

Evaluation verschiedener Methoden zur Datenaugmentierung für die Aspekt-basierte Sentiment Analyse im Deutschen

Thema:

Evaluation verschiedener Methoden zur Datenaugmentierung für die Aspekt-basierte Sentiment Analyse im Deutschen

Art:

BA, MA

BetreuerIn:

Jakob Fehle

Status:

ausgeschrieben

Stichworte:

Sentiment Analyse, NLP, Data Augmentation

angelegt:

2023-12-15

Hintergrund

Die Aspektbasierte Sentiment Analyse (ABSA) ist eine Methode im Bereich des Natural Language Processing (NLP), die es ermöglicht, Meinungen und Stimmungen in Texten gegenüber spezifischen Aspekten zu identifizieren. Im Gegensatz zur herkömmlichen Sentiment Analyse, die lediglich das allgemeine Sentiment eines Textes ermittelt, fokussiert sich die ABSA auf einzelne Aspekte eines Produkts oder einer Dienstleistung. Dadurch können detaillierte Einblicke in das Feedback zu Produkten und Dienstleistungen gewonnen und gezielt auf die Bedürfnisse und Vorlieben eingegangen werden. Im deutschen Sprachraum stellt sich jedoch die Herausforderung, dass nur wenige ausreichend große und brauchbare Datensätze für die Aspekt-basierte Sentiment Analyse existieren. Zudem ist die manuelle Annotation neuer Datensätze zeit- und kostenintensiv.

Die Augmentierung von Daten ist hierbei eine mögliche Lösung, um aus einer kleinen Menge an annotierten Beispielen eine große Menge an brauchbaren Trainingsdaten zu generieren. Für die Augmentierung existieren verschiedene Ansätze, wie z.B. Back-Translation, Noise-Induction oder Synonym-Replacement, welche bereits in anderen NLP-Bereichen erfolgreich eingesetzt werden konnten. Ziel ist es, kleine Datensätze durch zusätzliche Annotationen oder Samples zu erweitern, um eine effektive Anwendung von ABSA-Methoden zu ermöglichen.

Zielsetzung der Arbeit

Die Zielsetzung dieser Arbeit besteht in einer umfassenden Analyse geeigneter Techniken zur Datenaugmentierung im Kontext der ABSA im Deutschen. Hierzu sollen auch die Auswirkungen der Anwendung dieser Techniken auf potenzielle Methodiken zur Aspekt-basierten Sentiment Analyse untersucht werden. Die verschiedenen Techniken zur Datenaugmentierung sollen auf einen

bestehenden Datensatz angewandt werden (oder auf Teile eines vollständigen Datensatzes, um ein „Low Resource Setting“ zu simulieren), und der daraus resultierende, augmentierte Datensatz wird anschließend für das Training von Machine Learning-Modellen für die ABSA verwendet und evaluiert. Ein Vergleich der Ergebnisse soll dazu dienen, die effektivsten Techniken zu identifizieren.

Konkrete Aufgaben

- Literaturrecherche zu bestehenden Techniken der Datenaugmentierung im NLP-Bereich und insbesondere im Kontext der Sentiment Analyse.
- Auswahl eines geeigneten Datensatzes.
- Implementierung und Anwendung verschiedener Techniken der Datenaugmentierung auf den ausgewählten Datensatz.
- Entwicklung und Durchführung einer Methodik zur Evaluation der Effektivität der angewandten Techniken.

Erwartete Vorkenntnisse

Erfahrungen mit Python und NLP von Vorteil.

Weiterführende Quellen

- Feng, Steven Y., Varun Gangal, Jason Wei, Sarath Chandar, Soroush Vosoughi, Teruko Mitamura, and Eduard Hovy. 2021. “A Survey of Data Augmentation Approaches for NLP.” arXiv [cs.CL]. arXiv. <http://arxiv.org/abs/2105.03075>.
- Liesting, Tomas, Flavius Frasincar, and Maria Mihaela Trușcă. 2021. “Data Augmentation in a Hybrid Approach for Aspect-Based Sentiment Analysis.” In Proceedings of the 36th Annual ACM Symposium on Applied Computing, 828–35. SAC '21. New York, NY, USA: Association for Computing Machinery.
- Li, Guangmin, Hui Wang, Yi Ding, Kangan Zhou, and Xiaowei Yan. 2023. “Data Augmentation for Aspect-Based Sentiment Analysis.” International Journal of Machine Learning and Cybernetics 14 (1): 125–33.

From:

<https://wiki.mi.ur.de/> - MI Wiki

Permanent link:

<https://wiki.mi.ur.de/arbeiten/data-augmentation-absa?rev=1702858403>

Last update: **18.12.2023 00:13**

