Praktikumsbericht

Machine Learning for Natural Language Processing

am Lehrstuhl für Informatik X

Konstantin Herud Thomas Schaffroth Maximilian Meißner

Abgabedatum:	16. August 2020
Betreuer:	Prof. Dr. Andreas Hotho
	Daniel Schlör
	Albin Zehe
	Konstantin Kobs
	Tobias Koopmann

Julius-Maximilians-Universität Würzburg
Lehrstuhl für Informatik X
Data Science

Zusammenfassung

Dieses Dokument soll Studenten an unserem Lehrstuhl bei der Erstellung ihrer Abschlussarbeit unterstützen. Wir zeigen eine beispielhafte Gliederung einer Arbeit und beschreiben die Inhalten der einzelnen Kapitel. Zusätzlich geben wir an vielen Stellen auch Hinweise zur Benutzung von LaTeX für die Erstellung der Arbeit. Im Anhang **??** geben wir ein paar Hinweise zum Ablauf der Betreuung von Abschlussarbeiten an unserem Lehrstuhl.

Zur Handhabung dieses Pakets. In diesem Paket sind Vorlagen für verschiedene Dokumenttypen enthalten, die sie als Ausgangspunkt für ihre Arbeit verwenden können. Es gibt jeweils Vorlagen für deutsche und englische Arbeiten.

- `template_thesis_de.tex`, `template_thesis_en.tex`: Vorlage für Bachelorarbeit bzw. Masterarbeit
- `template_seminar_de.tex`, `template_seminar_en.tex`: Vorlage für Seminarausarbeitungen und Praktikumsberichte

Der Quelltext zu diesem Leitfaden ist ebenfalls im Paket enthalten. Diesen können Sie als praktisches Beispiel dafür verwenden, wie diese Dokumentenklasse angewandt wird.

Inhalt der Zusammenfassung. Schreiben Sie hier eine Zusammenfassung der Arbeit, vergleichbar mit dem Abstract auf wissenschaftlichen Papers. Sie dient dem Leser dazu, einen groben Überblick über die Inhalte zu gewinnen (Problemstellung, verwendeter Lösungsansatz, ggf. experimentelle Ergebnisse, gewonnene Erkentnisse). Der Umfang soll ca. eine halbe Seite betragen. Für Seminararbeiten ist diese Zusammenfassung nicht erforderlich.

Achtung: Bei Arbeiten auf Englisch fordern die Prüfungsordnungen, dass es eine deutsche Zusammenfassung gibt. Schreiben Sie in diesem Fall eine englische *und* eine deutsche Zusammenfassung (mit dem gleichen Inhalt). Die passenden LaTeX-Befehle dafür finden Sie in den englischsprachigen Vorlagen.

WARNUNG: Die vorliegende Version des Leitfadens ist eine Vorabversion, die noch nicht vollständig ist. Sie bezieht sich größtenteils auf die Ausarbeitung von Bachelor- und Masterarbeiten; Seminararbeiten unterscheiden sich davon etwas in Aufbau und Inhalt.

Inhaltsverzeichnis

1. Methodik

1.1. Task 2: Klassifikation

- z: Zeilen im Dokument
- t: Tokens pro Zeile (Padding kürzerer Zeilen)
- d: Dimension des Modells (z. B. 150)
- c: Anzahl Klassen

$$Y_{ztd} = \mathrm{LSTM}_t(X_{ztd}) \tag{1.1}$$

$$A_{zt} = \mathrm{softmax}_t\left(\sum_d Y_{ztd} W_d^1\right) \tag{1.2}$$

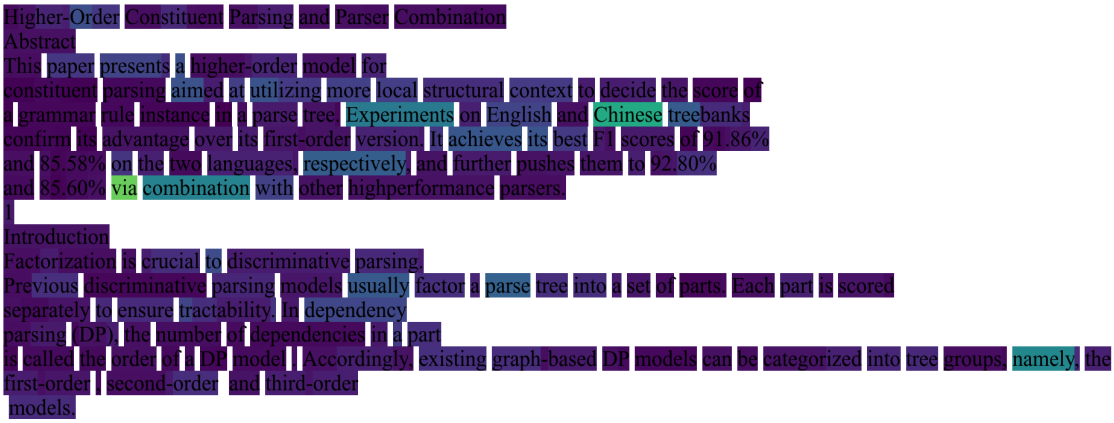
$$Y_{zd} = \sum_t Y_{ztd} A_{zt} \tag{1.3}$$

$$A_z = \mathrm{softmax}_z\left(\sum_d Y_{zd} W_d^2\right) \tag{1.4}$$

$$Y_d = \sum_z Y_{zd} A_z \tag{1.5}$$

$$Y_c = Y_d W_{dc}^3 \tag{1.6}$$

$$W_{L,C_i} = \left(\frac{|C_{\max}|}{|C_i|}\right)^{\alpha}, \quad \alpha = 0.6 \tag{1.7}$$



Wahrheit \ Voraussage	AU	CA	CN	FR	DE	IN	IL	JP	SG	CH	UK	USA
AU	6		8		1						7	6
CA	1	8	4					2		1	7	32
CN	1	1	188		1			2	1		1	17
FR			1	36	3			1			2	4
DE		2	2	2	47			1			12	11
IN			1			6					2	11
IL		1		1			14	2				10
JP			2	1				45		1	1	6
SG		1	6						9		2	8
CH			3		7					4	4	8
UK	3	5	1	6	1	2		1		2	66	18
USA	3	13	38	17	6	1	2	8	2	3	32	746

Tab. 1.1.: Konfusionsmatrix der Klassen Australien (AU), Canada (CA), China (CN), Frankreich(FR), Deutschland (DE), Indien (IN), Israel (IL), Japan (JP), Singapur (SG), Schweitz (CH), Großbritannien (UK), Amerika (USA). Leere Zellen implizieren Nullen.

A. Formelles/LaTeX

Erklärung

Hiermit versichere ich die vorliegende Abschlussarbeit selbstständig verfasst zu haben, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt zu haben, und die Arbeit bisher oder gleichzeitig keiner anderen Prüfungsbehörde unter Erlangung eines akademischen Grades vorgelegt zu haben.

Würzburg, den 2. August 2020

. . . . . .
Konstantin Herud Thomas Schaffroth Maximilian Meißner