

SoSe 2023

# NLP-gestützte Data Science

## Übung 2

Manuel Stoeckel

Prof. Dr. Alexander Mehler

Frist: 29. Juni 2023

### Formalia

Die Übungen der Veranstaltung *NLP-gestützte Data Science* dienen der Vertiefung der in der Vorlesung behandelten Themen und einem praktischen Einblick in die diskutierten Probleme und Ihrer Lösungen. Es wird insgesamt **drei** Übungsblätter geben, mit denen Sie bis zu **150 Punkte** erreichen können, welche Ihnen bei der Modulabschlussprüfung zu **einem Zehntel** als **Bonuspunkte** angerechnet werden. Ziehen Sie hierzu die entsprechenden Abschnitte Ihrer Prüfungsordnung zu rate.<sup>1</sup>

#### Allgemein gilt für Abgaben:

Die Abgabe erfolgt im **PDF Format** via OLAT. Bitte stellen Sie sicher, dass **Ihr Name und Ihre Matrikelnummer** auf jeder Abgabe vermerkt sind. Wir empfehlen Ihnen die Verwendung von  $\LaTeX$  zur Erstellung Ihrer Abgabedokumente. Dazu liegt ein Template in OLAT bereit. Es gelten die gängigen Regeln für Plagiate: stellen Sie sicher, dass Sie Ihre Abgaben selbstständig erstellt haben und etwaige Fremdinhalte entsprechend gekennzeichnet und korrekt zitiert haben.

#### Für Programmieraufgaben gilt gesondert:

Neben der Abgabe Ihrer Ergebnisse im geforderten Format, ist zusätzlich der gesamte Quellcode ZIP-komprimiert einzureichen. Bitte achten Sie darauf, dass Sie **keine Binaries oder Bibliotheken** mit abgeben (z.B. `virtualenv`, Python Byte-Code, `git`-Repositories, etc.). Der Quellcode ist zu kommentieren. Fremdcode ist entsprechend zu kennzeichnen, auch hier gelten die gängigen Regeln für Plagiate. Wir empfehlen die Verwendung von `conda` zur Umgebungsverwaltung, z.B. mit `miniconda`. Bitte verwenden Sie eine Version von Python  $\geq 3.10$ .

#### Bei Gruppenabgaben gilt gesondert:

Der individuelle Beitrag jedes Gruppenmitglieds muss aus der Abgabe ersichtlich sein.

---

<sup>1</sup>§36 Abs. 6 Satz 2 der Ordnung für den Bachelorstudiengang Informatik bzw. §35 Abs. 5 Satz 2 der Ordnung für den Masterstudiengang Informatik, jeweils in der Fassung vom 17. Juni 2019.

## Übung 2: Transformer

max. 50 P

In dieser Übung werden wir uns mit der Transformer-Architektur (Vaswani et al., 2017) beschäftigen. Für diese Übung ist der genaue Modellaufbau nicht von so großer Signifikanz wie bei Übung 1; vielmehr werden wir uns dieses Mal mit der Verwendung von fertigen Modellen und Analyse der Modellverhaltensweisen beschäftigen. Für dieses Blatt wird es keine *automatische* Evaluation geben; bitte kommentieren Sie daher Ihren Code ausreichend. Die Diskussion der Ergebnisse zu 2.1 und die Lösung zu 2.2 sind in einer der Abgabe beiliegenden PDF-Datei zu sammeln.

### 2.0 Einarbeitung

Die Transformer-Architektur ist vielseitig und in NLP aktuell allgegenwärtig. Entsprechend werden Sie unzählige Blogs, Videos und detaillierte Erklärungen zu jedem erdenklichen Anwendungsszenario nach kurzer Suche im Internet finden. Die folgenden Ressourcen sollen daher keine abschließende Sammlung darstellen, sondern dienen nur zur Orientierung.

#### Forschung

- › „Attention is All You Need“, Vaswani et al. (2017)
- › „BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding“, Devlin et al. (2018)
- › „Improving Language Understanding by Generative Pre-Training“, Radford et al. (2018)
- › „RoBERTa: A Robustly Optimized BERT Pretraining Approach“, Liu et al. (2019)
- › „ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators“, Clark et al. (2020)
- › „Learning Transferable Visual Models From Natural Language Supervision“, Radford et al. (2021)
- › „Zero-Shot Text-to-Image Generation“, Ramesh et al. (2021)
- › „High-Resolution Image Synthesis with Latent Diffusion Models“, Rombach et al. (2021)

#### Blogs

- › Jay Alammar's Blog: <https://jalammar.github.io/>
  - ›› [Seq2Seq with Attention](#)
  - ›› [The Illustrated Transformer](#)
- › Christopher Olah' Blog: <https://colah.github.io/about.html>
  - ›› [Attention \(on Distill\)](#)
- › Stephen Wolfram's Blog: <https://writings.stephenwolfram.com/>
  - ›› [What Is ChatGPT Doing ... and Why Does It Work?](#)

#### Notebooks & Tutorials:

- › 🤖 transformers Notebooks: <https://huggingface.co/docs/transformers/notebooks>
- › „Natural Language Processing with Transformers“, Tunstall (2021)
  - ›› Buch: <https://ubffm.hds.hebis.de/Record/HEB479538573>
  - ›› Notebooks: <https://github.com/nlp-with-transformers/notebooks>

## 2.1 Benchmarking Transformers: oLMpics

40 P

Neben dem in der Vorlesung vorgestellten Benchmark für LMs, *GLUE* (Wang et al., 2019), welcher sich auf die Performanz von Modellen nach dem *fine tuning* auf bestimmten *downstream tasks* fokussiert, existieren auch einige Evaluationsdatensätze, die einen eher intrinsischen Ansatz verfolgen. Dazu gehört der *oLMpics* Benchmark (Talmor et al., 2020), der acht Experimente umfasst, welche jeweils das „logische Denkenvermögen“ von Sprachmodellen testen sollen. In dieser Aufgabe werden Sie sich mit zwei dieser Experimente befassen und überprüfen, ob Sie die Ergebnisse von aus dem *oLMpics* Paper reproduzieren können.

- › Machen Sie sich mit der 😊 transformers Bibliothek<sup>2</sup> vertraut.
  - ›› Hier insbesondere mit Masked Language Models und wie Sie die wahrscheinlichste Vorhersage aus einer Auswahl von Tokens für ein maskiertes Token in einem Eingabesatz mit genau einem [MASK] Token erhalten können.
- › Lesen Sie das Paper von Talmor et al. (2020).
- › Laden Sie sich die dev Daten für die Always-Never und Age Comparison Experimente aus dem offiziellen GitHub Repository<sup>3</sup> herunter.
- › Implementieren Sie den MC-MLM Versuchsaufbau mit der 😊 transformers Bibliothek.
- › Reproduzieren Sie je den *zero shot* Test mit den Modellen roberta-base und bert-base-cased.
  - ›› Die Gewichte für die Modelle und eine Anleitung, wie Sie diese laden können, finden Sie einer kurzen Suche nach den Modellnamen hier: <https://huggingface.co/models>
- › Dokumentieren und interpretieren Sie Ihre Ergebnisse.

### Hinweise

- › Falls Ihnen kein persönlicher Rechner mit ausreichender Leistung zur Verfügung steht, können Sie Ihre Experimente auf den Rechnern der RBI durchführen.

## 2.2 Attention Beyond Sub-Words

2 × 5 P

Die in der Vorlesung vorgestellten *Transformer*-Modelle verwenden hauptsächlich *Self Attention*, eine Variante des *Attention*-Mechanismus, bei dem *Query*-, *Key*- und *Value*-Vektoren alle den selben Ursprung haben: in der Regel handelt es sich dabei dann je um Repräsentationen von *sub-word units*.

- › Überlegen Sie sich **zwei** Modellalternativen, in denen der *Attention*-Mechanismus nicht ausschließlich auf der Ebene von Wörtern oder *sub-word units* zum Einsatz kommt.
- › Beschreiben Sie die möglichen Quellen für Ihre Repräsentationen und zu welchem Zweck diese eingesetzt werden könnten.

### Hinweise

- › Wenn Sie Ansätze aus der Literatur aufgreifen, müssen diese korrekt zitiert und ausreichend diskutiert werden.

---

<sup>2</sup><https://huggingface.co/docs/transformers/index>

<sup>3</sup><https://github.com/alontalmor/oLMpics>

## Bonus: oLMpics – Perturbed Language Age Comparison

1 P

- › Ersetzen Sie die Wörter age und than in den Prompts des Age Comparison Tests durch Nonsenswörter wie blah und da.
- › Wiederholen Sie das Experiment und dokumentieren Sie etwaige Veränderungen in den Ergebnissen.

## Bonus: oLMpics – Alle Tests

4 × 1 P

Führen Sie die Experimente aus 2.1 auch für die anderen MC-MLM Tests durch:

- › Objects Comparison
- › Taxonomy Conjunction
- › Antonym Negation
- › Multi-Hop Composition

## Literatur

- Clark, Kevin et al. (2020). „ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators“. In: *CoRR* abs/2003.10555. arXiv: [2003.10555](https://arxiv.org/abs/2003.10555).
- Devlin, Jacob et al. (2018). „BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding“. In: *CoRR* abs/1810.04805. arXiv: [1810.04805](https://arxiv.org/abs/1810.04805).
- Liu, Yinhan et al. (2019). „RoBERTa: A Robustly Optimized BERT Pretraining Approach“. In: *CoRR* abs/1907.11692. arXiv: [1907.11692](https://arxiv.org/abs/1907.11692).
- Radford, Alec et al. (2018). „Improving Language Understanding by Generative Pre-Training“. In: URL: [https://cdn.openai.com/research-covers/language-unsupervised/language\\_understanding\\_paper.pdf](https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf).
- Radford, Alec et al. (2021). „Learning Transferable Visual Models From Natural Language Supervision“. In: *CoRR* abs/2103.00020. arXiv: [2103.00020](https://arxiv.org/abs/2103.00020).
- Ramesh, Aditya et al. (2021). „Zero-Shot Text-to-Image Generation“. In: *CoRR* abs/2102.12092. arXiv: [2102.12092](https://arxiv.org/abs/2102.12092).
- Rombach, Robin et al. (2021). „High-Resolution Image Synthesis with Latent Diffusion Models“. In: *CoRR* abs/2112.10752. arXiv: [2112.10752](https://arxiv.org/abs/2112.10752).
- Talmor, Alon et al. (Dez. 2020). „oLMpics-On What Language Model Pre-training Captures“. In: *Transactions of the Association for Computational Linguistics* 8, S. 743–758. ISSN: 2307-387X. DOI: [10.1162/tac1\\_a\\_00342](https://doi.org/10.1162/tac1_a_00342). eprint: [https://direct.mit.edu/tac1/article-pdf/doi/10.1162/tac1\\_a\\_00342/1923716/tac1\\_a\\_00342.pdf](https://direct.mit.edu/tac1/article-pdf/doi/10.1162/tac1_a_00342/1923716/tac1_a_00342.pdf).
- Tunstall, Lewis (2021). *Natural Language Processing with Transformers*. Hrsg. von Leandro von Werra und Thomas Translators Wolf. O'Reilly Media, Inc. ISBN: 9781098103231. URL: <https://ubffm.hds.hebis.de/Record/HEB479538573>.
- Vaswani, Ashish et al. (2017). „Attention is All you Need“. In: *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*. Hrsg. von Isabelle Guyon et al., S. 5998–6008. URL: <https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html>.
- Wang, Alex et al. (2019). „GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding“. In: *International Conference on Learning Representations*. URL: <https://openreview.net/forum?id=rJ4km2R5t7>.