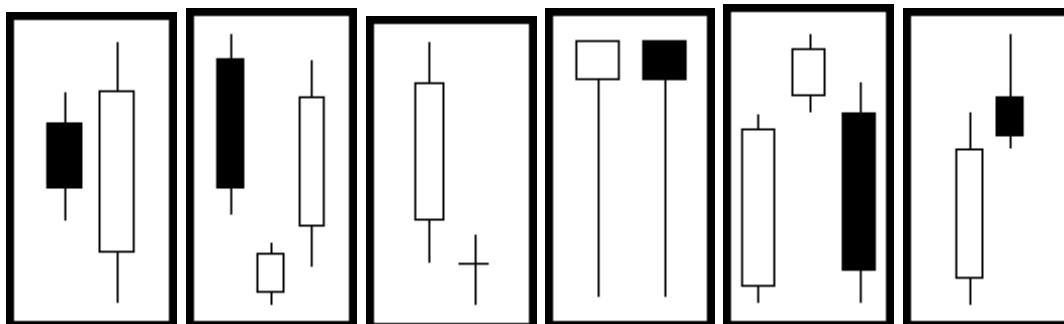
3.1 Svíčkové formace

Často opakující se posloupnost tří svíček značí změnu trendu trhu se nazývá svíčková formace. Různé kombinace různých svíček nám nabízejí různé signály, na základě kterých se otevírá či uzavírá (většinou v kombinaci s dalšími signály) obchod. Tyto svíčkové formace se používají pro podporu rozhodování, proto se je budeme snažit využít pro predikci. Svíčkové formace se často v obměnách opakují [53, 30, 37, 56].

Svíčkových formací existuje mnoho, ale pro ilustraci uvedu několik příkladů takových vzorců. První tři formace na obrázku 3.1 signalizují býčí (stoupající) trend a zbylé signalizují medvědí (klesající) trend.

Z těchto obrázků lze vidět, že tyto formace závisí na velikosti, pozici a barvě těla. a právě tyto parametry se snažím popsat reálnou hodnotou pro vstup do neuronové sítě. Formace jsou stejné pro stoupající i pro klesající trend, pouze jsou zrcadlově převrácené podobně jako některé z obrázků 3.1.

Protože svíčkové formace nikdy nejsou identické, je nutné vstupy zobecnit. Hodnoty $Open_t$, $Close_t$, $High_t$, Low_t jdou vidět na obrázku 1.4. Většina svíčkových formací je složena ze tří svíček, u kterých rozhoduje hlavně délka těla, pozice těla vůči stínům a směr svíčky (Obrázek 1.4). Tyto tři parametry d_t , p_t , s_t jsem se rozhodl použít jako vstupy.



Obrázek 3.1: Svíčkové formace. Převzato z [56]

Délka těla (d_t) je rozměr od otvírací ceny ($open$) po zavírací cenu ($close$) a může nabývat hodnot $\langle 0, 1 \rangle$.

$$d_t = \frac{|Open_t - Close_t|}{|High_t - Low_t|} \quad (3.1)$$

Pozice těla (p_t) je rozměr od maximální ceny ($high$) po střed těla svíčky a může nabývat hodnot $\langle 0, 1 \rangle$.

$$p_t = \frac{High_t - Close_t + \frac{|Open_t - Close_t|}{2}}{|High_t - Low_t|} \quad (3.2)$$

Směr svíčky (s_t) je záporné nebo kladné číslo 1 podle toho jestli je větší vstupní cena nebo zavírací. Graficky se u svíčkových grafů tato hodnota znázorňuje jako černé, nebo bílé tělo svíčky.

$$s_t = \begin{cases} 1 & \text{pro } 0 < (Open_t - Close_t) \\ -1 & \text{pro } 0 > (Open_t - Close_t) \end{cases} \quad (3.3)$$

Hodnota $PriceMiddle_t$ se spočítá jako střední hodnota maxima a minima ceny v časovém úseku.

$$PriceMiddle_t = \frac{High_t + Low_t}{2} \quad (3.4)$$

3.2 Index relativní síly (RSI)

Z anglického Relative Strength Index, index měřící dynamiku vývoje ceny [8]. Pohybuje se v rozmezí 0 až 100. Trh je považován za překoupený, pokud je index víc než 70, přeprodáný pokud je index menší než 30. Je definován následovně:

$$AG_t(n) = \frac{AG_{t-1}(n)(n-1) + currentGain}{n} \quad (3.5)$$

$$AL_t(n) = \frac{AL_{t-1}(n)(n-1) + currentLoss}{n} \quad (3.6)$$

$$AG_0(n) = \frac{\sum_{i=0}^n currentGain}{n} \quad (3.7)$$

$$AL_0(n) = \frac{\sum_{i=0}^n currentLoss}{n} \quad (3.8)$$

$$RSI_t(n) = 100 - \frac{100}{1 + \frac{AG_t(n)}{AL_t(n)}} \quad (3.9)$$

Hodnoty AG_t a AL_t jsou průměrný zisk resp. průměrná ztráta. Hodnoty $currentGain$ a $currentLoss$, jsou definovány jako aktuální zisk, resp. ztráta. Vypočítají se jako rozdíl aktuálního close a předchozího close [49].

3.3 Commodity Channel Indicator (CCI)

Indikátor CCI měří odchylku ceny od průměrné hodnoty [3]. Vyšší hodnota ukazuje, že cena je výše než průměr, nízká poukazuje na opak. Hodnota tohoto indikátoru se obvykle pohybuje v rozsahu -250 až 250 a je definován následovně:

$$MD(n) = \frac{|TP_{t-n} - SMA_t(n)| + |TP_{t-(n-1)} - SMA_t(n) + \dots + |TP_t - SMA_t(n)|}{n} \quad (3.10)$$

$$CCI_t(n) = \frac{TP_t - SMA_t(n)}{0,015 \cdot MD(n)} \quad (3.11)$$

Hodnota $MD(n)$ je střední odchylka pro n historických hodnot, TP_t je typická hodnota v čase t :

$$TP_t = \frac{high + low + close}{3} \quad (3.12)$$

a $SMA_t(n)$ je jednoduchý klouzavý průměr hodnot TP .

3.4 Acceleration/Deceleration Technical Indicator (AC)

Indikátor AC měří rozdíl mezi klouzavými průměry. Změna směru indikátoru značí změnu trendu trhu. Vypočítá se jako rozdíl klouzavých průměrů z období 5 a 34 historických hodnot střední hodnoty ceny [1]. V mém případě po 30 minutovém intervalu.

$$AO_t = SMA_t^{PriceMiddle}(5) - SMA_t^{PriceMiddle}(34) \quad (3.13)$$

$$AC_t = AO_t - SMA_t^{AO}(5) \quad (3.14)$$

3.5 Moving Average Convergence/Divergence Indicator (MACD)

MACD je jedním z nejvíce populárních indikátorů v technické analýze. Jeho výpočet vyžaduje tři exponenciální klouzavé průměry (EMA), které jsou zpožděné indikátory pokračujícího nebo obracejícího se trendu. Tyto zpožděné indikátory jsou konvertovány na hybnost oscilátoru odvozeného od rozdílu pomalého klouzavého průměru vůči rychlému klouzavému průměru. Rychlý znamená kratší klouzavý průměr, standardně 12, a pomalý delší periodu, standardně 26 [18, 22].

$$MACD_t = EMA_t^{12-day} - EMA_t^{26-day} \quad (3.15)$$

$$EMA_t = \frac{2}{n}P_t + (1 - \frac{2}{n})EMA_{t-1} \quad (3.16)$$

3.6 Normalizace technických vstupů

Všechny technické vstupy jsou normalizovány pomocí z-score normalizace. Z-score vyjadřuje vzdálenost dané hodnoty od střední hodnoty množiny. Vstupní data normalizujeme tak,

že od každé hodnoty odečteme střední hodnotu vstupních dat a poté výsledek vydělíme velikostí směrodatné odchylky [15, 38]:

$$x' = \frac{x - \bar{x}}{\sigma_x} \quad (3.17)$$

, kde střední hodnotu \bar{x} spočítáme jako průměr hodnot z trénovací množiny [38]:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (3.18)$$

a směrodatná odchylka σ_x spočítáme podle vztahu [38]:

$$\sigma_x = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad (3.19)$$

Kapitola 4

Fundamentální vstupy

4.1 Titulky zpráv

Podobně jako autor článku [44] jsem se rozhodl dolovat titulky zpráv ze zpravodajských serverů zaměřených na finanční trhy. Titulky zpráv jsou staženy jako textová data, způsob získání a rozsah je popsán v kapitole 7.2.

Po stažení máme texty titulků s časovými razítky (tabulka 4.1). Texty jsou agregovány po 30 minutových intervalech. Množiny slov D_t pro časový údaj t obsahují slova z celého 30 minutového intervalu. Příklad těchto titulků můžeme vidět v tabulce 4.1.

4.2 Forexní kalendář

Textová data jsem rozšířil o další zdroj. Textová data makroekonomických událostí nazývaný také forexní kalendář. Kalendář obsahuje nejdůležitější události a vyhlášení vydaná vládou a národními bankami. Například procento nezaměstnanosti nebo hrubý domácí produkt. Podobný zdroj dat je použit v práci [21]. Texty jsou také vygenerovány po 30 minutových intervalech a přidány do množiny D_t . Příklad těchto událostí můžeme vidět v tabulce 4.2.

4.3 Zpracování textových dat

Vstupní data jsou promazána o slovník stopwords S . Slovník byl stažen se serveru¹. Tento slovník obsahuje slova, které v sobě nenesou žádnou informaci pro analýzu. Například se

¹<http://www.ranks.nl/stopwords>

2014-11-19 11:45:34, Trade Idea Update: EUR/USD - Hold long entered at 1.2460
2014-11-19 11:43:43, Trade Idea Update: USD/JPY - Buy at 117.00
2014-11-19 11:29:28, Bank Of Japan Board Probably Still Divided Despite 8-1 Vote
2014-11-19 11:25:41, USD/JPY will hit 1.20 in a month - RBS
2014-11-19 11:23:44, Technical Analysis for Crosses

Tabulka 4.1: Ukázka stažených titulků.

2014-10-03 12:30:00,GMT,usd,USD Unemployment Rate,High,5.9%,6.1%,6.1%
2014-10-03 12:30:00,GMT,usd,Underemployment Rate ,Low,11.8%,12.0%
2014-10-03 13:45:00,GMT,usd,USD Markit US Composite PMI,Low,59.0,58.8
2014-09-24 14:00:00,GMT,usd,USD New Home Sales,Medium,504K,430K,427K
2014-09-24 14:00:00,GMT,usd,USD New Home Sales (MoM),Low,18.0%,4.4%,1.9%

Tabulka 4.2: Ukázka stažených událostí kalendáře.

10093 trade
9605 us
8393 daily
8389 technical
8266 outlook
7430 dollar
7235 analysis
6712 eur/usd
6129 data
6098 usd/jpy
5995 market
5420 euro
4880 gbp/usd
4782 forex
4339 orders
4228 idea
4119 trading

Tabulka 4.3: Ukázka nejčastějších slov v datové sadě

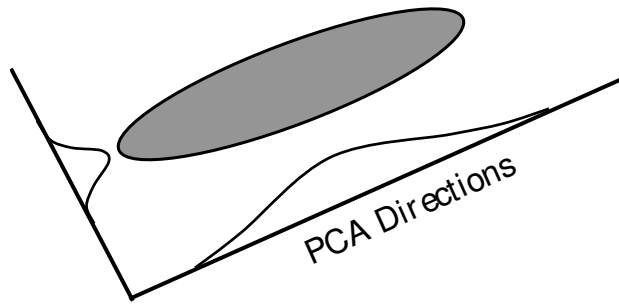
jedná o slova: *a*, *am*, *the*, *for*, *some* a jiné.

$$T_t = D_t \setminus S \quad (4.1)$$

Z vstupních dat je vytvořen slovník nejčastějších slov W (tabulka 4.3) s 500 slovy. Na základě tohoto slovníku je vytvořen vektor jedniček a nul indikující zda se slovo v daném intervalu vyskytuje či nikoliv.

$$F_{ti} = \begin{cases} 1 & \text{pro } W_i \in T_t \\ 0 & \text{pro jinak} \end{cases} \quad (4.2)$$

Tyto vektory výskytů slov jsou příliš dlouhé pro zpracování a proto je nutná redukce dimenze dat. Pro redukci dimenzionality jsem použil techniku Principal component analysis (PCA).



Obrázek 4.1: PCA pro 2-rozměrná data. Převzato z [20].

4.4 Principal component analysis

PCA vyhledává takovou lineární transformaci ortogonálních bází, kde první báze ukazuje směr největší variability v trénovacích datech N -rozměrného prostoru vstupních vektorů. Druhý bázový vektor je kolmý na první a udává druhý směr největší variability v datech. Limit této techniky spočívá v předpokladu, že vstupní data mají Gaussovo rozložení [40, 50, 20].

Na obrázku 4.1 je demonstrován efekt pro 2-rozměrná data. Šedá elipsa představuje rozložení dat, osy zobrazují nové souřadnice (směr) získané pootočením originálních os použitím PCA transformace [40, 50, 20].

Bázové vektory PCA transformace jsou dány vlastními vektory kovarianční matice, která je spočítána z trénovacích dat podle následující rovnice:

$$\mathbf{C} = \frac{1}{N} \sum_{i=1}^N (x_i - m)(x_i - m)^T \quad (4.3)$$

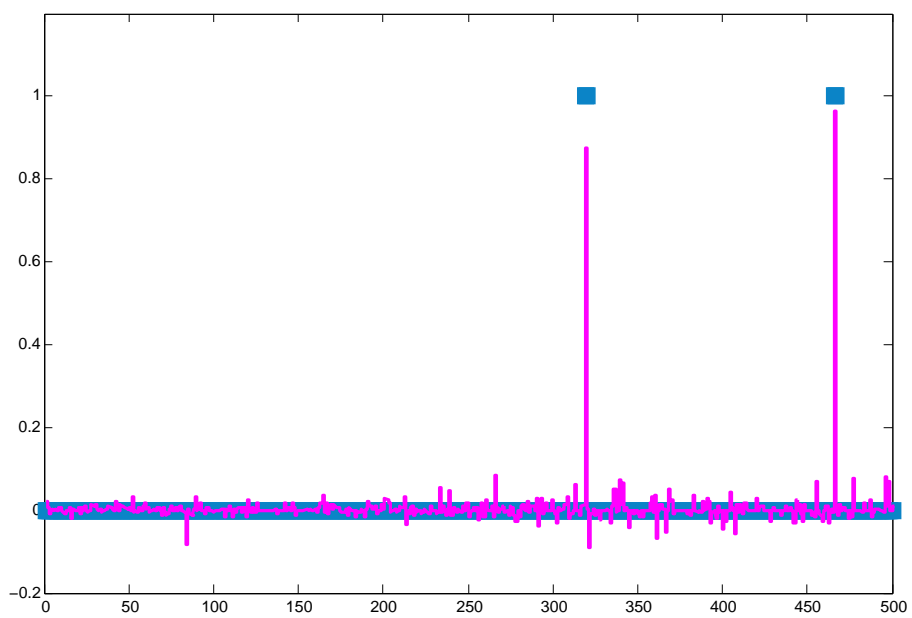
kde C je kovarianční matice získaná z N vstupních vektorů k dispozici pro trénování, x_i je i -tý trénovací vektor a m je odhadovaný střední vektor:

$$\mathbf{m} = \frac{1}{N} \sum_{i=1}^N x_i \quad (4.4)$$

Vlastní hodnota spojená s jedním vlastním vektorem je reprezentována množstvím variability získané projekcí vstupního vektoru do daného vlastního vektoru. Proto pouze několik vlastních vektorů odpovídajících největším vlastním hodnotám jsou použity jako báze pro účely redukce dimenze [20].

4.4.1 Konkrétní použití

V mém případě byla redukována dimensionalita pomocí techniky PCA z původních 500 na 50 dimenzí. Dále budeme jednotlivé vektory po redukcí značit f_t v závislosti na čase. Na grafu 4.2 můžeme vidět hodnoty jednoho vektoru před redukcí dimenze (modré body) a po redukcí a zpětné rekonstrukci na dimenzi 500 (fialová křivka). Původní vektor obsahuje pouze dvě jedničky ty jsou znázorněny čtverci v hodnotě 1. Ostatní hodnoty jsou rovné nule a zobrazené na ose. Redukovaný vektor znázorněný fialovou křivkou má většinou hodnot v rozmezí -0.1 až 0.1 kromě dvou hodnot, kde je požadovaná hodnota co nejbližší jedné. Můžeme vidět, že tento případ 50 dimenzí dokáže popsat.



Obrázek 4.2: Příklad redukce dimenze. Závislost hodnoty vektoru na dané položce.

Kapitola 5

Obchodní model

O průběh experimentu se stará platforma Meta Trader, která počítá zisk i s odečtením poplatků brokerovi. Samotný obchodní model pouze zadává pokyny k otevření, uzavření obchodu. Celkový zisk je spočítán jako rozdíl mezi otevřením a uzavřením obchodu.

Strategie pracuje pouze na začátku intervalu. Tedy pokud je použit interval 30min, strategie pracuje jedenkrát za 30min. Provede zjištění výstupu neuronové sítě na základě aktuálních vstupů, poté podle výstupu vstoupí do obchodu, pokud je přesáhnut práh. Pokud je již otevřený nějaký obchod, musí být přesáhnut práh v opačném směru obchodování a starý obchod je ukončen. Pokud uběhl týden od posledního trénování sítě, je spuštěno trénování nové. Tento postup je zobrazen na diagramu 5.1.

5.1 Otevření obchodu

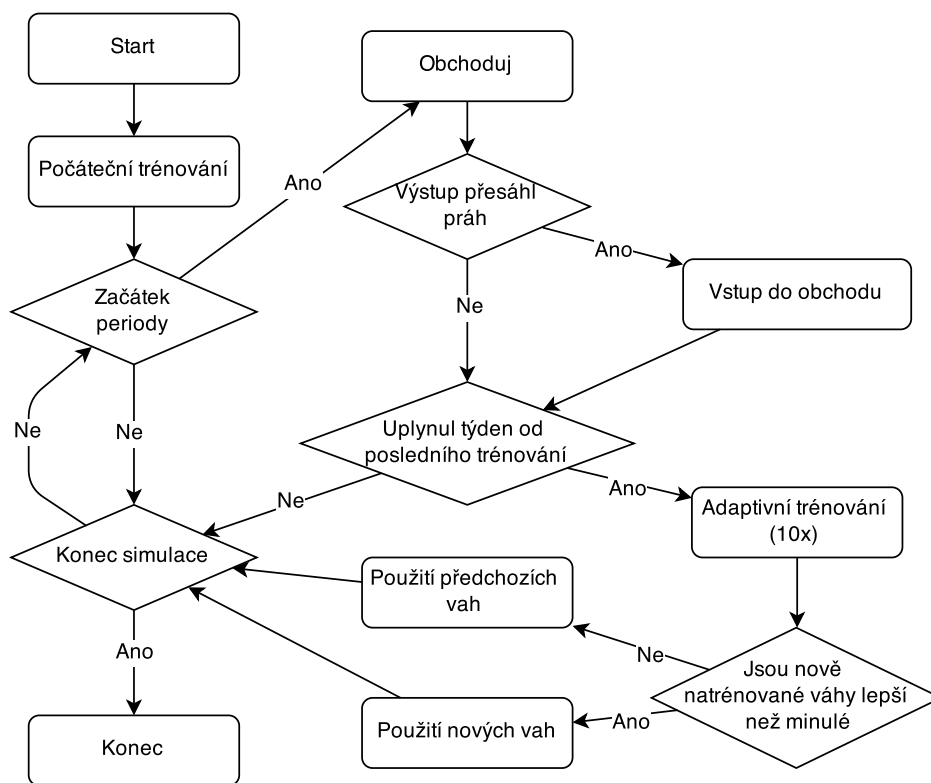
Na základě výstupu se rozhoduje, zda se bude vstupovat do obchodu. Pokud výstup přesáhne prahy, vstoupím do obchodu. Prahů jsou dva: pro kladný a pro záporný směr a jsou stejné v absolutní hodnotě. Podle toho se rozhoduje, zda nakoupit nebo prodat. Obchod může být ziskový nebo ztrátový. Hodnota prahu byla po několika experimentech stanovena na 0,4.

5.2 Ukončení obchodu

Dosud jsem mluvil pouze o rozhodování pro vstup do obchodu, ale důležité je i to, kdy se má obchod uzavřít [46]. Pro strategii založenou na trénování, kde je potřeba mít jasně dané jestli byl obchod ziskový nebo ztrátový, je vhodné použití statických "stoploss" a "takeprofit". Ty se nejčastěji uvádějí v bazický bodech (bts) [28]. To znamená, že se obchod ukončí po dosažení požadovaného zisku (takeprofit) respektive ztráty (stoploss). Konstanty byly po několika experimentech stanoveny na 20 bts.

5.3 Ověření ziskovosti nového trénování

Ukázalo se, že je potřeba ověřit v průběhu experimentu, že je neuronová síť dobře natrénována. Toho nelze docílit pomocí platformy Meta Traderu. V důsledku toho jsem vytvořil funkci na ověření zisku obchodního modelu na minulém vývoji, která spočítá zisk za uplynulé období s daným nastavením vah neuronové sítě.



Obrázek 5.1: Vývojový diagram obchodního modelu.

Před začátkem adaptivního trénování je vypočten zisk za uplynulé období se stávajícími váhami. Váhy jsou zálohovány. Poté je síť natrénována a vypočte se zisk za uplynulé období s novými váhami. Nové váhy jsou uplatněny pouze, pokud je zisk vyšší. V opačném případě se obchoduje dále se starými váhami. Přístup můžeme pozorovat na diagramu obchodní strategie 5.1. Takto technika zabraňuje tomu, aby adaptivní trénování uškodilo dobře natrénované strategii. Tato technika lépe propojuje odezvu sítě a její chyby s požadovaným výstupem a obchodní model a reálné obchodování. Výpočet zisku za minulé období se spočítá pomocí funkce CalculateProfit 2.

5.3.1 Násobná inicializace neuronové sítě

Toto ověřování zisku nově natrénované sítě mě vedlo na další postup. Počáteční náhodná inicializace sítě má jistý vliv na naučení neuronové sítě a poté také na její odezvu. Proto jsem se rozhodl po vzoru [39], trénovat několik neuronových sítí z nichž vybírám pouze tu, která má nejvyšší zisk spočítaný pomocí algoritmu 2 CalculateProfit. Testoval jsem 5 a 10 neuronových sítí. Experimenty můžeme vidět v kapitole 8.

5.3.2 Funkce hodnotící váhy

Podstata funkce CalculateProfit 2 spočívá v procházení nejbližší historie vývoje ceny trhu. Testuje jestli neuronová síť ($y_t = nn(w, t)$) s testovanými vahami ($w[]$) přesáhla hodnotu prahu (*Threshold*). Pokud ano, zapamatuje si čas (t), ve kterém k tomu došlo. Dále testuje jestli v otevřených obchodech (*Orders[]*) nedošlo ke změně, nebo-li jestli obchod nedosáhl (*StopLoss*) nebo (*TakeProfit*) a neuzavřel se. Pokud k tomu dojde, připočítá se zisk nebo ztráta k celkovému zisku (*profit*) a odečte poplatek. Poté se posune v čase dokud nedojde do aktuálního času. Jako výsledek funkce vrátí celkový zisk za uplynulé období (*profit*).

Algoritmus 2: Funkce CalculateProfit, Výpočet zisku za uplynulé období pro zadané parametry.

Input: ($w[]$, *Threshold*, *StopLoss*, *TakeProfit*, *Charges*)

Output: *profit*

```

1: for  $t = -1000$  to 0 do
2:    $y_t = nn(w, t)$ 
3:   if  $c > Threshold$  then
4:      $Orders[] = t$ 
5:   end
6:   foreach Order in  $Orders[]$  do
7:     if  $Open[Order] - StopLoss > Low[t]$  then
8:        $profit = profit - Charges - StopLoss$ 
9:     end
10:    if  $Open[Order] + TakeProfit < High[t]$  then
11:       $profit = profit - Charges + TakeProfit$ 
12:    end
13:  end
14: end
15: return profit

```

Kapitola 6

Implementace

6.1 Struktura systému

Na schéma 6.1 můžeme vidět propojení mezi jednotlivými moduly systému. Platforma Metatrader poskytuje technická data a také simulaci obchodování. V rámci této platformy běží obchodní systém. Ten obstarává rozhodování obchodování a přípravu vstupů pro neuronovou síť. Knihovna neuronové sítě je napojená na obchodní systém, obstarává trénování vah a také odezvu sítě. Další data, které obchodní systém zpracovává, jsou vektorová data výskytu slov a kovarianční matice pro redukci dimensionalitu pomocí PCA. Tato data jsou vytvořena v konzolovém programu respektive v Matlabu. Pro stažení textových dat z Internetu jsem vytvořil skripty v jazyce php.

6.2 Obchodní model

Pro implementaci svého přístupu jsem si zvolil platformu MetaTrader 4, který má vestavěný systém spouštění automatických obchodních systému (tzv. strategií). Tyto strategie se implementují v jazyku MQL4. MQL4 je zkratkou pro jazyk MetaQuotes Language 4. MetaQuotes¹ je společnost, která postavila základ pro systém MetaTrader Trading [11].

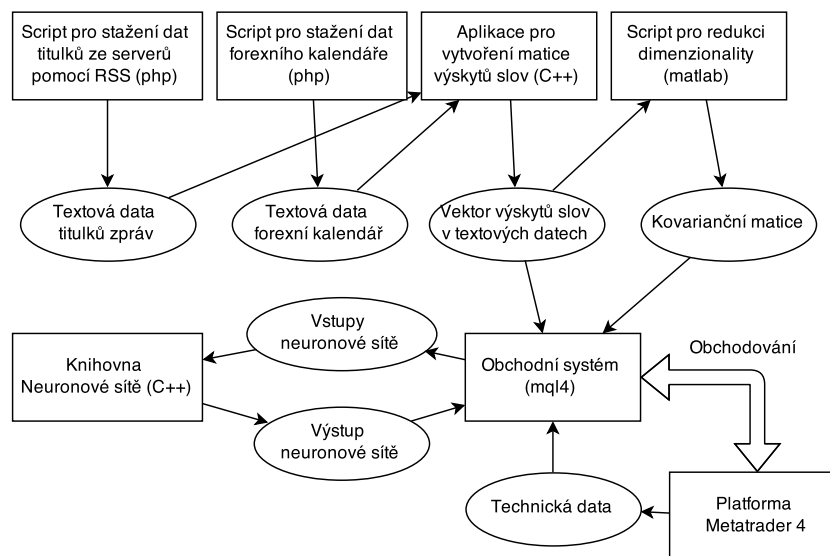
6.2.1 Programovací jazyk

Implementoval jsem svou strategii v jazyku MQL4, který je podtřídou jazyka C++. Umožňuje nám vytvářet vlastní strategie, indikátory a skripty. Strategie umožňují automatické obchodování, tedy výpočet historie a následného vstupu do obchodu. Indikátory jsou grafické prvky vykreslované přímo do grafu, zobrazující například klouzavý průměr. Skripty jsou jednorázově vykonané funkce, například pro zavření všech otevřených obchodů. Pro vyvíjení jsem využíval velmi dobře zpracovaného editoru MetaEditor. Ten obsahuje zvýrazňování textu, nápovědu, doplňování klíčových slov a vestavěný kompilátor [11].

6.2.2 Platforma

Pro simulování obchodu jsem měl několik možností. Z nich jsem se rozhodoval mezi použitím matematického prostředí Matlab nebo obchodní platformy MetaTrader 4. Rozhodl jsem se nakonec pro MetaTrader, který má několik výhod proti Matlabu:

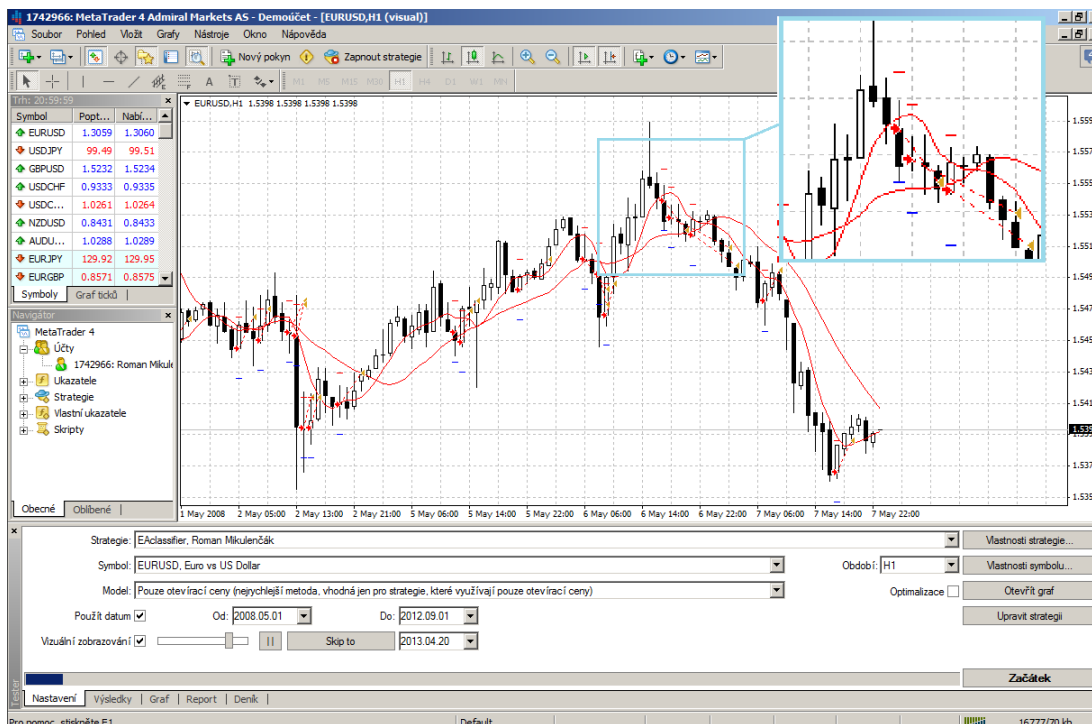
¹<http://www.metaquotes.net/>



Obrázek 6.1: Struktura systému.

- Patří mezi nejpoužívanější software pro obchodování [45]
- Je celosvětově uznávanou a využívanou platformou pro reálné obchodování [7]
- Jeho součástí je simulace obchodování automatického systému včetně poplatků [2]
- Má zabudovanou optimalizaci obchodních systémů, která spouští simulaci několikrát po sobě pro všechny optimalizované parametry [2]
- Jednoduché získání historických dat, kde je možné stáhnout data ze serveru brokera nebo od společnosti MetaQuotes, která tento program vyvíjí [2]
- Data lze exportovat a importovat [2]
- Grafické zobrazení průběhu simulace
- Možnost testovat strategie na reálných aktuálních datech v demo účtu, nebo i spuštění na reálném účtu
- Možnost vyvíjet strategie v prostředí MetaEditor, které má již v sobě zabudované již proměnné obsahující ceny trhu, indikátory jako klouzavé průměry, funkce pro vstup do obchodu a jiné [5]
- Není nutné se starat o udržování obchodu, načítání dat, ani nic podobného, vše je již součástí programu
- Výhodou je také, že je zcela zdarma [53]

Na obrázku 6.2 můžeme vidět program MetaTrader 4. Uprostřed v největší části je zobrazen grafický výsledek testování strategie. Jedná se o svíčkový graf se zobrazenými obchody, zobrazující vstup do obchodu v podobě pravé šipky a ukončení obchodu levé šipky. Také zde jdou vidět klouzavé průměry jako spojitě funkce. Ve spodní části se nachází



Obrázek 6.2: Náhled na program MetaTrader 4. Lze zde vidět svíčkový graf měn EURUSD v období 1.5.2008 až 7.5.

tester strategií [2]. Vlevo lze vidět aktuální hodnoty aktiv. Níže pak seznam účtů, strategií a indikátorů. V testeru strategií se spouští simulace obchodování automatického obchodního systému. Lze vybrat libovolný časový interval a obchodované aktivum, pro které máme dostupná data.

6.3 Knihovna neuronové sítě

Neuronová síť včetně Backpropagation algoritmu je napsaná v jazyce C++ v podobě dynamické knihovny [12], která je připojena ke strategii. Knihovna vychází z otevřeného zdrojového kódu². Dále je upravována pro možné použití a spojení se strategií v MQL4 jazyce.

6.4 Skript pro stažení dat titulků

Pro stažení fundamentálních dat jsem musel vytvořit skript. Skript stahuje data ze serveru feedly.com, který zprostředkovává historická data přibližně 18 měsíců zpětně z protokolů RSS, protože servery zprostředkovávající zprávy přes protokol RSS nabízejí pouze posledních pár zpráv.

Skript byl vytvořen podle dokumentace feedly.com api³. Pro skript byl použit skriptovací jazyk php.

Spuštění se provádí příkazem:

```
php feed.php > file.csv
```

²<http://www.codeproject.com/Articles/13582/Back-propagation-Neural-Net>

³<https://developer.feedly.com/>

kde soubor `file.csv` bude přepsán staženými daty. Výběr z kterého serveru se budou data stahovat se provádí zakomentováním/odkomentováním příslušného řádku na začátku skriptu.

6.5 Skript pro stažení dat forexního kalendáře

Podobně jsem vytvořil script pro stažení dat forexního kalendáře z serveru `dailyfx`⁴.

Spuštění se provádí příkazem:

```
php calendar.php > calendar.csv
```

kde soubor `calendar.csv` bude vytvořen s daty forexního kalendáře.

6.6 Program na vytvoření vektoru výskytu slov

Program na vytvoření vektoru výskytu slov je konzolová aplikace napsaná v jazyce C++. Aplikace otevře postupně všechny zdrojové soubory textů, načte data a jejich časové značky. Dále je agreguje po půl hodinách a vytvoří slovník nejčastějších slov. Ze slovníku je vybráno 500 nejčastějších. Také odstraní stopslova. Pro každý časový interval je zkontrolováno zda data obsahují nebo ne dané slovo. Od toho se poté odvíjí 0 nebo 1 ve výsledném výstupním souboru.

6.7 Skript na vytvoření kovarianční matice

Pro techniku PCA jsem použil Matlab. Script načte data vytvořená programem pro vytvoření vektoru výskytu slov. Zpracuje data funkcí `pca` a výsledná data uloží do souboru. Výsledný soubor kovarianční matice je poté načítána obchodní strategií a je pomocí ní snižován počet vstupů na 50 dimenzí.

⁴<http://www.dailyfx.com/calendar>

Kapitola 7

Data

Pro trénování bylo použito období od 1.6.2013 do 1.1.2014 a pro testování období od 1.1.2014 do 1.6.2014.

7.1 Technická data

Jako datovou sadu jsem si vybral trh EUR/USD, který se vyznačuje největší likviditou. Je to trh, na kterém obchoduje nejvíce traderů [10]. První dvě písmena ve zkratce trhu znamenají název země a třetí je název měny. Jedná se o měnový pár Euro vůči Dolaru [6]. Data jsem stáhl od společnosti Admiral Markets¹ v rámci vytvořeného demo účtu. MetaTrader si je stáhl automaticky, když byla zapotřebí. Data byly vyexportována a jsou uložena na přiloženém CD.

Použil jsem data se vzorkováním po 30 minutách. Každý interval je popsán 5 hodnotami $Open_t$, $Close_t$, $High_t$, Low_t , které zároveň vyjadřují svíčkové grafy. Grafické znázornění těchto dat můžeme vidět na obrázku 1.4. Poslední hodnota je objem obchodů. Příklad datové sady můžeme vidět v tabulce 7.1:

V tabulce 7.2 můžeme vidět statistické hodnoty daných dat. Ze statistických hodnot můžeme vidět, že průměrné hodnoty v jednotlivých sloupcích se téměř neliší. Tedy rozdíly mezi otevírací a uzavírací cenou v celé datové sadě jsou zanedbatelné.

Na grafu 7.1 můžeme vidět závislost jednoho vstupu neuronové sítě na požadovaném výstupu. Můžeme vidět, že data jsou velmi chaotická a mají spíše charakter rovnoměrného rozložení. Proto je obtížné s těmito daty pracovat.

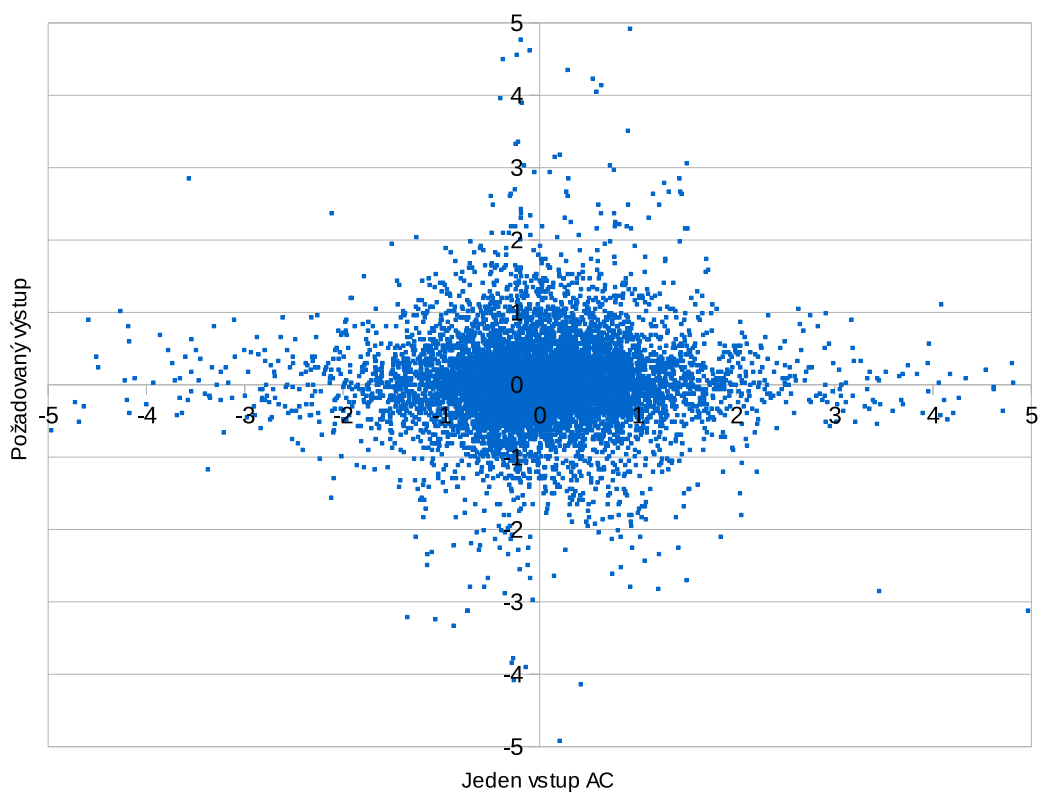
¹<http://www.admiralmarkets.cz/>

2014.04.03,17:30,1.3338,1.3342,1.3333,1.3336,820
2014.04.03,18:00,1.3336,1.3336,1.3322,1.3325,744
2014.04.03,18:30,1.3325,1.3328,1.3323,1.3328,701
2014.04.03,19:00,1.3328,1.3334,1.3327,1.3333,620
2014.04.03,19:30,1.3333,1.3334,1.3329,1.3331,510

Tabulka 7.1: Ukázka technických dat.

	Open	High	Low	Close
Minimum	1,277	1,2776	1,2755	1,277
Maximum	1,3964	1,3993	1,3952	1,3964
Průměr	1,3534	1,3539	1,3530	1,3534
Směrodatná odchylka	0,0254	0,0253	0,0255	0,0254
Suma	16572	16578	16566	16572

Tabulka 7.2: Statické hodnoty technické datové sady.



Obrázek 7.1: Závislost vstupu AC na požadovaném výstupu neuronové sítě.

Server	Od data	Počet titulků	Adresa
marketwatch	31.05.13 21:01	26353	feeds.marketwatch.com
fxstreet	31.05.13 14:00	49385	xml.fxstreet.com
dailyfx	31.05.13 12:20	11833	www.dailyfx.com
dailyforex	05.03.09 06:37	1573	www.dailyforex.com
ft	26.05.13 20:03	2221	www.ft.com
actionforex	31.05.13 17:10	37277	feeds.actionforex.com
forexfactory	31.05.13 20:20	22804	rss.forexfactory.net
forexlive	31.05.13 17:53	37026	www.forexlive.com
Celkem		188472	

Tabulka 7.3: Seznam serverů a počtu stažených titulků.

7.2 Fundamentální data

Použitá fundamentální data se dají rozdělit do dvou tříd. A to na titulky zpráv a makroekonomická vyhlášení nazývané také forexní kalendář.

7.2.1 Titulky zpráv

Textové data titulků jsem čerpal z webových serverů zabývajících se měnovým trhem. Tyto servery zprostředkovávají zprávy z oblasti trhu v reálném čase prostřednictvím protokolu RSS. Titulků bylo celkem staženo 188472.

Zdroj

Data byla stažena ze serverů uvedených v tabulce tabulce 7.3. Data zprostředkovala služba feedly.com². Celkový počet stažených titulků je 188472.

Struktura

Struktura těchto dat je dána jako časový údaj následovaný údajem textovým. Ukázku titulků můžeme vidět na příkladu v tabulce 4.1.

7.2.2 Forexní kalendář

Zdroj

Data forexních událostí byla stažena ze serveru dailyfx³ a bylo jich staženo 31048. Tento server umožňuje stažení csv strukturovaného souboru.

Struktura

Příklad stažených dat kalendáře můžeme vidět v tabulce 4.2. Každý řádek dat odpovídá jedné události, která je opatřena časovým razítkem a hodnotami vyhlášení.

²<https://feedly.com/>

³<http://www.dailyfx.com/>

7.2.3 Nejčastější slova

Z nejčastěji vyskytujících se slov byl vytvořen slovník. Prvních několik slov můžeme vidět na seznamu popisujícím počet výskytů a slovo 4.3. Počet slov slovníku byl stanoven na 500.

Všechny datové sady můžeme nalézt na příloženém CD.

Kapitola 8

Experimenty a výsledky

Experimenty jsem prováděl v rámci platformy Meta traderu. Na obrázku 8.1 můžeme vidět část průběhu experimentu. V horní části je svíčkový graf s proběhlými obchody. Modrý histogram znázorňuje výstupy neuronové sítě a fialový histogram požadované výstupy, které jsou známy až se zpožděním. Na tyto hodnoty je také trénována neuronová síť. Červené čáry označují práh, po jeho překročení se vstoupí do obchodu.

Následující grafy jsou závislosti kumulovaného zisku (osa y) tedy prostředků na virtuálním účtu vůči času (osa x).

První srovnání můžeme vidět na grafu 8.2, kde vidíme strategii založenou na náhodném obchodování. Pokud by strategie klesla pod tuto hranici, dá se o ní prohlásit, že je, zcela nefunkční.

8.1 Srovnání fixní sítě s adaptivním dotrénováním

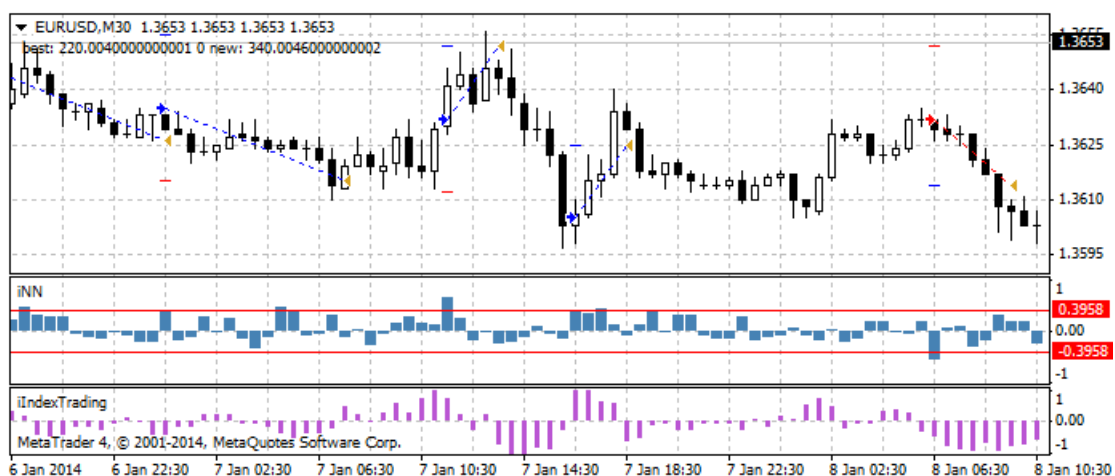
Na následujícím grafu 8.2 můžeme vidět srovnání po použití dotrénování neuronové sítě pomocí adaptivního trénování. Přibližně od 1.4. přestává být neuronová síť dobře natrénovaná, jak můžeme vidět na modré křivce pro technické vstupy a červené křivce pro vstupy technické i fundamentální. Proto je potřeba neuronovou síť dotrénovávat, aby byla vždy natrénována na aktuální změny na trhu.

Ovšem jak můžeme pozorovat, adaptivní trénování ne vždy pomáhá. Někde došlo ke zlepšení, například od data 1.5. se technické vstupy s adaptivním trénováním (zelená křivka) zlepšily. Ale někde se může zdát, že škodí původnímu relativně dobře natrénovanému stavu. Tuhle skutečnost přikládám špatnému propojení učícího algoritmu s obchodním modelem. Celková chyba trénovacích vzorků od požadovaného výstupu se snížila, ale očekávaný zisk tomu neodpovídá. Proto jsem se rozhodl přidat techniku ověření zisku experimentovanou v následující sekci 8.2.

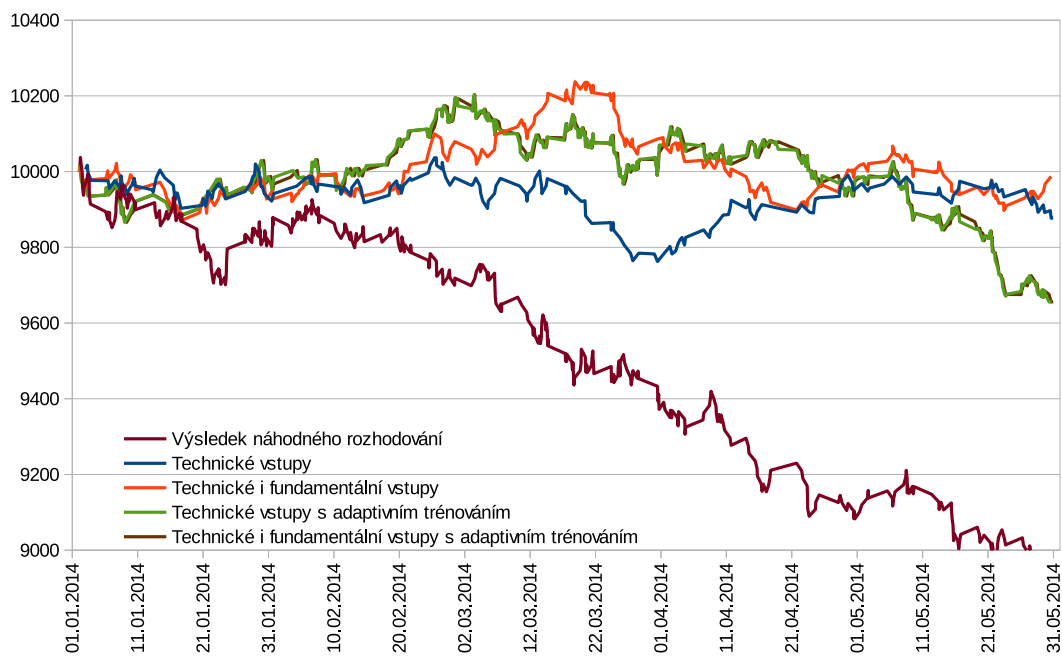
8.2 Ověření zisku nově natrénovaných vah

Jelikož adaptivní trénování ne vždy přispělo k zisku, rozhodl jsem se zpětně počítat zisk na minulém vývoji s nově natrénovanými váhami a výsledek porovnat se ziskem získaným pomocí původních vah. Tato technika (sekce 5.3) výrazně přispěla ke zlepšení výsledků, které můžeme vidět na grafu 8.3.

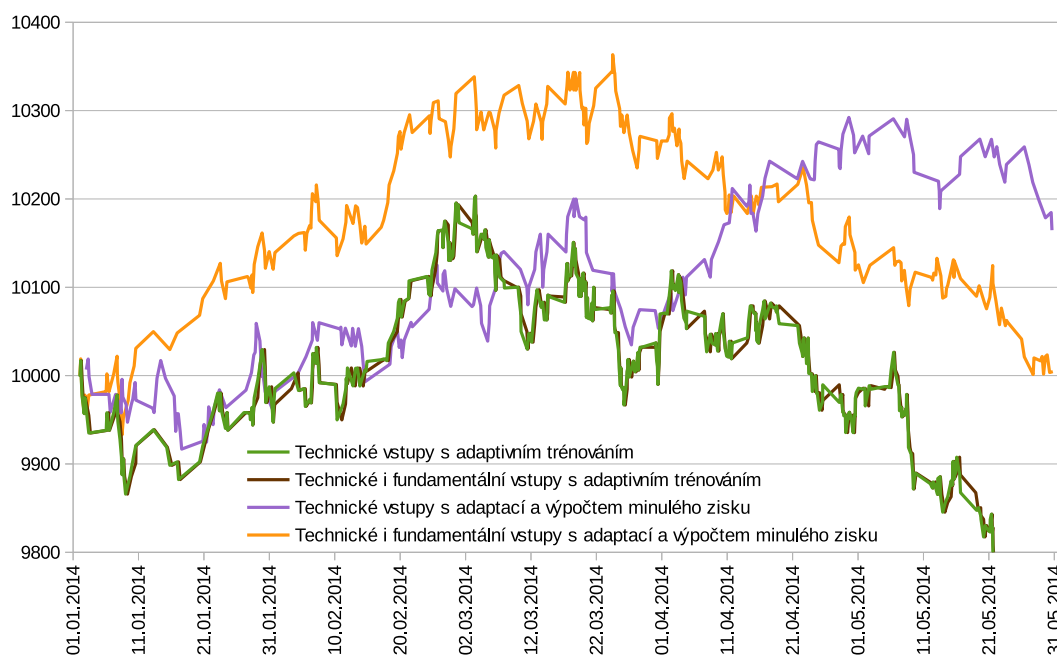
Oba testované vstupy, technické (fialová křivka) i kombinace technických a fundamentálních vstupů (oranžová křivka) s ověřováním zisku, téměř v celém testovaném období do-



Obrázek 8.1: Příklad experimentu.



Obrázek 8.2: Porovnání experimentů s adaptivním a bez adaptivního dotrénování.



Obrázek 8.3: Porovnání s ověřením výpočtu zisku na minulém vývoji vůči pouze adaptivnímu dotrénování.

sáhnou lepšího zisku než pouze adaptivní trénování bez zpětného ověření (zelené a hnědá křivka).

Dále také můžeme pozorovat, že kombinace vstupů dosáhla lepších výsledků až do data 11.4. než pouze technická analýza.

8.3 Násobná inicializace neuronové sítě

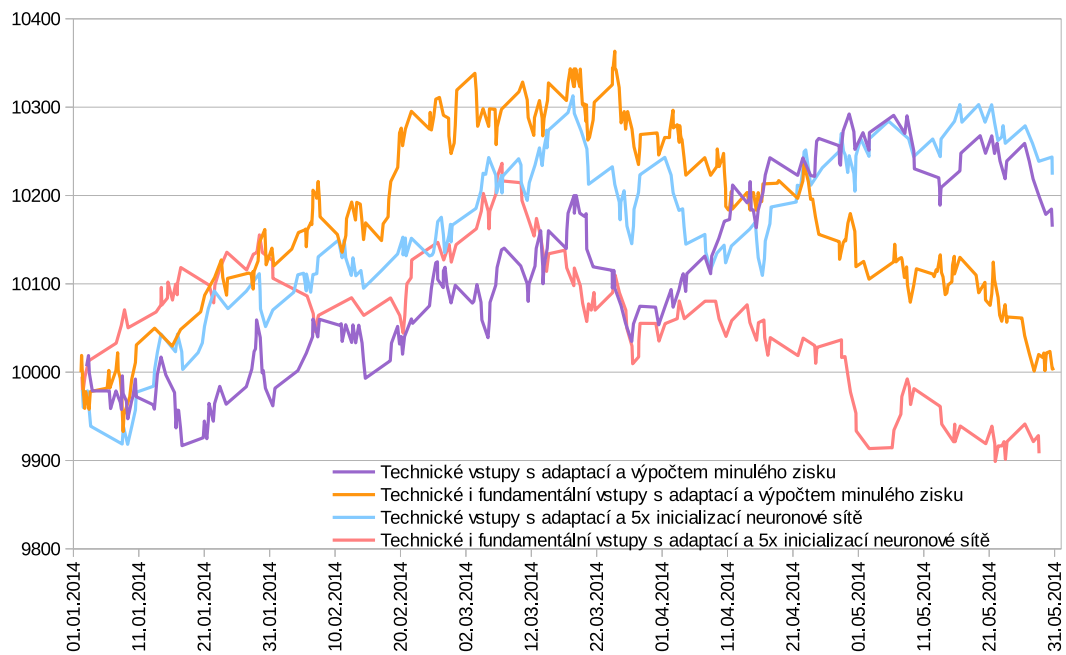
Na grafu 8.4 a 8.5 můžeme vidět porovnání pěti respektive desetinásobné inicializace neuronové sítě, jak je popsáno v sekci 5.3.1. U technických vstupů vyšla lépe pětínásobná a u kombinace vstupů vyšla lépe desetinásobná. Strategie s fundamentálními daty dosáhla výrazně vyššího zisku než technická. A byla více zisková až do data 20.4., kde technická data dokázala obrátit ztrátu a adaptovat se lépe na změnu v trhu.

8.4 Srovnání strategií

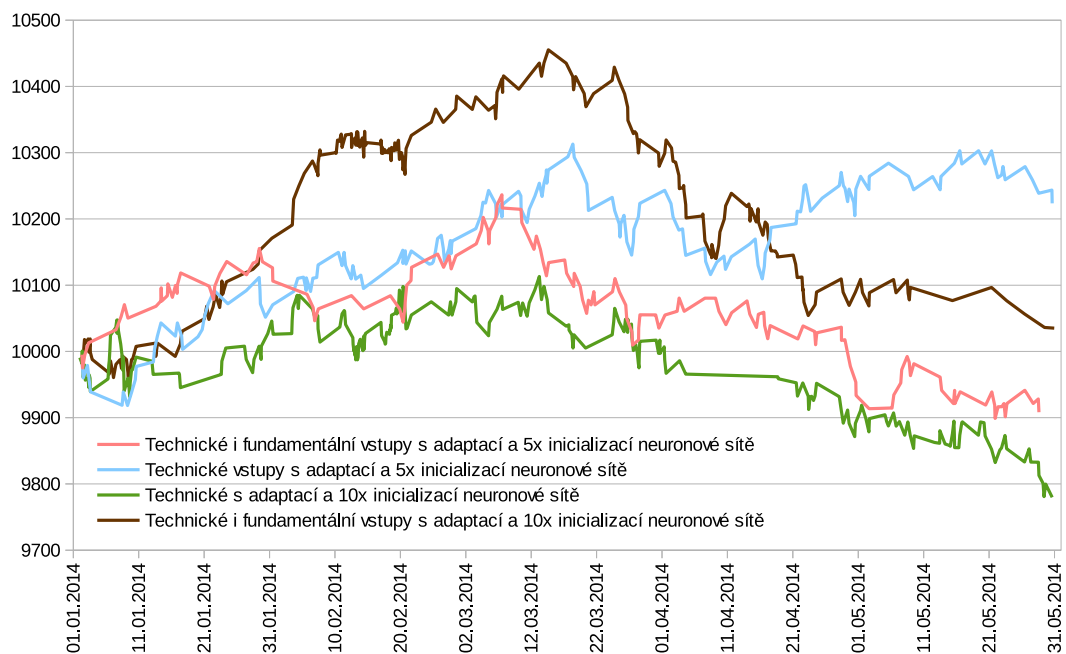
Na grafu 8.6 můžeme vidět porovnání různých strategií. Jako referenční strategií jsem použil svou strategií z bakalářské práce [41] založenou na rozpoznávání svíčkových formací s využitím lineárního klasifikátoru s trénováním pomocí gradient/descent s hodnotící funkcí, která simuluje testování strategie na předchozím čase.

Z předchozích experimentů jsem vybral experimenty, které dopadly nejlépe. A to pětínásobnou inicializaci neuronové sítě a desetinásobnou inicializaci neuronové sítě. První je založená pouze na technických datech a druhá je kombinací technických a fundamentálních dat.

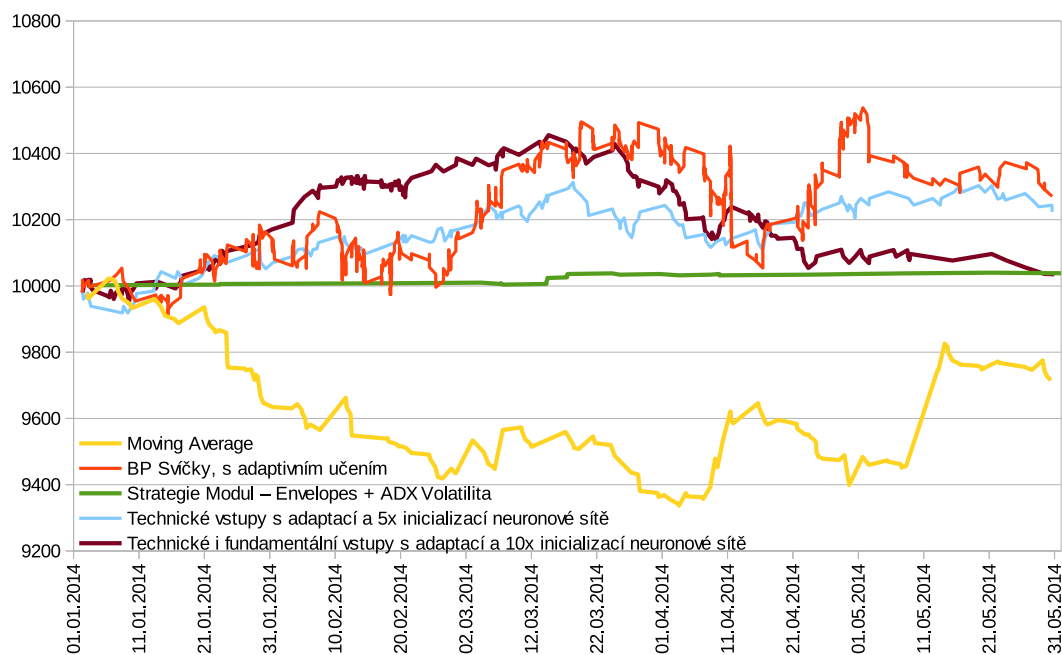
Další testovanou strategií je strategie Moving Average. Tato strategie je příkládána



Obrázek 8.4: Porovnání pětinasobné inicializace sítě, vůči pouhému ověření zisku nově natrénovaných vah.



Obrázek 8.5: Porovnání pětinasobné inicializace sítě vůči desetinásobné inicializaci.



Obrázek 8.6: Porovnání obchodních strategií.

k programu Metatrader a je založena na protnutí aktuální hodnoty kurzu s průměrnou hodnotou.

Poslední testovaná strategie je Modul - Envelopes + ADX Volatilita. Tato strategie je založená na překročení kurzu indikátor Envelopes [33].

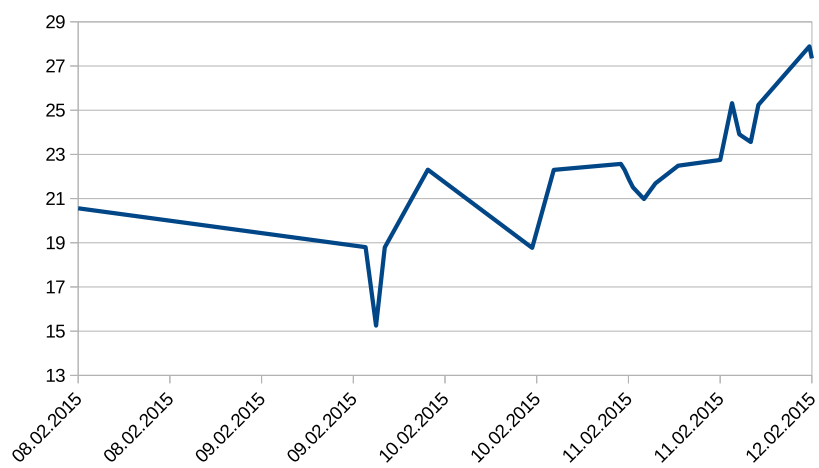
V porovnání můžeme vidět, že nejlépe si vedla strategie kombinací vstupů fundamentálních i technických a to až do data 20.3. kde téměř všechny strategie zaznamenaly poměrně velkou ztrátu. Tuto skutečnost příkládám nevhodnému pohybu na trhu.

Přibližně od 21.4. strategie založené pouze na technických datech nabírají znovu zisk.

Nejhůře si vedla strategie Moving Average, která je v celém testovaném období ztrátová. Strategie založená na indikátoru Envelopes je zisková, ale její zisk není velký ve srovnání s ostatními strategiemi.

8.5 Forward test na demo účtu

V krátkém časovém intervalu, jsem také testoval obchodní strategií na demo účtu společnosti Admiral Markets. Vstupy byly použity pouze technické, kvůli složitosti použití fundamentálních textových na aktuálních datech. V případě jejich využití by bylo zapotřebí vytvořit další program, který by tyto data stahoval v reálném čase a připravoval je pro strategii. Výsledek běhu můžeme vidět na grafu 8.7. Strategie po provedení pár ztrátových obchodů se dostala do zisku. Na konci testovaného období vykázala zisk přes 7 dolarů.



Obrázek 8.7: Forward test na demo účtu.

Kapitola 9

Závěr

9.1 Souhrn

Cílem této práce bylo vytvořit predikční obchodní systém. Na základě nastudované literatury jsem se rozhodl obchodovat pomocí svíčkových formací a dalších technických indikátorů. Tuto strategii jsem rozšířil o rozhodování pomocí textových dat. Predikci provádí neuronová síť jako regresivní funkce.

Mohu konstatovat, že strategie založená na neuronových sítích s danými vstupy dokáže do jisté míry předpovídat vývoj trhu. Ale protože vývoj ceny trhu už ze své podstaty obsahuje množství šumu a chaosu, není možné zaručit nebo garantovat zisk.

Z experimentů je patrné, že po většinu testovacího období fundamentální data přispěla ke zvýšení výnosnosti strategie. A také s těmito vstupy dosáhla největšího zisku, ale v určité chvíli si tato strategie nedokázala poradit se změnou trhu a zaznamenala propad.

Obchodní strategii jsem také spustil na demo účtu jako tzv. forward test. Tedy na nově získávaných datech, kde se strategie ukázala jako zisková.

9.2 Možné pokračování práce

Další rozšíření bych navrhl využití rekurentních neuronových sítí, které se přímo nabízí kvůli časové ose a redukci dimensionalit. Tato neuronová síť by na vstupu nepotřebovala historii několika posledních ukazatelů, ale pouze posledního. Díky tomu by se snížil počet vstupů a teoreticky i rychlost výpočtu.

Dále bych navrhl vyzkoušení jiného obchodního modelu založeného spíše na dlouhodobějších obchodech s rozhodováním ukončení obchodu. Delší držení otevřeného obchodu vede k větším ziskům, ale také zvyšuje riziko větší ztráty.

K zajímavému rozšíření by vedlo zpracování čísel z titulků. Například pomocí regulárních výrazů a následné přímé připojení těchto čísel na neuronovou síť. Čísla se ale v textech vyskytují v různých formátech a významech, a proto by bylo potřeba vymyslet roztržení do různých skupin. Také by bylo zapotřebí experimentovat co vkládat na neuronovou síť, pokud není žádné číslo v textu obsaženo.

Pro obchodování na demo účtu v reálném čase s využitím fundamentální analýzy by bylo vhodné vytvořit program, který by stahoval data automaticky. Tento program by musel propojovat funkčnost zde vytvořených skriptů pro stažení dat z internetu a programu zpracování dat pomocí vektoru výskytů slov.

Literatura

- [1] Acceleration/Deceleration Technical Indicator (AC).
http://ta.mql4.com/indicators/bills/acceleration_deceleration.
- [2] Automatické obchodní systémy. <http://www.xtb.cz/obchodni-systemy/metatrader/automaticke-obchodni-systemy>.
- [3] Commodity Channel Index (CCI).
http://stockcharts.com/school/doku.php?id=chart_school:technical_indicators:commodity_channel_index_cci.
- [4] *MetaTrader: Uživatelský manuál*.
- [5] MQL4 Documentation. <http://docs.mql4.com/>.
- [6] Měnové páry. <http://www.fxstreet.cz/3-cast-menove-pary.html>.
- [7] Obchodní systémy: MetaTrader verze 4.
<http://www.invest-centrum.cz/obchodni-systemy/>.
- [8] Relative Strength Index (RSI).
http://stockcharts.com/school/doku.php?id=chart_school:technical_indicators:relative_strength_index_rsi.
- [9] Slovník burzovních pojmů. <http://www.maklerska-zkouska.cz/slovník/>.
- [10] Strategie EUR/USD. <http://www.forexfriends.cz/4-59-126-pruvodce-.aspx>.
- [11] WELCOME. In *"MQL4 COURSE"*, 2005-10-19 [cit. 2013-04-25].
- [12] Create your first DLL for Metatrader using Microsoft Visual C++. 2012.
URL <http://www.mgaforex.com/2012/04/10/9/>
- [13] TikZ: Diagram of a perceptron. 2012.
URL <http://tex.stackexchange.com/questions/104334/tikz-diagram-of-a-perceptron>
- [14] Neural Networks and PyBrain. 2014.
URL http://nbviewer.ipython.org/github/cs109/2014/blob/master/labs/Lab11_Notes.ipynb
- [15] Al Shalabi, L.; Shaaban, Z.: Normalization as a Preprocessing Engine for Data Mining and the Approach of Preference Matrix. In *Dependability of Computer Systems, 2006. DepCos-RELCOMEX '06. International Conference on*, May 2006, s. 207–214, doi:10.1109/DEPCOS-RELCOMEX.2006.38.

- [16] Balahur, A.; Steinberger, R.; Goot, E. v. d.; aj.: Opinion Mining on Newspaper Quotations. In *Web Intelligence and Intelligent Agent Technologies, 2009. WI-IAT '09. IEEE/WIC/ACM International Joint Conferences on*, ročník 3, Sept 2009, s. 523–526, doi:10.1109/WI-IAT.2009.340.
- [17] Beneš, V.: *Burzy a burzovní obchody*. Informatorium, 1992, ISBN 80-85427-25-7.
- [18] Bill, H.; Soo, K. Y.: *A test of MACD trading strategy*. Diplomová práce, 2006.
- [19] BRYCHTA, J.: Fundamentální analýza. Složitá, ale nezbytná. <http://www.finance.cz/zpravy/finance/273472-fundamentalni-analyza-slozita-ale-nezbytna/>, 2010-07-26 [cit. 2013-02-21].
- [20] Černocký, J.: *Temporal processing for feature extraction in speech recognition, shortened version of habilitation thesis*. Edice Habilitační a inaugurační spisy, sv. 112, Publishing house of Brno University of Technology VUTUM, 2003, ISBN 80-214-2395-1, s. 1–30.
URL http://www.fit.vutbr.cz/research/view_pub.php?id=7240
- [21] Chatrath, A.; Miao, H.; Ramchander, S.; aj.: Currency jumps, cojumps and the role of macro news. *Journal of International Money and Finance*, ročník 40, č. 0, 2014: s. 42 – 62, ISSN 0261-5606, doi:<http://dx.doi.org/10.1016/j.jimonfin.2013.08.018>.
URL <http://www.sciencedirect.com/science/article/pii/S0261560613001204>
- [22] Chong, T. T.-L.; Li, C.; Yu, H. T.: Structural Change in the Stock Market Efficiency after the Millennium: The MACD Approach. *Economics Bulletin*, ročník 7, č. 12, 2008: s. 1–6.
- [23] Elder, A.: *Tradingem k bohatství*. Impossible, s.r.o., 1993, ISBN 80-239-7048-8.
- [24] Fraser, N.: The Biological Neuron.
<http://vv.carleton.ca/~neil/neural/neuron-a.html>, September 21, 1998.
- [25] Ghiassi, M.; Skinner, J.; Zimbra, D.: Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with Applications*, ročník 40, č. 16, 2013: s. 6266 – 6282, ISSN 0957-4174, doi:<http://dx.doi.org/10.1016/j.eswa.2013.05.057>.
URL <http://www.sciencedirect.com/science/article/pii/S0957417413003552>
- [26] Gurney, K.: *Neural networks for perceptual processing: from simulation tools to theories*. The Royal Society, 2007, ISBN 1471-2970, ISSN 0962-8436, s. 339–353, doi:10.1098/rstb.2006.1962.
- [27] Gutarra, E.: RMS and Backpropagation for Feedforward Neural Networks. 29 January 2011.
- [28] Hayes, A.: Basis Point (BPS).
<http://www.investopedia.com/terms/b/basispoint.asp>, 2015-03-15.
- [29] Horner, R.: *Forex tradingem k maximálním ziskům*. Computer press, a.s., 2008, ISBN 978-80-87071-87-8.

- [30] Izumi, Y.; Yamaguchi, T.; Mabu, S.; aj.: Trading Rules on the Stock Markets using Genetic Network Programming with Candlestick Chart. In *Proceedings of the 2006 IEEE Congress on Evolutionary Computation*, editace G. G. Yen; L. Wang; P. Bonissone; S. M. Lucas, Vancouver: IEEE Press, 6-21 July 2006, ISBN 0-7803-9487-9, s. 8531–8536.
- [31] Jílek, J.: *Finanční a komoditní deriváty v praxi*. GRADA Publishing, 2010, ISBN 978-80-247-3696-9.
- [32] Koza, R.: Objem obchodu jako indikátor.
http://www.traders.cz/ta_indikatory/objem_obchodu_jako_indikator.html, 2004-03-04 [cit. 2013-02-21].
- [33] Kučera, M.: *Modul systému MetaTrader pro automatické obchodování na finančních trzích*. Diplomová práce, 2012.
- [34] Lee, A.: *Klíče k pochopení cenných papírů*. Computer press, a.s., ISBN 80-85605-59-7.
- [35] Maks, I.; Vossen, P.: A lexicon model for deep sentiment analysis and opinion mining applications. *Decision Support Systems*, ročník 53, č. 4, 2012: s. 680 – 688, ISSN 0167-9236, doi:<http://dx.doi.org/10.1016/j.dss.2012.05.025>, 1) Computational Approaches to Subjectivity and Sentiment Analysis 2) Service Science in Information Systems Research : Special Issue on {PACIS} 2010.
 URL <http://www.sciencedirect.com/science/article/pii/S0167923612001364>
- [36] MarkPouglas: *Trading in the zone*. Impossible, s.r.o., 2010, ISBN 970-80-254-7066-4.
- [37] Martiny, K.: Unsupervised Discovery of Significant Candlestick Patterns for Forecasting Security Price Movements. In *KDIR*, editace A. L. N. Fred; J. Filipe; A. L. N. Fred; J. Filipe, SciTePress, 2012, ISBN 978-989-8565-29-7, s. 145–150.
 URL <http://dblp.uni-trier.de/db/conf/ic3k/kdir2012.html#Martiny12>
- [38] Meško, D.: *Normalizace dat pro neuronovou síť GAME*. Diplomová práce, České vysoké učení technické v Praze, Fakulta elektrotechnická, 2008.
- [39] Mikolov, T.: *Statistical Language Models Based on Neural Networks*. Dizertační práce, 2012.
 URL http://www.fit.vutbr.cz/research/view_pub.php.cs?id=10375
- [40] Mikolov, T.; Plchot, O.; Glembek, O.; aj.: PCA-based Feature Extraction for Phonotactic Language Recognition. In *Proc. Odyssey 2010 - The Speaker and Language Recognition Workshop*, International Speech Communication Association, 2010, ISBN 978-80-214-4114-9, s. 251–255.
 URL http://www.fit.vutbr.cz/research/view_pub.php?id=9317
- [41] Mikulenčák, R.: *Elektronické obchodování s cennými papíry*. 2013.
- [42] Musílek, P.: *Trhy cenných papírů*. Ekopress, s.r.o., 2011, ISBN 978-80-86929-70-5.
- [43] Nassirtoussi, A. K.; Aghabozorgi, S.; Wah, T. Y.; aj.: Text mining for market prediction: A systematic review. *Expert Systems with Applications*, ročník 41, č. 16, 2014: s. 7653 – 7670, ISSN 0957-4174, doi:<http://dx.doi.org/10.1016/j.eswa.2014.06.009>.
 URL <http://www.sciencedirect.com/science/article/pii/S0957417414003455>

- [44] Nassirtoussi, A. K.; Aghabozorgi, S.; Wah, T. Y.; aj.: Text mining of news-headlines for {FOREX} market prediction: A Multi-layer Dimension Reduction Algorithm with semantics and sentiment. 2015, ISSN 0957-4174, s. 306 – 324, doi:<http://dx.doi.org/10.1016/j.eswa.2014.08.004>.
URL <http://www.sciencedirect.com/science/article/pii/S0957417414004801>
- [45] Ondřej Hartman, L. T.: *První kroky na forexu*. Victoria Publishing, 2009, ISBN 978-80-251-2006-4.
- [46] Putna, L.: *Predikce vývoje kurzu pomocí umělých neuronových sítí*. Diplomová práce, Vysoké učení technické v Brně, FIT, 2011.
- [47] Rejnuš, O.: *Finanční Trhy*. Keypublishing, s.r.o., 2008, ISBN 978-80-87071-87-8.
- [48] Sekanina, L.: BIN Biologií inspirované počítače. (přednáška), 2014-04-21.
- [49] Shen, L.; Loh, H. T.: Applying rough sets to market timing decisions. *Decision Support Systems*, ročník 37, č. 4, 2004: s. 583 – 597, ISSN 0167-9236, doi:[http://dx.doi.org/10.1016/S0167-9236\(03\)00089-7](http://dx.doi.org/10.1016/S0167-9236(03)00089-7), data mining for financial decision making.
URL <http://www.sciencedirect.com/science/article/pii/S0167923603000897>
- [50] Smith, L. I.: A tutorial on principal components analysis. Technická zpráva, Cornell University, USA, February 26 2002.
URL http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf
- [51] Sovierzoski, M.; Argoud, F.; de Azevedo, F.: Evaluation of ANN Classifiers During Supervised Training with ROC Analysis and Cross Validation. In *BioMedical Engineering and Informatics, 2008. BMEI 2008. International Conference on*, ročník 1, May 2008, s. 274–278, doi:10.1109/BMEI.2008.251.
- [52] TUREK, L.: *Technická analýza*. Praha: Czechwealth, 2010.
- [53] Štýbr, D.: *Začínáme investovat a obchodovat na kapitálových trzích*. GRADA Publishing, 2011, ISBN 978-80-247-3648-8.
- [54] WWW stránky: Obchodní systémy. <http://www.xtb.cz/obchodni-systemy>.
- [55] WWW stránky: Čárové a sloupkové grafy. <https://www.highsky.cz/technicka-analyza/carove-a-sloupcove-grafy>.
- [56] WWW stránky: Candlestick - tajemství grafu "Made in Japan". http://www.financnik.cz/komodity/fin_obchod/candlestick-svickove-grafy.html, 2005-03-07 [cit. 2013-02-21].
- [57] WWW stránky: Kapitálové trhy. <http://www.grantcapital.cz/zacnete-investovat/kapitalove-trhy>, 2011 [cit. 2013-02-21].
- [58] Zbořil, F.: SCF Soft Computing. (přednáška), 2013.

Příloha A

Slovník pojmů

AC

Indikátor Acceleration/Deceleration Technical Indicator (AC) měří rozdíl mezi klouzavými průměry. Změna směru indikátoru značí změnu trendu trhu. Vypočítá se jako rozdíl klouzavých průměrů z období 5 a 34 historických hodnot střední hodnoty ceny (sekce 3.4) [1].

Aktivum

Souhrnně se takto označuje produkt, který je možné na burze koupit. Může se jednat o akcii, měnový pár, zlato atd. Cokoliv s čím se obchoduje [53].

Akcie

Akcie představuje podíl na majetku společnosti, která tento cenný papír vydala [34]. Držitel akcie má nejrůznější práva vlastníka: například právo podílet se na zisku společnosti formou dividendy, právo účastnit se valné hromady akcionářů a hlasovat na ní nebo se podílet na likvidačním zůstatku společnosti v případě její likvidace [57].

Broker

Broker je zprostředkovatel obchodů, který za určitý poplatek sjednává obchody na cizí účet. V případě burzy se jedná o obchodníka s cennými papíry (právníckou osobu), který má oprávnění obchodovat s investičními nástroji [57].

bts

Jednotka bazického bodu (bts). Používá se pro vyjádření hodnoty "stoploss" a "takeprofit" [28].

CCI

Indikátor Commodity Channel Indicator (CCI) měří odchylku ceny od průměrné hodnoty (sekce 3.3) [3].

Deriváty

Deriváty jsou transakce, jejichž podkladovým aktivem jsou běžně obchodované komodity (například kukuřice, zlato, atd.), měny, cenné papíry či indexy nebo úrokové míry, od kterých jsou tyto obchody odvozeny. Tyto obchody jsou uzavírány, na rozdíl od klasických

obchodů s odkladným účinkem vypořádání v daném termínu. Základním obchodovaným typem jsou futures a opce [57].

Futures kontrakty

Futures kontrakty představují závazek prodat či koupit určitá podkladová aktiva (např.: akcie, komodity, akciové indexy, měny atd.) ve sjednaném množství v předem určený čas a za předem stanovenou cenu [57]. Může se například jednat o výměru pevné částky hotovosti v jedné měně za doposud neznámou částku v měně jiné [31].

Likvidita

Na likvidním aktivu probíhá hodně obchodů. Není problém se vstupem ani výstupem z trhu. Když je likvidita nízká, může například jediný obchod pohnout s celým trhem o velký skok [53]. Obecně platí, že pokud měnové trhy mají okamžitou likviditu, není problém se vstupem ani výstupem z trhu [47].

Lot

Lot je jednotka minimálního obchodovatelného množství cenných papírů [9].

MACD

Moving average convergence/divergence Indicator (MACD) je jedním z nejpoužívanějších indikátorů v technické analýze. Jeho výpočet vyžaduje tři exponenciální klouzavé průměry (EMA), které jsou zpožděné indikátory pokračujícího nebo obracejícího se trendu [18, 22].

Opce

Opce je finanční nástroj z kategorie derivátů, nabyvateli dává právo, ne však povinnost, nakoupit nebo prodat určité základní nástroje za určitou cenu, v určité době. Vystavitel opce má povinnost prodat nebo koupit základní nástroj - podle dispozice nabyvatele opce [57].

Open interest

Termínem open interest se označuje celkový počet opcí a/nebo kontraktů futures, které v daný den ještě nejsou uzavřeny či doručeny. Počet zadaných nákupních objednávek typu market před otevřením trhu [52].

Objem obchodů

Objem obchodů je velikost celkového obchodovaného množství, který provází pohyb ceny trhu [32].

Perioda

Periodou je časový úsek, který svíčka nebo sloupec grafu zobrazuje [56].

RSI

Index relativní síly z anglického relative strength index (RSI), index měřící dynamiku vývoje ceny (sekce 3.2) [8].

Stop loss

Jedná se o úroveň maximální možné ztráty pro jeden určený obchod. Je vhodné ho nastavit u každého obchodu. Je možné ho nastavit statický, nebo dynamický [53].

Trend

Trend je obecné směřování pohybu trhu nebo ceny aktiva v určitém směru po delší dobu [52].

Take profit

Nastavuje se jako hodnota zisku z obchodu. Pokud cena dosáhne hodnoty, uzavře se obchod. Nevyužívá se příliš často v dlouhodobém obchodování, ale spíše je více využívám krátkodobými obchodníky [53].

Vstup a výstup z obchodu

Vstup do obchodu znamená nákup nebo prodej obchodního aktiva. Výstup znamená ukončení otevřeného obchodu [29].

Příloha B

Obsah CD a návod k použití

Příložené CD obsahuje tyto složky:

- AOS - Obchodní strategie
- BackPropDll - Zdrojové kódy obchodní strategie
- data - Stažená data
- doc - Zdrojové kódy LaTeXu diplomové zprávy
- experiments - Provedené experimenty
- MetaTrader - Instalace programu MetaTrader
- scripts - Skripty pro stažení a transformaci dat
- VectorExtract - Program pro čištění a zpracování textových dat

B.1 AOS

Na CD přiloženém k této diplomové práci se nachází zdrojové kódy predikčních systémů:

- EAclassifier - Obchodní systém založený na rozpoznávání svíčkových formací
- Random - Obchodní systém založený na náhodném rozhodování
- Moving Average - Demonstrační obchodní systém přikládáný k programu MetaTrader
- eNeuro2 - Obchodní systém s využitím neuronových sítí technické i fundamentální analýzy
- modul - Tato strategie je založená na překročení kurzu indikátor Envelopes.

Zdrojové soubory jsou uloženy ve složce *AOS* a jsou napsány jako strategie spustitelné v programu MetaTrader. Před spuštěním je nutné je zkompileovat v programu *metaeditor.exe* přikládáném k programu MetaEditor.

B.2 BackPropDll

Složka BackPropDll obsahuje knihovnu neuronové sítě. Otevřením souboru *BackPropDll.sln* ve visual studiu je možná úprava a kompilace. Knihovna je napsaná v jazyku C++.

B.3 Data

B.3.1 Technická data

Exportovaná technická data se nacházejí v adresáři *tech*. A jsou v textových souborech, kde jeden řádek má následující formát:

YYYY-MM-DD,HH:MM,Open,High,Low,Close,Volume

Například:

2011.08.24,13:00,1.4433,1.4476,1.4427,1.4448,1301

Řádky musí být u uvedeném formátu a chronologicky seřazeny, jinak je program MetaTrader neotevře. Data jsou stažena od společnosti Admiral Markets¹ a MetaQuotes².

B.3.2 Správa dat

Správa dat v programu MetaTrader se nachází v dialogovém oknu *Správce historie*. Je zde možné importovat a exportovat data v libovolném čase a aktivu. Je možné stahovat data od brokera automaticky rolováním grafu nebo od společnosti MetaQuotes³, která vyvíjí program MetaTrader. To můžeme přímo ve správě dat vybráním aktiva a kliknutím na tlačítko *Stáhnout*.

B.3.3 Fundamentální data

Stažená textová data jsou uložena ve složce *fun*. Soubory jsou pojmenovány podle serveru s příponou *.csv*.

Ve složce *transform* jsou uložena transformovaná data.

- *pcamatrix* - Kovariační matice vytvořená pomocí skriptu *dppca.m*
- *vector* - Soubor s vektory výskytů slov zpracovaný pomocí programu *VectorExtract*
- *vectortrain.matrix* - Soubor vektorů pro vstup do skriptu *dppca.m*
- *dictionary.txt* - Seznam nejčastějších slov

B.4 MetaTrader

Program MetaTrader je přiložen ve stejnojmenné složce a pro jeho instalaci budete potřebovat připojení na internet. Můžete nainstalovat buď *am4setup.exe* od společnosti Admiral Markets, který jsem využil já, nebo *mt4setup.exe* od výrobců programu MetaQuotes.

¹<http://www.admiralmarkets.cz/>

²<http://www.metaquotes.net/>

³<http://www.metaquotes.net/>

B.5 Scripts

Složka scripts obsahuje skripty:

- dppca.m - obsahuje script pro Matlab pro redukci dimensionalitu PCA
- calendar.php - script v jazyce php pro stažení dat forexního kalendáře
- feed.php - script v jazyce php pro stažení dat titulů ze zpravodajských serverů

B.6 VectorExtract

Složka obsahuje program pro úpravu a čištění textových dat. Jednotlivé funkce programu je možné přepínat podmocí konstanty *MODE*.

B.7 Instalace a spuštění

Pro vyzkoušení a spuštění strategie budeme muset nainstalovat MetaTrader. Můžeme si jej stáhnout s příloženého CD, nebo stáhnout z internetu například od společnosti Admiral Markets⁴.

Obsah složky AOS zkopírujeme do umístění pro zdrojové kódy meta traderu. Dané umístění je závislé na daném operačním systému. Obvykle bývá podobné:

```
c:\Users\user\AppData\Roaming\MetaQuotes\Terminal\  
96D85E71DE03515148B6C958482DB02B\MQL4\.
```

Daná složka již obsahuje zkompilevanou dll knihovnu, proto ji není třeba kompilovat ve Visual studiu.

Z příložených souborů vybereme strategii a poté tyto soubory otevřeme a zkompilujeme v programu *metaeditor.exe*, který je součástí instalace MetaTraderu [4].

Nyní již můžeme spustit MetaTrader (*terminal.exe*) a v testeru strategií vybrat zvolenou strategii, zvolit období, pro které se má simulace provádět.

Pokud chceme využít i fundamentální data, musíme je dodat do složky *Tester/Files* v adresáři MetaTraderu.

Před spuštěním simulace je ještě nutné si obstarat data. Můžeme importovat data použité pro testování, přes dialogové okno *správce historie*, taktéž příložené na CD, nebo využít data, které se vám stáhnou automaticky samy, pokud budeme přihlášení k účtu.

Simulaci spustíme tlačítkem *Začátek*.

⁴<http://www.admiralmarkets.cz/>