

Slovenská technická univerzita v Bratislave
Fakulta informatiky a informačných technológií

FIIT-0000-00000

Michal Fašánek

Pokročilé spracovanie sekvenčných dát pomocou umelých neurónových sietí

Diplomová práca

Vedúci práce: Ing. Michal Barla, PhD

Január, 2016

Slovenská technická univerzita v Bratislave
Fakulta informatiky a informačných technológií

FIIT-0000-00000

Michal Fašánek

Pokročilé spracovanie sekvenčných dát pomocou umelých neurónových sietí

Diplomová práca

Študijný program:	Informačné systémy
Študijný odbor:	Informačné systémy
Miesto vypracovania:	Ústav Aplikovanej Informatiky
Vedúci práce:	Ing. Michal Barla, PhD
Január, 2016	

>>>> ASSIGNMENT <<<<
>>>> ZADANIE <<<<

Anotácia

Slovenská technická univerzita v Bratislave

FAKULTA INFORMATIKY A INFORMAČNÝCH TECHNOLOGIÍ

Študijný program: Informačné systémy

Diplomová práca: Pokročilé spracovanie sekvenčných dát pomocou ume-
lých neurónových sietí

Autor: Michal Fašánek

Vedúci práce: Ing. Michal Barla, PhD

Január, 2016

Analýza dát z používateľského správania k zdrojom na webe je v súčasnosti populárna téma, vzhľadom na svoj potenciál zlepšovať služby poskytované návštevníkom webu. Najnovšie prístupy skúmajú aj možnosti aplikácie metód strojového učenia. Medzi týmito prístupmi si získavajú popularitu hlboké mnohovrstvové samoučiace sa neurónové siete a rôzne architektúry rekurentných neurónových sietí. Využívajú princípy učenia bez učiteľa, pomocou ktorých dokážeme v jednotlivých vzorkách dát identifikovať podstatné črty. V tejto práci sa zameriavam na možnosti využitia rekurentných neurónových sietí s dlhou krátkodobou pamäťou(LSTM) pri analýze dát z platobných brán pre používanie online spravodajských portálov. Takáto analýza poskytuje náhľad do používateľského správania a z toho vyplývajúce možnosti spätnej väzby voči návštevníkom online spravodajských portálov.

Annotation

Slovak University of Technology Bratislava

FACULTY OF INFORMATICS AND INFORMATION TECHNOLOGIES

Study program: Informačné systémy

Master thesis: Advanced processing of sequential data by artificial
neural networks

Author: Michal Fašánek

Supervisor: Ing. Michal Barla, PhD

2016, January

Data analysis of user behaviour in accessing web sources has become a popular topic, due to its potential in improvement of services offered to the visitors of web. Most recent approaches examine possibilities of applying methods machine-learning. Among these approaches, deep belief networks and recurrent neural networks are gaining popularity. They work with principles of unsupervised learning to identify important patterns in given data. In this paper, I focus on using recurrent neural networks with long short-term memory(LSTM) to analyze data from paywall of online news portals. Such analysis provides insight into the user behaviour and resulting feedback possibilities to the users of online news portals.

POĎAKOVANIE

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.....

ČESTNÉ PREHLÁSENIE

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua....

.....

Michal Fašánek

Obsah

1	Úvod	1
2	Analýza	2
2.1	Zvolená problémová oblasť	2
2.1.1	Úbytok zákazníkov	2
2.1.2	Moderovanie úbytku zákazníkov	3
2.1.2.1	Reaktívny prístup	3
2.1.2.2	Proaktívny prístup	3
2.2	Časť	3
2.2.1	Číslovaný zoznam	4
2.2.2	Citácia	4
2.2.3	Návestia & Referencie	4
2.2.4	Príklady	5
3	Dizajn	7
3.1	Časť	7
4	Výsledky	9
4.1	Časť	9
5	Záver	11
	Literatúra	A-1
A	Technická dokumentácia	A-1
A.1	Implementácia	A-1

B	Používateľská dokumentácia	B-1
B.1	Inštalácia	B-1
B.1.1	Spustenie aplikácie	B-1
C	Electronic medium	C-1

Zoznam obrázkov

2.1	Name figure	4
-----	-----------------------	---

Zoznam tabuliek

2.1	Information about users	6
-----	-----------------------------------	---

Zoznam ukážok

2.1	Príklad 1	5
2.2	Názov	5
2.3	Manager	6

Kapitola 1

Úvod

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat.

Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi. Nam liber tempor cum soluta nobis eleifend option congue nihil imperdiet doming id quod mazim placerat facer possim assum. Typi non habent claritatem insitam; est usus legentis in iis qui facit eorum claritatem. Investigationes demonstraverunt lectores legere me lius quod ii legunt saepius.

Claritas est etiam processus dynamicus, qui sequitur mutationem consuetudinum lectorum. Mirum est notare quam littera gothica, quam nunc putamus parum claram, anteposuerit litterarum formas humanitatis per seacula quarta decima et quinta decima. Eodem modo typi, qui nunc nobis videntur parum clari, fiant sollemnes in futurum.

Kapitola 2

Analýza

V tejto časti sa venujeme dôkladnej analýze podkladov. Jednotlivé časti sú popísané v rozsahu relevantnom pre túto prácu. Analýza je štrukturovaná na nasledovné časti:

- Zvolená problémová oblasť
- Dáta sprístupnené pre prácu
- Dostupné metódy
- Výskum v danej oblasti

2.1 Zvolená problémová oblasť

2.1.1 Úbytok zákazníkov

V tejto práci sa zameriavame na predikciu úbytku zákazníkov(churn rate) pri predplatiteľských službách. V súčasnej dobe sa do popredia biznis prístupov stále viac dostávajú prístupy riadenia vzťahov zo zákazníkmi(customer relationship management). Ukazuje sa totiž, že na trhu s dostatočným pokrytím poskytovateľov cieľovej služby je niekoľkonásobne drahšie získať si nového ako udržať si existujúceho zákazníka. Tento prístup však vyžaduje rozsiahlu znalosť dostupnej zákazníckej základne, ktorou poskytovateľ disponuje.

Predikcia úbytku zákazníkov sa venuje spracovaniu dostupných informácií o zákazníckej aktivite, službách ktoré využívajú a vývoja ich správania v čase.

Výsledkom takejto analýzy je štatistika poskytujúca informácie o jednotlivých zákazníkoch a ich šanci na presun k inému poskytovateľovi. Z týchto dát je následne odvoditeľné, aké percento zákazníkov odíde ku konkurencii a aký to bude mať dopad na finančné príjmy od ktorých je poskytovateľ závislý.

2.1.2 Moderovanie úbytku zákazníkov

Vo vzťahu k úbytku zákazníkov definuje CRM dva základné prístupy, ktorými je možné moderovať úbytok.

2.1.2.1 Reaktívny prístup

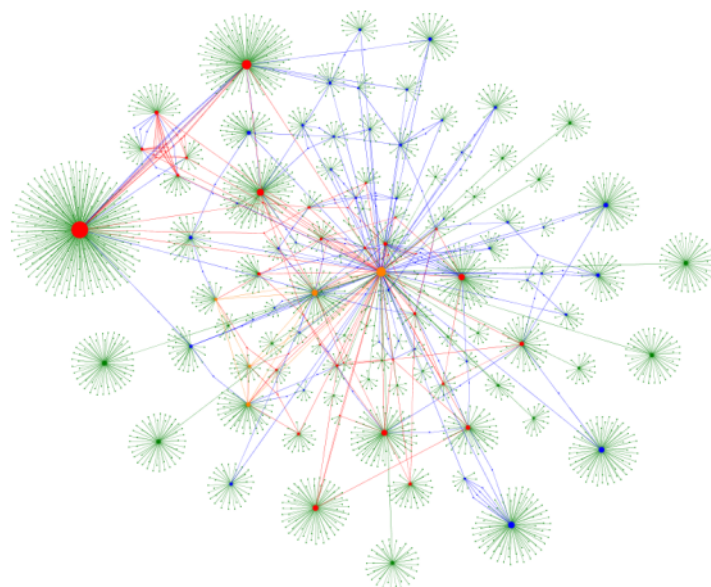
Motivácia zákazníka pre zotrvanie s pôvodným poskytovateľom služby nastáva, až keď sa zákazník explicitne rozhodne pre prechod ku konkurenčnému poskytovateľovi. V tomto okamihu začína poskytovateľ na svojho zákazníka apelovať výhodnými ponukami, zľavami alebo inými spôsobmi motivácie pre zotrvanie u poskytovateľa. Takýto prístup sa ukazuje ako ľahko zneužitelný ostatnými zákazníkmi, ktorí by inak nemali motiváciu pre prechod ku konkurencii. Predikcia úbytku zákazníkov v tomto prístupe nemá nijakú významnú úlohu.

2.1.2.2 Proaktívny prístup

Pri úspešnej predikcii záujmu zákazníka o prechod ku konkurenčnému poskytovateľovi je možné efektívne jeho zámer smerovať pozitívnou motiváciou. Tento prístup však predpokladá vysokú úspešnosť predikčných metód. Pri nesprávnej identifikácii zákazníckeho správania je totiž možné nielen nezabrániť zákazníkovi v presune ku konkurenčnému modelu, ale aj investíciou finančných prostriedkov do skupiny zákazníkov, ktorá by naďalej generovala zisk aj bez významnejšej motivácie, resp. nevrátila by rozdielom v úbytku motivačné náklady, ktoré na ňu daný poskytovateľ vynaložil.

2.2 Časť

V tejto časti sa venujeme



Obr. 2.1: *Name figure*

2.2.1 Číslovaný zoznam

1. cieľ 1
 - (a) cieľ 1.a
 - (b) cieľ 1.b
2. cieľ 2
3. cieľ 3

2.2.2 Citácia

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat [?].

2.2.3 Návestia & Referencie

Vid. sekcia 2.2.4.

Vid. ukážka ??.

Vid. číslovanie 2.1.

Vid. tabuľka ??.

2.2.4 Príklady

```
<table class="metric_index">
  <tr>
    <th>Lines of code</th>
    <th>Value</th>
  </tr>
  <% if (filenum and modulenum) then %>
    <tr>
      <td class="name">Number of files</td>
      <td class="value"><%=filenum%></td>
    </tr>
    <tr>
      <td class="name">Number of modules</td>
      <td class="value"><%=modulenum%></td>
    </tr>
  <% end %>
  <tr>
    <td class="name">Lines Total</td>
    <td class="value"><%=LOC.lines%></td>
  </tr>
  <!--
      skryty zdrojovy kod
      podobne zobrazenie ostatnych metrik riadkov
  -->
</table>
```

Ukážka 2.1: Príklad 1

```
local parser = require 'leg.parser'
local rules = require 'metrics.rules'
-- << skryty zdrojovy kod >> --
local capture_table = {}
grammar.pipe(LOC_capt.captures, AST_capt.captures)
grammar.pipe(block_capt.captures, LOC_capt.captures)
-- << viacero rovnakych volani s tabulkami captures inych modulov >> --
grammar.pipe(capture_table, cyclo_capt.captures)
local lua = lpeg.P(grammar.apply(parser.rules, rules.rules, capture_table))
```

```

local patt = lua / function (...)
    return {...}
end
local result = patt:match(code)[1]

```

Ukážka 2.2: *Názov*

Ukážka 2.3: *Manager*

```

int a;

```

Number of males	51
Number of woman	57
Gender not given	27
Average age	21,83

Tabuľka 2.1: *Information about users*

Kapitola 3

Dizajn

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat.

3.1 Časť

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi. Nam liber tempor cum soluta nobis eleifend option congue nihil imperdiet doming id quod mazim placerat facer possim assum. Typi non habent claritatem insitam; est usus legentis in iis qui facit eorum claritatem. Investigationes demonstraverunt lectores legere me lius quod ii legunt saepius. Claritas est etiam processus dynamicus, qui sequitur mutationem consuetudinum lectorum. Mirum est notare quam littera gothica, quam nunc putamus parum claram, anteposuerit litterarum formas humanitatis per seacula quarta decima et

quinta decima. Eodem modo typi, qui nunc nobis videntur parum clari, fiant sollemnes in futurum.

Kapitola 4

Výsledky

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat.

4.1 Časť

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi. Nam liber tempor cum soluta nobis eleifend option congue nihil imperdiet doming id quod mazim placerat facer possim assum. Typi non habent claritatem insitam; est usus legentis in iis qui facit eorum claritatem. Investigationes demonstraverunt lectores legere me lius quod ii legunt saepius. Claritas est etiam processus dynamicus, qui sequitur mutationem consuetudinum lectorum. Mirum est notare quam littera gothica, quam nunc putamus parum claram, anteposuerit litterarum formas humanitatis per seacula quarta decima et

quinta decima. Eodem modo typi, qui nunc nobis videntur parum clari, fiant sollemnes in futurum.

Kapitola 5

Záver

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi sit amet arcu. Fusce pharetra dapibus elit. Duis malesuada. Proin at elit vitae quam cursus tristique. Quisque fermentum. Praesent dictum. Nullam vehicula. Nunc pharetra dolor ut velit. Sed pulvinar, est sed congue tempor, nibh arcu cursus enim, quis consequat magna lacus sed pede. In sagittis. Etiam volutpat, velit id tincidunt egestas, augue ligula auctor eros, sit amet viverra sapien tortor at odio. In diam libero, fringilla ut, adipiscing condimentum, ultricies at, dui. Phasellus vitae risus.

Pellentesque vulputate ante ut diam. Sed adipiscing malesuada odio. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Nam a leo. Praesent velit. Aenean vehicula accumsan quam. Nulla dolor lorem, imperdiet a, ullamcorper hendrerit, ultrices at, urna. Integer placerat ligula id purus. Sed id nisl. Pellentesque tincidunt neque in lacus. In non quam et felis suscipit viverra.

Príloha A

Technická dokumentácia

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat.

A.1 Implementácia

Modul abc Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi. Nam liber tempor cum soluta nobis eleifend option congue nihil imperdiet doming id quod mazim placerat facer possim assum.

Modul def Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation

ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue dui dolore te feugait nulla facilisi. Nam liber tempor cum soluta nobis eleifend option congue nihil imperdiet doming id quod mazim placerat facer possim assum. Typi non habent claritatem insitam; est usus legentis in iis qui facit eorum claritatem. Investigationes demonstraverunt lectores legere me lius quod ii legunt saepius. Claritas est etiam processus dynamicus, qui sequitur mutationem consuetudinum lectorum. Mirum est notare quam littera gothica, quam nunc putamus parum claram, anteposuerit litterarum formas humanitatis per seacula quarta decima et quinta decima. Eodem modo typi, qui nunc nobis videntur parum clari, fiant sollemnes in futurum.

Príloha B

Používateľská dokumentácia

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat.

B.1 Inštalácia

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi.

B.1.1 Spustenie aplikácie

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat. Duis autem vel eum iriure

dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue dui dolore te feugait nulla facilisi.

Príloha C

Electronic medium

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat:

/Application

- implementácia opisovaného riešenia

/Documentation

- bakalárska práca spolu s anotáciami v slovenskom a anglickom jazyku

/Documentation/Latex

- latex zdrojové súbory dokumentácie

/Documentation/BibTeX

- BibTeX súbor s použitými referenciami

/Documentation/Resources

- dostupné použité zdroje

/Resources

- vstupné/testovacie dáta opisované v dokumente

/Source/Dependencies

- inštalačné súbory pre knižnice, ktoré potrebuje aplikácia

read.me - popis obsahu média v slovenskom a anglickom jazyku