

Tento článek citujte takto:

Dana Nejedlová: Předpovídání indexu PX 50 neuronovou sítí. In: Liberecké informatické fórum, Technická univerzita v Liberci, listopad 2006, s. 176 – 183, ISBN 80-7372-121-X.

Dana Nejedlová

Technická univerzita v Liberci

Hospodářská fakulta

Katedra informatiky

Hálkova 6

461 17 Liberec

PŘEDPOVÍDÁNÍ INDEXU PX 50 NEURONOVOU SÍTÍ

1. Úvod

Burzovní indexy vytvářejí časové řady, jejichž vývoj je ovlivněn mnoha faktory. Některé z nich jsou náhodné (politický vývoj, počasí), některé mají jisté zákonitosti (psychologie davu na burze). Zákonitosti umožňují předpovídat budoucí hodnoty časové řady z předchozích hodnot. Celkové množství, složitost a vzájemná interakce náhodných a zákonitých faktorů však činí proces předpovídání velice složitým.

Z toho důvodu se pro předpovídání využívají statistické metody, které jsou schopny odstranit vliv náhodných jevů a postihnout významné zákonitosti. Příkladem takové statistické metody je regrese, při které se pomocí metody nejmenších čtverců hledají koeficienty modelu, například lineárního trendu, dat. Model s nalezenými koeficienty je potom možné extrapolovat do budoucnosti, čímž se předpoví budoucí hodnoty časové řady.

Tento článek popisuje předpovídání nové hodnoty časové řady z několika předchozích hodnot pomocí neuronové sítě a výsledek porovnává s předpovědí pomocí lineární regrese. Závěrem je zjištění, že neuronová síť dosahuje lepších odhadů.

2. Neuronová síť

2.1. Historie a využití

Možnost zpracování informací pomocí umělých neuronových sítí se ve vědě zkoumá od 40. let 20. století. Teprve od 80. let 20. století se však objevují praktické aplikace neuronových sítí.

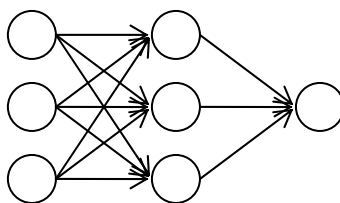
Vhodná pro neuronovou síť je úloha, ve které je třeba klasifikovat výskyty entit reprezentovatelné vektory podle složitých pravidel, z nichž některé si mohou i odporovat. Nejvýznamnější vlastností neuronových sítí je totiž schopnost generalizace neboli zobecnění, kdy se síť naučí jen ta pravidla, která statisticky v datech převládají.

Uplatnění neuronových sítí je nevhodné v úlohách, kdy je třeba, aby počítač podal člověku naprosto přesný výsledek. V takovém případě je vhodnější vytvořit systém explicitních (přímo člověkem pochopitelných) pravidel.

Systémy explicitních pravidel bývají vytvářeny v procesu znalostního inženýrství, kdy inženýr získává pravidla od experta. V oboru umělé inteligence existují také techniky pro získávání explicitních pravidel z dat, například rozhodovací stromy a Bayesovy sítě. Pravidla, které je schopna získat neuronová síť, jsou pravidla implicitní, což znamená, že jsou výsledkem interakce velkého množství vah v síti a vstupních dat a nedají se přepsat do explicitní formy. Neuronová síť tedy může správně klasifikovat aniž by výsledky zdůvodňovala.

2.2. Neuronová síť pro předpovídání indexu PX 50

Ve výzkumu popsaném v tomto článku je použita dopředná třívrstvá neuronová síť učící se pomocí pravidla backpropagation (zpětné šíření chyby). Zjednodušené schéma této sítě ukazuje obrázek č. 1.



Obr. 1. Model dopředné neuronové sítě

Pramen: Vlastní

Dopředná se zde popisovaná síť jmenuje proto, že v ní informace postupují ze vstupní levé vrstvy do výstupní pravé vrstvy, která poskytuje výsledek. Síť na ob-

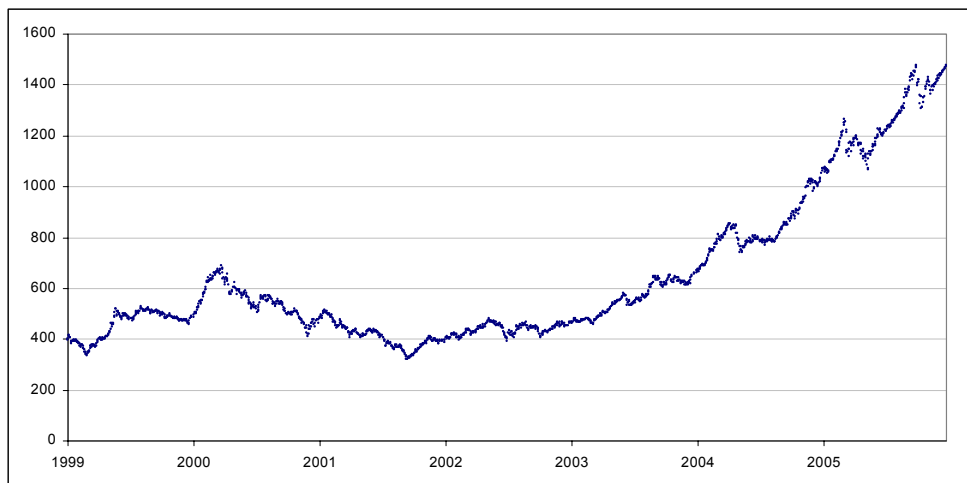
rázku č. 1 má 3 neurony ve vstupní vrstvě, 3 neurony v prostřední vrstvě a 1 neuron ve výstupní vrstvě. Síť použitá pro předpovídání indexu PX 50 má 10 neuronů ve vstupní vrstvě, 10 neuronů v prostřední vrstvě a 1 neuron ve výstupní vrstvě. Mezi sousedními vrstvami je každý neuron propojen s každým, jak také ukazuje obrázek č. 1.

Během svého učení síť čte informace o hodnotě indexu PX 50 z deseti předchozích dnů pomocí neuronů na vstupní vrstvě a hodnotu indexu PX 50 z jedenáctého dne na výstupní vrstvě. Pomocí algoritmu backpropagation dochází k postupné úpravě vah sítě tak, aby při zpracování deseti vstupních údajů došlo k co nejmenší odchylce předpovědi od hodnoty jedenáctého dne.

3. Data a jejich reprezentace

Index PX 50 byl do 20. 3. 2006 oficiálním indexem Burzy cenných papírů Praha, a.s., kdy byl spojitě nahrazen indexem PX. Způsob výpočtu indexu PX je v dokumentu [1]. Hodnoty časové řady indexu PX 50 za roky 1999 až 2005 byly vzaty z ročenek Burzy cenných papírů Praha, které jsou volně dostupné na internetové adrese [2].

Vývoj indexu PX 50 od roku 1999 do roku 2005 je znázorněn na obrázku č. 2.



Obr. 2. Vývoj indexu PX 50 v závěrečných hodnotách každého burzovního dne

Pramen: Vlastní

Data pro neuronovou síť byla vytvořena ze 1759 hodnot dostupné časové řady. Vzniklo 1749 vzorů pro neuronovou síť. To znamená, že prvních 11 hodnot bylo použito pro 1. vzor a ostatní vzory vznikly posunutím čteného úseku řady o jeden

den. 1505 vzorů popisujících období 1999 až konec roku 2004 bylo použito pro trénování sítě. Zbylých 244 vzorů za rok 2005 bylo použito pro testovací data. Cílem výzkumu je vytrénovat síť na letech 1999 až 2004 tak, aby byla schopna co nejlepších předpovědí jedenáctých dnů podle deseti předchozích dnů v roce 2005.

Všechny neurony v síti vysílají signály v rozsahu hodnot od 0 do 1. Z toho důvodu bylo prvních 10 hodnot každého vzoru znormalizováno podle vzorce (1)

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}, \quad (1)$$

kde x_{min} a x_{max} je minimální a maximální hodnota v deseti následujících dnech, podle kterých se předpovídá jedenáctý den.

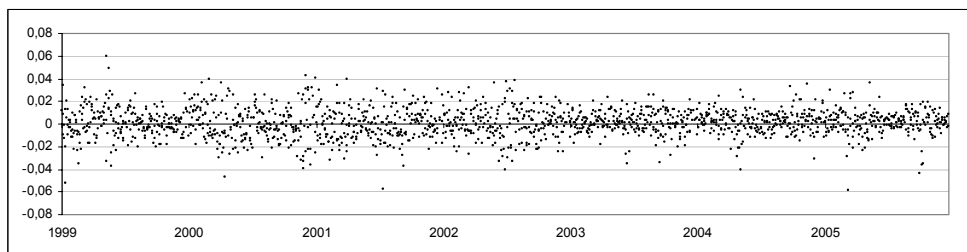
Jedenáctá hodnota byla znormalizována podle vzorce (2)

$$x_{norm} = \frac{I + 0,06}{0,12}, \quad (2)$$

kde I je mezidenní index růstu indexu PX 50 vypočtený podle vzorce (3).

$$I = \frac{x_{11} - x_{10}}{x_{10}} \quad (3)$$

Analýzou dostupných dat (viz obrázek č. 3) bylo zjištěno, že index I se pohybuje v rozmezí hodnot $\pm 0,06$, což znamená, že výsledek vzorce (2) bude v rozmezí od 0 do 1. Je možné se domnívat, že rovnoměrnost rozložení hodnot I na obrázku č. 3 je důsledkem pravidel obchodování na burze týkajících se nepřekročitelnosti povoleného rozpětí akciových kurzů během obchodů v daném dni.



Obr. 3. Mezidenní index růstu indexu PX 50 vypočtený podle vzorce (3)

Pramen: Vlastní

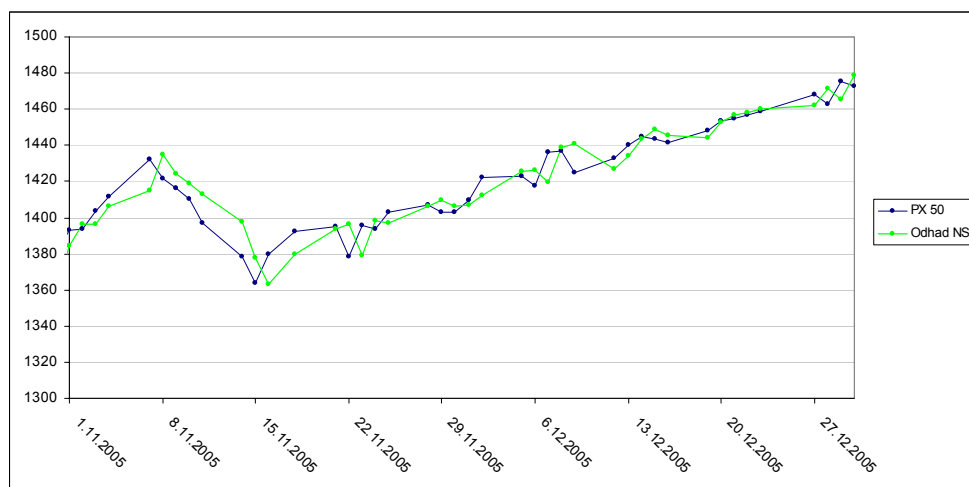
4. Zpracování a výsledky

Pro zpracování byla použita neuronová síť získaná vlastním vývojem v programovacím jazyce Turbo Pascal. Bylo provedeno 30 pokusů s učením neuronové sítě z počátečních náhodně stanovených vah. Výsledkem těchto pokusů

bylo zjištění, že úspěšnost tréninku sítě je především závislá na počátečních náhodných vahách. Například někdy došla síť k uspokojivému výsledku po 80 cyklech a někdy jí k tomu nestačilo ani 1000 cyklů. Hodnota počátečních vah byla důležitější než drobné modifikace ve způsobu učení sítě (trénování či netrénování strmosti sigmoidální funkce v neuronech, změny vah nejprve sečíst za celý cyklus a potom teprve přičíst k vahám či přičítat po zpracování jednotlivých vzorů).

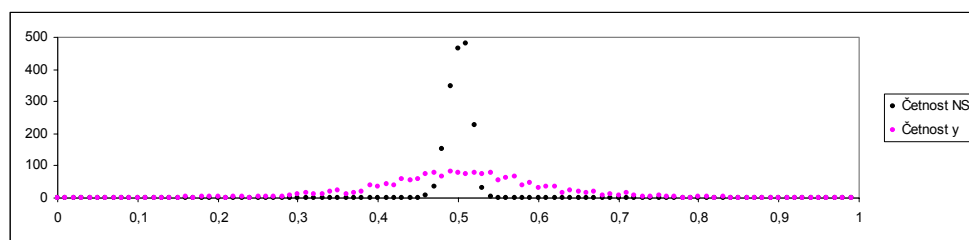
V každém cyklu síť přečetla všechny trénovací vzory a upravila si podle nich váhy a dále přečetla všechny testovací vzory a vypočetla celkovou chybu na nich. Celková chyba na trénovacích vzorech v průběhu učení postupně klesá. Celková chyba na testovacích datech také nejdříve klesá. Trénování sítě bylo ukončeno, když celková chyba na testovacích datech začala stoupat. To totiž znamená, že se síť začíná učit zákonitosti specifické pouze pro trénovací data, zatímco před tím se učila zákonitosti společné pro trénovací i testovací data. Zbytek této kapitoly ukazuje výsledky nejúspěšnějšího pokusu.

Obrázek č. 4 ukazuje detail předpovědí neuronové sítě v porovnání se skutečnými hodnotami v posledních dvou měsících testovacích dat. Je z něj vidět, že předpovídané hodnoty jsou velice blízké hodnotám předchozího dne. To je důsledkem toho, že hodnoty vycházející z výstupního neuronu sítě jsou velice blízké střední hodnotě 0,5, jak ukazuje obrázek č. 5. Cesta od hodnot na vodorovné ose obrázku č. 5 k hodnotám na svislé ose obrázku č. 4 je přes aplikaci vzorečků (2) a (3).



Obr. 4. Detail předpovědí neuronové sítě v porovnání se skutečnými hodnotami indexu PX 50

Pramen: Vlastní



Obr. 5. Histogram výstupních signálů výstupního neuronu (NS) a skutečných dat (y)

Pramen: Vlastní

Zjištění, že předpovědi neuronové sítě téměř kopírují předchozí den, vedlo k porovnání výkonu sítě s odhadováním podle předchozího dne. Nejdůležitější výsledky výzkumu shrnuje tabulka č. 1.

Tabulka 1. Průměrná odchylka skutečných a předpověděných hodnot indexu PX 50 na jeden vzor

Odhad / Data	Trénovací	Testovací	Chyba celkem
Neuronová síť	6,821	13,598	8,113
Předchozí den	6,849	13,832	8,188
Lineární trend	9,747	20,952	11,958

Pramen: Vlastní

Tabulka č. 1 ukazuje, že neuronová síť byla přesnější než odhad říkající, že 11. den bude stejný jako 10. den a tento jednoduchý odhad je zase mnohem přesnější než extrapolace lineárním trendem na 11. den podle 10 předchozích dnů. Hodnoty v tabulce č. 1 jsou ve stejném měřítku a jednotkách, jako hodnoty na svislé ose obrázku č. 2 a č. 4.

Ze statistického hlediska je neuronová síť parametrický model. Jeho významnou vlastností je počet parametrů. To je počet všech vah sítě, které se při jejím tréninku přepočítávají. Síť, pomocí níž byly získány nejlepší výsledky z tabulky č. 1, měla 10 krát 10 vah mezi 1. a 2. vrstvou a 10 vah mezi 2. a 3. vrstvou + 11 prahových hodnot (dodatečné váhy) pro neurony ve 2. a 3. vrstvě, tj. 121 parametrů.

5. Závěr

Výzkum popsany v tomto článku ukázal, že neuronová síť může být užitečným nástrojem pro predikci budoucí hodnoty v časové řadě burzovního indexu. Její generalizační schopnosti ukázaly, že dobrou metodou odhadu příští hodnoty je prosté použití poslední známé hodnoty. Oproti tomuto odhadu je však neuronová

síť na použitých testovacích datech přesnější. Predikci pomocí extrapolace lineárního trendu také nebylo dosaženo lepších výsledků.

Další výzkum se zaměří na tabelování závislosti výkonu sítě na délce časové řady, podle které předpovídá následující hodnotu. Výsledkem bude zjištění, jak dlouho trvá vliv jedné hodnoty na budoucí vývoj časové řady.

Podobnost charakteru zde zkoumaného indexu PX 50 s jinými časovými řadami z finančních trhů dovoluje předpokládat možnost bezproblémové aplikace neuronových sítí všude tam, kde je třeba provádět velké množství rychlých odhadů na základě předchozího vývoje finančního ukazatele. Srovnání odhadu neuronové sítě a odhadu podle předchozího dne na testovacích datech v tabulce č. 1 ukazuje, že pokud by existovala situace, kdy by zlepšení odhadu indexu o 0,2 přinášelo zisk, neuronová síť by mohla pomoci. Z toho také plyne oprávněnost domněnky, že ve vývoji zkoumané časové řady opravdu existují zákonitosti, kdy předchozích 10 dnů ovlivňuje její další vývoj.

Autorka by chtěla touto cestou poděkovat studentu Janu Kašparovi z Pedagogické fakulty Technické univerzity v Liberci za nalezení zdroje [2].

Literatura

- [1] http://ftp.pse.cz/Info.bas/Cz/PX_manual_060222.pdf
- [2] <http://ftp.pse.cz/Statist.dta/Year/>

SUMÁŘ

Příspěvek popisuje aplikaci neuronové sítě na předpovídání burzovního indexu PX 50. Zkoumanou časovou řadu tvořilo 1759 hodnot indexu PX 50 za roky 1999 až 2005. Každá hodnota reprezentovala jeden burzovní den. Úkolem bylo předpovědět 11. hodnotu podle 10 předchozích. Celá časová řada byla rozdělena na trénovací a testovací data. Neuronová síť se nejdříve naučila předpovídat řadu pro roky 1999 až 2004 (trénovací data) a potom byla aplikována na řadu z roku 2005 (testovací data). Zde dosahovala lepších výsledků než odhad pomocí lineárního trendu a odhad pomocí poslední známé hodnoty.

SUMMARY

This contribution describes the application of neural network to the prediction of the PX 50 stock exchange indices. The studied time series consisted of 1759 values of the PX 50 indices for the years from 1999 to 2005. Each value represented a single exchange day. The task was to predict the 11th value according to 10 preceding values. The whole time series was divided into training data and test data. The neural network has learned to predict time series of the years from 1999 to 2004 (training

8

data) and then it was applied to the time series of the year 2005 (test data). The network has reached better results than prediction by means of linear trend and by means of the last known value on the test time series.