Friedrich-Schiller-Universität Jena
Institut für Informatik
Studiengang Informatik, B.Sc.

Transferring Relevance Judgments with Pairwise Preferences

Bachelorarbeit

Fabian Hofer
geb. am: 22.08.2002 in Eisenach

Matrikelnummer 199111

1. Gutachter: Prof. Dr. Matthias Hagen
2. Gutachter: Maik Fröbe

Datum der Abgabe: 31. Februar 2024

Erklärung

Hiermit versichere ich, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Jena 07743, Germany, 31. Februar 2024

. . . . . .
Fabian Hofer

Literaturverzeichnis

Hyung Won Chung, Le Hou, Shayne Longpre, Barret Zoph, Yi Tay, William Fedus, Eric Li, Xuezhi Wang, Mostafa Dehghani, Siddhartha Brahma, Albert Webson, Shixiang Shane Gu, Zhuyun Dai, Mirac Suzgun, Xinyun Chen, Aakanksha Chowdhery, Sharan Narang, Gaurav Mishra, Adams Yu, Vincent Y. Zhao, Yanping Huang, Andrew M. Dai, Hongkun Yu, Slav Petrov, Ed H. Chi, Jeff Dean, Jacob Devlin, Adam Roberts, Denny Zhou, Quoc V. Le, and Jason Wei. Scaling instruction-finetuned language models. *CoRR*, abs/2210.11416, 2022. doi: 10.48550/ARXIV.2210.11416. URL `https://doi.org/10.48550/arXiv.2210.11416`.

Maik Fröbe, Janek Bevendorff, Lukas Gienapp, Michael Völske, Benno Stein, Martin Potthast, and Matthias Hagen. Copycat: Near-duplicates within and between the clueweb and the common crawl. In Fernando Diaz, Chirag Shah, Torsten Suel, Pablo Castells, Rosie Jones, and Tetsuya Sakai, editors, *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021*, pages 2398–2404. ACM, 2021. doi: 10.1145/3404835.3463246. URL `https://doi.org/10.1145/3404835.3463246`.

Lukas Gienapp, Maik Fröbe, Matthias Hagen, and Martin Potthast. Sparse pairwise re-ranking with pre-trained transformers. In Fabio Crestani, Gabriella Pasi, and Éric Gaussier, editors, *ICTIR '22: The 2022 ACM SIGIR International Conference on the Theory of Information Retrieval, Madrid, Spain, July 11 - 12, 2022*, pages 72–80. ACM, 2022. doi: 10.1145/3539813.3545140. URL `https://doi.org/10.1145/3539813.3545140`.

Sean MacAvaney and Luca Soldaini. One-shot labeling for automatic relevance estimation. In Hsin-Hsi Chen, Wei-Jou (Edward) Duh, Hen-Hsen Huang, Makoto P. Kato, Josiane Mothe, and Barbara Poblete, editors, *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2023, Taipei, Taiwan, July 23-27, 2023*, pages 2230–2235. ACM, 2023. doi: 10.1145/3539618.3592032. URL `https://doi.org/10.1145/3539618.3592032`.

Joel Mackenzie, Matthias Petri, and Alistair Moffat. A sensitivity analysis of the MSMARCO passage collection. *CoRR*, abs/2112.03396, 2021. URL `https://arxiv.org/abs/2112.03396`.

Christopher D. Manning and Hinrich Schütze. *Foundations of Statistical Natural Language Processing*. MIT Press, 2001. ISBN 978-0-262-13360-9.

Ronak Pradeep, Rodrigo Frassetto Nogueira, and Jimmy Lin. The expando-mono-duo design pattern for text ranking with pretrained sequence-to-sequence models. *CoRR*, abs/2101.05667, 2021. URL `https://arxiv.org/abs/2101.05667`.

Shengyao Zhuang, Honglei Zhuang, Bevan Koopman, and Guido Zuccon. A setwise approach for effective and highly efficient zero-shot ranking with large language models. In Grace Hui Yang, Hongning Wang, Sam Han, Claudia Hauff, Guido Zuccon, and Yi Zhang, editors, *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2024, Washington DC, USA, July 14-18, 2024*, pages 38–47. ACM, 2024. doi: 10.1145/3626772.3657813. URL `https://doi.org/10.1145/3626772.3657813`.

Kapitel 1

The Structure of a Thesis in Computer Science

A thesis in computer science follows the structure of academic writing and consists of several chapters, each chapter contains several sections, and each section can contain several subsections. Sections an subsections contain several paragraphs of text as well as, for example, lists, tables, and figures. It is recommended to start your thesis by studying the scientific toolbox [1].

1.1 References

Elements like sections, figures, or tables can be referenced (see Section 1.1) by assigning a `\label{label-name}` after the `\section` command or within a `\begin{...}` environment and referencing the label with `\ref{label-name}`.

1.2 Links and Citations

In scientific work, all external sources must be cited or linked.

Scientific Articles Papers, Textbooks, theses, or other scientific work should always be cited as `Article, InProceedings, or Book` via bibtex. To cite a work, add a corresponding bibtex-entry to your `literature.bib` and cite this entry with `\cite{bib-key}`: Manning and Schütze [2001], `\citep{bib-key}`: [Manning and Schütze, 2001], `\citet{bib-key}`: Manning and Schütze [2001], or `\citeauthor{bib-key}`: Manning and Schütze.

[1] Find our notes on scientific work at `https://webis.de/lecturenotes.html#part-scientific-toolbox`

Non-scientific Articles Journalistic articles, books, blog-posts, and various web sources with a known author and title should generally be cited as a `Article, Book, or Misc` bibtex-entry. For web-sources, provide a url and a date of last access.

Other Sources Other sources, like images, illustrations, libraries, or source code can be cited by proving a link an a date of last access in a footnote on the page.

Kapitel 2

How to Display Results

Results should be displayed primarily in tables (see Table 2.1), figures (see Figure 2.1), enumerations, and itemizations.

2.1 Figures and Tables

To demonstrate the usage of figures and tables, here are some examples like Figure 2.1 and some table with some numbers (Table 2.1) that for some reason deserves to be on an extra page.

The Table 2.2 and Table 2.3 show how every figure and table should be referenced in the text.

2.2 Equations and Code Listings

Equations should be contained within an `\begin{equation}` environment, as with Equation 2.1. Formulas can be set inline with a $ math environment $, so `$ f(x)=\frac{P(x)}{Q(x)} $` produces $f(x) = \frac{P(x)}{Q(x)}$.

$$f(x) = \frac{P(x)}{Q(x)} \quad \text{and} \quad f(x) = \tfrac{P(x)}{Q(x)} \tag{2.1}$$

Code or pseudocode can be included within listings[1].

[1] Code listings are explained in detail in the Overleaf LaTeX guide: `https://www.overleaf.com/learn/latex/Code_listing`

A

Abbildung 2.1: The first letter in the Roman alphabet.

Tabelle 2.1: Tables have their captions above, figures below.

Some numbers			
	1999	2000	2001
Distance (km)	23.0	18.0	42.0
Awesomeness (aws)	3.2	8.1	2.4

Tabelle 2.2: A matrix showing which attributes an entity has (✔), partially has ((✔)), or does not have (✖).

Entities	**Class 1**		**Class 2**	
	Attribute1	Attribute2	Attribute3	Attribute4
Entity1	✔	✖	(✔)	✖
Entity2	✖	(✔)	✔	(✔)
Entity3	(✔)	✔	✖	✔
Entity4	✔	✖	(✔)	✖
Entity5	✖	(✔)	✔	(✔)

Tabelle 2.3: A heatmap.

Entities	**Attribute1**	**Attribute2**	**Attribute3**	**Attribute4**
Entity1	0.72	0.05	0.20	0.20
Entity2	1.00	0.11	0.12	0.91
Entity3	0.63	0.36	0.68	0.27
Entity4	0.69	0.48	0.28	0.59
Entity5	0.12	0.69	0.82	0.42
Entity6	0.23	0.07	0.45	0.85