

KIV/VSS

Skrytý markovský model pro predikci cen akcií

Martin Forejt - A20N0079P
mforejt@students.zcu.cz

23. 1. 2021

Obsah

1	Zadání	2
1.1	Skrytý markovský model pro predikci cen akcií	2
2	Úvod	3
3	Skrytý markovský model	4
3.1	Definice	4
3.2	Problémy a jejich řešení	4
4	Existující nástroje	6
4.1	hmmlearn	6
4.2	pomegranate	6
4.3	PyHHMM	6
5	Predikce cen akcií	7
6	Implementace	8
7	Výsledky	9
8	Uživatelská příručka	10
8.1	Požadavky	10
8.2	Spuštění ze zdrojových kódů	10
8.3	Spuštění z wheel balíčku	10
8.4	Parametry programu	10
9	Závěr	12

1 Zadání

Cílem je podívat se na stávající literaturu a vyprat si pro vás zajímavý simulační problém, který zkusíte reimplementovat. V minulých letech se objevila řada věcí - celulární automat pro modelování šíření požáru, vzniku mraků nebo pohybu kapaliny v terénu, predikce cen akcií postavené na markovských modelech, simulace různého chování hmyzu a podobně. Není nutné přijít s vlastním, originálním řešením, může jít (a očekává se že půjde) o replikační studii, moc se jich nedělá.

Součástí zadání pak bude přehled state-of-the-art - co v dané oblasti už existuje, jaké nástroje je možné využít a podobně.

1.1 Skrytý markovský model pro predikci cen akcií

Cílem semestrální práce by měla být analýza odborné literatury na téma predikce cen akcií za pomoci skrytého markovského modelu (dále HMM) a jeho následná implementace.

Práce bude vycházet ze článku Hidden Markov Model for Stock Trading (Nguyen N., 2018) [1], který představuje aplikaci HMM pro obchodování s akcemi (jako příklad je použit index S&P 500) na základě jejich predikce. Autor začíná použitím 4 kritérií pro odhad predikční chyby, aby určil optimální počet stavů pro HMM, vybraný čtyřstavový HMM implementovaný dle [2] je použit k predikci měsíčních uzavíracích cen indexu S&P 500.

Výsledkem práce by měla implementace modelu v podobě desktopové aplikace (v pythonu), která dle historických dat (ať už poskytnuta uživatelem, či získána automaticky např. z finance.yahoo.com) natrénuje model a následně zobrazí v grafu jak trénovací data, tak predikovaná data.

V první části by měl mít implementovaný HMM fixní počet stavů, následně by mohl být rozšířen tak, aby dle kritérií pro odhad predikční chyby vybral vhodný počet stavů, který se může lišit dle konkrétního akciového titulu.

2 Úvod

Cílem obchodníků s akciemi je nakoupit akcie za nízkou cenu a prodat je za cenu vyšší. Nicméně, kdy je nejlepší čas na nákup nebo prodej akcií je náročná otázka. Investice do akcií tak mohou mít obrovský výnos nebo výrazné ztráty v důsledku vysoké volatility cen akcií.

Dle [2] bylo v poslední době publikováno mnoho výzkumů pro nalezení optimálního modelu pro predikci chování akciových trhů. Většina výzkumů využívala techniky jako autoregression moving area (ARMA) nebo multiple regression modely. Objevují se také techniky založené na umělé inteligenci využívající neuronové sítě, fuzzy logiku a další.

V této práci bude použit skrytý markovský model (dále HMM), který pro předpověď ceny akcií použili Hassan a Nath [2] nebo Nguyet Nguyen [1] a naše práce bude z těchto dvou vycházet.

Hassan a Nath ve své práci [2] použili HMM pro předpověď ceny akcie na konci následujícího dne pro několik akcií aerolinek. Nguyet Nguyen svou prací [1] na tuto práci navazuje a přidává použití kritérií pro odhad predikční chyby, podle kterých vybírá počet stavů HMM.

3 Skrytý markovský model

Skrytý markovský model (HMM) je konečný automat s fixním počtem stavů, představen v 70. letech jako nástroj pro rozpoznávání řeči. Tento model založený na statistických metodách najde mnoho aplikací jako je rozpoznávání řeči, analýza sekvenční DNA, rozpoznání ručně psaného textu a další.

3.1 Definice

HMM je charakterizován:

- N - počet stavů modelu
- M - počet symbolů na stav
- T - délka pozorované sekvenční
- O - pozorovaná sekvenční O_1, O_2, \dots, O_T
- $A = \{a_{ij}\}$ - pravděpodobnost přechodu ze stavu i do stavu j
- $B = \{b_j(O_t)\}$ - pravděpodobnost pozorování O_t ve stavu j
- $\pi = \{\pi_i\}$ - pravděpodobnost bytí ve stavu i v čase $t=1$
- $\lambda = \{A, B, \pi\}$ - HMM model

Pokud jsou pravděpodobnosti spojitě rozděleny, máme spojitý HMM. V této práci předpokládáme, že pravděpodobnost pozorování je Normální (Gaussovo) rozdělení, potom $b_i(O_t) = N(O_t = \mu_i, \sigma_i)$, kde μ_i a σ_i jsou střední hodnota a rozptyl rozdělení pro stav S_i , potom jsou parametry HMM definovány jako:

$$\lambda = \{A, \mu, \sigma, \pi\}$$

3.2 Problémy a jejich řešení

Najdeme tři problémy při řešení problémů pomocí HMM:

1. Pro pozorování $O = \{O_t, t = 1, 2, \dots, T\}$ a model $\lambda = \{A, \mu, \sigma, \pi\}$ najít pravděpodobnost pozorování $P(O|\lambda)$
2. Pro pozorování $O = \{O_t, t = 1, 2, \dots, T\}$ a model $\lambda = \{A, \mu, \sigma, \pi\}$ najít sekvenci stavů, která nejlépe popisuje pozorování

3. Pro pozorování $O = \{O_t, t = 1, 2, \dots, T\}$ zkalibrovat parametry modelu $\lambda = \{A, \mu, \sigma, \pi\}$
1. Pravděpodobnost pozorování najdeme pomocí algoritmu Forward nebo Backward
2. Sekvenci stavů najdeme pomocí algoritmu Viterbi
3. Model zkalibrujeme pomocí algoritmu Baum-Welch

V naší práci využijeme algoritmus Forward pro řešení prvního problému a algoritmus Baum-Welch pro řešení třetího problému (tento algoritmus využívá Forward a Backward). Druhý problém řešit nepotřebujeme.

4 Existující nástroje

Pro použití HMM v pythonu (ve kterém bude aplikace implementována) existuje několik nástrojů, zde následuje seznam vybraných:

- `hmmlearn`¹
- `pomegranate`²
- `PyHHMM`³

4.1 `hmmlearn`

Knihovna pro práci s HMM v pythonu, prvotní verze programu byly testovány s touto knihovnou, ale celkem často padala.

4.2 `pomegranate`

Knihovna pro práci s HMM pythonu, narozdíl od `hmmlearn` fungovala spolehlivě a byla použita v prvotních fázích vývoje. Její integrace v programu zůstala a lze ji použít na místo našeho implementovaného modelu (viz uživatelská příručka). Je napsána v cythonu a obsahuje paralelizace, proto je mnohem výkonnější než naše implementace.

4.3 `PyHHMM`

Další knihovna pro práci s HMM v pythonu, s jednoduchou a transparentní implementací, některé její části byly použity i v naší implementaci.

¹<https://github.com/hmmlearn/hmmlearn>

²<https://github.com/jmschrei/pomegranate>

³<https://github.com/fmorenopino/HeterogeneousHMM>

5 Predikce cen akcií

TODO info jak pomoci HMM predikovat ceny

6 Implementace

TODO info o implementaci

7 Výsledky

TODO info o výsledcích, porovnání s články, porovnání s knihovnama

8 Uživatelská příručka

8.1 Požadavky

Pro spuštění aplikace je potřeba mít nainstalovaný python 3.

8.2 Spuštění ze zdrojových kódů

Projekt využívá pro správu závislostí nástroj Poetry, který je možné nainstalovat pomocí příkazu:

```
pip install poetry
```

Následně pomocí poetry nainstalujeme všechny závislosti definované v souboru *pyproject.toml* pomocí příkazu:

```
poetry install
```

Vlatní aplikaci můžeme spustit pomocí příkazu (parametry viz níže):

```
poetry run start
```

8.3 Spuštění z wheel balíčku

K dispozici je také *.wheel* soubor *hmm_stock_forecast-1.0-py3-none-any.whl*, který můžeme nainstalovat i bez Poetry pomocí příkazu:

```
python -m pip install hmm_stock_forecast-1.0-py3-none-any.whl
```

A následně spustit (parametry viz níže):

```
python -m hmm_stock_forecast.main
```

8.4 Parametry programu

Aplikace očekává následující parametry:

- -t ticker akciového titulu (např. AAPL)
- -f cesta k csv souboru (4 sloupce: Open, Low, High, Close)
- -s datum počátku dat, která se mají stáhnout
- -e datum konce dat, která se mají stáhnout
- -w velikost okna viz výše (defaultně 120)

- -m typ modelu:
 - HMM - naše implementace modelu (defaultně)
 - pomegranate - využití modelu knihovny pomegranate

Spuštění je tedy možné buď s -f nebo -t, v tom případě jsou povinné i datумы pro stažení dat z yahoo finance.

9 Závěr

TODO

Reference

- [1] Nguyen N. *Hidden Markov Model for Stock Trading*. International Journal of Financial Studies. 2018; <https://doi.org/10.3390/ijfs6020036>.
- [2] M. R. Hassan and B. Nath, *Stock market forecasting using hidden Markov model: a new approach* 5th International Conference on Intelligent Systems Design and Applications (ISDA'05), 2005, pp. 192-196, doi: 10.1109/ISDA.2005.85.