

# Modelovanie časových radov z monitorovacieho systému Hawkular

Bc. Pavol Loffay<sup>1</sup>

23. novembra 2015

**Abstrakt:** Práca spracováva predikciu časových radov prevzatých z monitorovacieho systému Hawkular<sup>a</sup>. Tento systém dokáže monitorovať Java aplikácie, alebo fyzický hardvér na ktorom je spustený. Z množiny pozorovaných metrík boli vybraté nasledujúce: vyťaženie Java hromady (heap), miesto na disku a počet voľných databázových spojení. Tieto časové rady som analyzoval a cieľom bolo zostaviť model, ktorý najlepšie popisoval priebeh danej časovej rady. V práci som postupoval podľa Box – Jenkinsovej metodológie.

**Kľúčové slová:** Časová rada; ARIMA; Hawkular; ACF

**JEL klasifikácia:** C53

---

<sup>a</sup>Dostupné na <http://www.hawkular.org>

## 1 Úvod

Monitorovanie dôležitých business aplikácií, ako sú napríklad bankové systémy alebo rôzne servery na ktorých bežia služby ktoré sú využívané 24/7 je veľmi dôležité. Väčšina monitorovacích systémov dokáže upozorniť administrátora na vyťaženie pri prekročení hraničnej hodnoty. V monitorovacom systéme Hawkular je možné zapnúť predikciu, takže administrátor bude upozornený vopred ak by náhodou malo dôjsť k prekročeniu spomenutej hraničnej hodnoty. Používané modely časových rád v systéme Hawkular sú varianty exponenciálneho vyrovnania. Modely ARIMA nemohli byť použité z dôvodu nestacionarity dát a náročnosti na výpočet – systém je schopný analyzovať aj niekoľko tisíc model naraz.

V tejto práci sa zameriam na konštrukciu optimálneho ARIMA modelu, ktorý následne porovnam s exponenciálnym vyrovnaním konkrétne Holtovou metódou s lineárnym trendom.

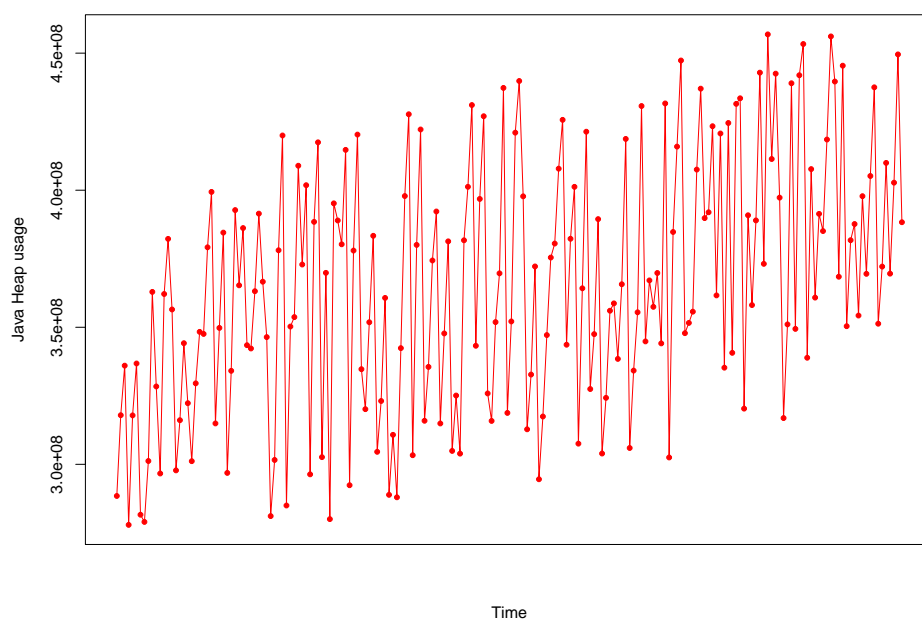
## 2 Analýza Java Heap metriky

V tejto kapitole budeme analyzovať časovú radu, ktorá popisuje vyťaženosť Java heap-u v čase.

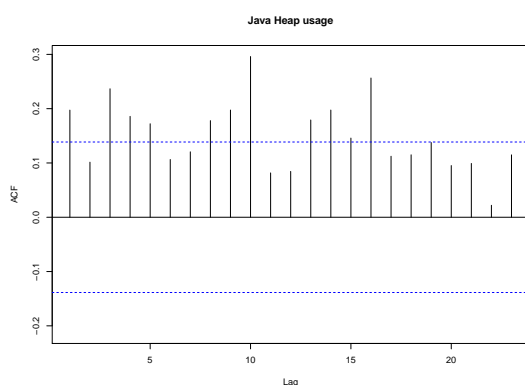
Graf časovej rady:

---

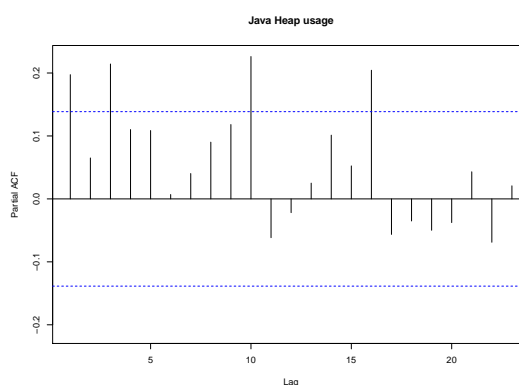
<sup>1</sup>Masarykova univerzita, Fakulta informatiky, obor: Service Science Management Engineering, p.loffay@mail.muni.cz



Obr. 1: Vyťaženosť Java Heap-u v čase.



(a) Autokorelačná funkcia.



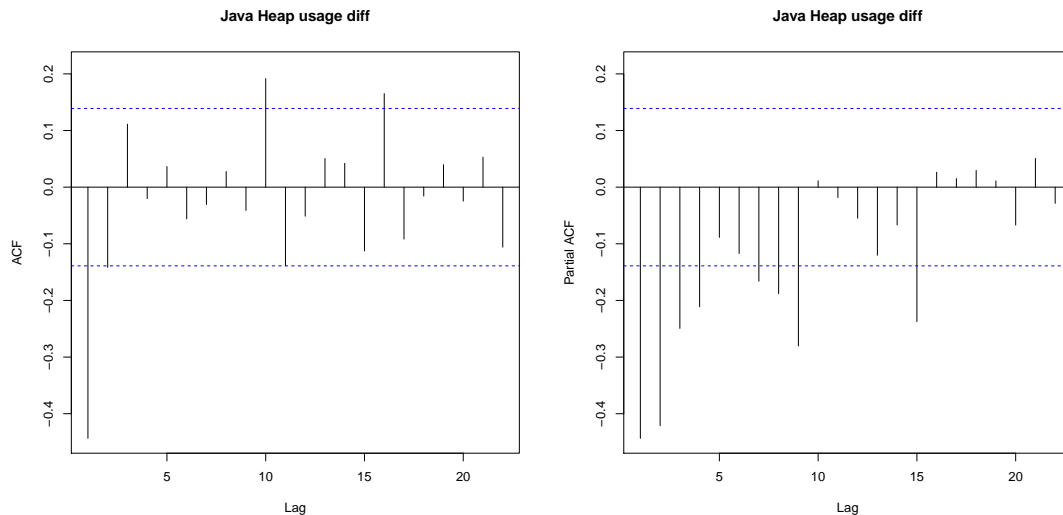
(b) Parciálna autokorelačná funkcia.

## 2.1 Identifikácia ARIMA modelu

Z obrázka 1 je možné usúdiť, že časová rada obsahuje mierny lineárny trend, čo by znamenalo že je nestacionárna. Následne si vykresíme autokorelačnú (zkrátene) a parciálnu autokorelačnú funkciu (zkrátene PACF).

Z obrázka 2a je vidieť že hodnoty ACF sú kladné ale avšak relatívne blízke nule, stým že významne neklesajú k nule. V grafe ACF 2a taktiež nie sú prítomné prítomné periodicky posunuté vysoké hodnoty, takže rada neobsahuje sezónnosť.

Keďže hodnoty ACF so zvyšujúcim spozdením neklesali k nule rozhodli sme sa urobiť ADF test na test stacionarity.



Obr. 3: ACF a PACF diferencovanej časovej rady.

```
> adfTest(as.numeric(b$avg), lags = 0, type="nc")
Title:
Augmented Dickey-Fuller Test
Test Results:
PARAMETER:
Lag Order: 0
STATISTIC:
Dickey-Fuller: -0.976
P VALUE:
0.3042
```

Z vyššie uvedeného ADF testu môžeme vidieť, že nulovú hypotézu časová rada nie je stacionárna by sme na hladine významnosti 0.05 ne zamietli. Takže naša časová rada je nestacionárna. Pre overenie skúsime KPSS test, ktorého nulová hypotéza je že časová rada je stacionárna.

```
> kpss.test(as.numeric(b$avg))
KPSS Test for Level Stationarity
data: as.numeric(b$avg)
KPSS Level = 2.5916, Truncation lag parameter = 3, p-value = 0.01
```

Na hladine významnosti by sme nulovú hypotézu zamietli. Oba testy nám ukázali, že časová rada nie je stacionárna. Následne môžeme pomocou KPSS testu testovať, či je daná časová rada trend

```
> kpss.test(as.numeric(b$avg), null='Trend')
KPSS Test for Trend Stationarity
data: as.numeric(b$avg)
KPSS Trend = 0.073128, Truncation lag parameter = 3, p-value = 0.1
```

stacionárna:

```
> ndiffs(as.numeric(b$avg))
[1] 1
```

```
> diff = diff(as.numeric(b$avg))
```

Z výstupu je jasné, že rada je trend stacionárna takže nulovú hypotézu na hladine 0.05 nezamietame.

Ďalej pokračujeme vykreslením ACF a PAC. Z obrázka parciálnej autokorelačnej funkcie ?? môžeme usúdiť, že do úvahy by spadali model  $AR(15)$  alebo  $MA(1)$ . Keďže ACF je možné obmedziť krivkou U, tak je lepšie vybrať model  $MA(1)$  [1].

Alternatívny spôsob voľby modelu je pomocou informačných kritérií. Tento spôsob je vhodný pre plne automatizované spracovanie [1] napríklad v ekonometrických softvéroch. K identifikácii modelu ARMA(p, q) sa priktupuje ako k minimalizácii funkcie 2.1

$$(\hat{p}, \hat{q}) = \arg \min_{(k, l)} A(k, l) \quad (2.1)$$

A je vhodné kritérium pre ktorého konštrukciu musíme odhadnúť model ARMA(k,l). Pri minimalizácii postupujeme postupne inkrementujeme obidva parametre k, l. V tejto práci sme zvolili Akaikeho informačné kritérium:

$$A(k,l) = AIC(k,l) = \ln \hat{\sigma}_{k,l}^2 + \frac{2(k+l+1)}{n} \quad (2.2)$$

Ako je vidieť z 2.2 kritérium prenalizuje veľké rády k a l.  $\hat{\sigma}_{k,l}^2$  je smerodajná odchylka reziduí.

### **3 Ekonomický model**

### **4 Ekonometrický model**

### **5 Odhadnutý model**

### **6 Interpretace výsledků**

### **7 Záver**

[1]

## **PodĎakovanie**

Na záver by som chcel poďakovať Ing. Danielovi Němcovi, Ph.D. za návrh na vypracovanie tejto témy a za veľmi príjemné a užitočné konzultácie. Ďalej by som chcel poďakovať Ester Železnákovéj za gramatickú korektúru textu.

## **Literatúra**

[1] Tomáš, C.: *Finanční ekonometrie*. Ekopress, 2008, ISBN 978-80-86929-43-9.

## **A Prílohy**