Master's Thesis

Semantic approaches to citation recommendation

Tarek Saier

Examiners: Prof. Dr. Georg Lausen
Dr. Michael Färber

Albert-Ludwigs-University Freiburg

Faculty of Engineering

Department of Computer Science

Chair of Databases and Information Systems

April 30th, 2019

Writing Period

15. 10. 2018 – 30. 04. 2019

First Examiner

Prof. Dr. Georg Lausen

Second Examiner & Supervisor

Dr. Michael Färber

Master-Thesis

Semantic approaches to citation recommendation

Tarek Saier

Gutachter: Prof. Dr. Georg Lausen
Dr. Michael Färber

Albert-Ludwigs-Universität Freiburg

Technische Fakultät

Institut für Informatik

Lehrstuhl für Datenbanken und Informationssysteme

30. April 2019

Bearbeitungszeit

15. 10. 2018 – 30. 04. 2019

Erstgutachter

Prof. Dr. Georg Lausen

Zweitgutachter & Betreuer

Dr. Michael Färber

Declaration

I hereby declare, that I am the sole author and composer of my thesis and that no other sources or learning aids, other than those listed, have been used. Furthermore, I declare that I have acknowledged the work of others by providing detailed references of said work. I hereby also declare, that my Thesis has not been prepared for another examination or assignment, either wholly or excerpts thereof.

Place, Date

Signature

Abstract

foo bar

Zusammenfassung

fu bar

Contents

1 Introduction **1**

1.1 Example Section 1

2 Related Work **3**

3 Background **5**

4 Approach **7**

4.1 Problem Definition 7

5 Experiments **9**

6 Conclusion **11**

7 Acknowledgments **13**

Bibliography **13**

List of Figures

1 Caption that appears in the figlist 6

List of Tables

1 Table caption 9

List of Algorithms

1 Stochastic Gradient Descent: Neural Network 6

1 Introduction

Chapters are divided into sections.

1.1 Example Section

Copypasta of useful stuff below.

- Put a tilde (nbsp) in front of citations [1].
- **(TODO: Do this!)**
- **(EXTEND: Write more when new results are out!)**
- **(DRAFT: Hacky text!)**
- Chapter 1
- the colors of the Uni
  - UniBlue
  - UniRed
  - UniGrey
- a command for naming matrices **G**, and naming vectors **a**. This overwrites the default behavior of having an arrow over vectors, sticking to the naming conventions normal font for scalars, bold-lowercase for vectors, and bold-uppercase for matrices.

- named equations:

$$d(a, b) = d(b, a) \tag{1}$$

symmetry

- Use “these” for citing, not "these"
- If an equation is at the end of a sentence, add a full stop. If it’s not the end, add a comma: $a = b + c \quad (1)$,
- `https://en.wikipedia.org`
- Do not overuse footnotes[1] if possible.

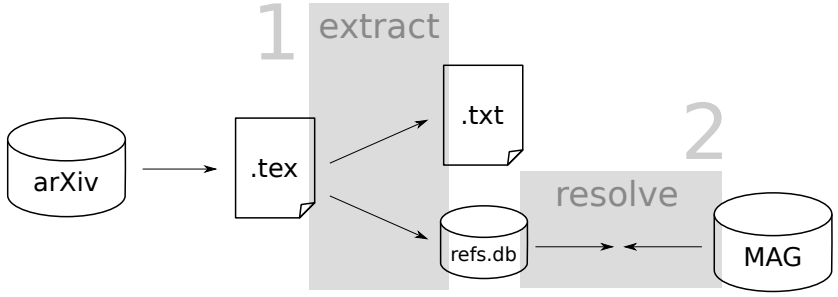
[1] `https://en.wikipedia.org`

2 Related Work

lots. pick wisely.

3 Background

explain all the things.



(a) subfloat

Figure 1: Caption that appears under the fig—do I want this in bold tho?

Algorithm 1 Stochastic Gradient Descent: Neural Network

Create a mini batch of m samples $\mathbf{x}_0 \dots \mathbf{x}_{m-1}$
foreach sample $\mathbf{x}$ **do**
  $\mathbf{a}^{\mathbf{x},0} \leftarrow \mathbf{x}$ ▷ Set input activation
  foreach Layer $l \in \{1 \dots L-1\}$ **do** ▷ Forward pass
    $\mathbf{z}^{\mathbf{x},l} \leftarrow \mathbf{W}^l \mathbf{a}^{\mathbf{x},l-1} + \mathbf{b}^l$
    $\mathbf{a}^{\mathbf{x},l} \leftarrow \varphi(\mathbf{z}^{\mathbf{x},l})$
  end for
  $\boldsymbol{\delta}^{\mathbf{x},L} \leftarrow \nabla_{\mathbf{a}} C_{\mathbf{x}} \odot \varphi'(\mathbf{z}^{\mathbf{x},L})$ ▷ Compute error
  foreach Layer $l \in L-1, L-2 \dots 2$ **do** ▷ Backpropagate error
    $\boldsymbol{\delta}^{\mathbf{x},l} \leftarrow ((\mathbf{W}^{l+1})^T \boldsymbol{\delta}^{\mathbf{x},l+1}) \odot \varphi'(\mathbf{z}^{\mathbf{x},l})$
  end for
end for
foreach $l \in L, L-1 \dots 2$ **do** ▷ Gradient descent
  $\mathbf{W}^l \leftarrow \mathbf{W}^l - \frac{\eta}{m} \sum_{\mathbf{x}} \boldsymbol{\delta}^{\mathbf{x},l} (\mathbf{a}^{\mathbf{x},l-1})^T$
  $\mathbf{b}^l \leftarrow \mathbf{b}^l - \frac{\eta}{m} \sum_{\mathbf{x}} \boldsymbol{\delta}^{\mathbf{x},l}$
end for

4 Approach

approach approach.

4.1 Problem Definition

define define.

5 Experiments

experiment experiment

Data set	#Papers	Cit. context	Disciplines	Full text	Ref. IDs
arXiv CS	90K	1 sentence	CS	yes	DBLP
CiteSeerX /RefSeer	1M	400 chars	all	no	no
PubMed Central OA[1]	2.3M	extractable	Biomed./Life Sci.	yes	mixed
Scholarly v2[2]	100K	extractable	CS	yes	no
ACL-ARC	11k	extractable	CS/comp. ling.	yes	no
ACL-AAN	18k	extractable	CS/comp. ling.	yes	no

Table 1: Table caption. foo bar...

[1] `https://www.ncbi.nlm.nih.gov/pmc/tools/openftlist/`
[2] `http://www.comp.nus.edu.sg/~sugiyama/SchPaperRecData.html`

6 Conclusion

conclude conclude.

7 Acknowledgments

acknowledge acknowledge.

- advisers
- examiner
- person1 for the x
- person2 for the y

Bibliography

[1] M. J. Moravcsik and P. Murugesan, "Some results on the function and quality of citations," *Social Studies of Science*, vol. 5, no. 1, pp. 86–92, 1975.