

Die Annotation komplexer Trainingsdaten in Low-Resource Sprachen am Beispiel der Aspekt-basierten Sentiment Analyse

Thema:

Die Annotation komplexer Trainingsdaten in Low-Resource Sprachen am Beispiel der Aspekt-basierten Sentiment Analyse

Art:

BA, MA

BetreuerIn:

Jakob Fehle

Status:

Entwurf

Stichworte:

Sentiment Analyse, Annotation, NLP

angelegt:

2023-08-05

Hintergrund

Die Sentiment Analyse ist ein Forschungsgebiet der natürlichen Sprachverarbeitung, das sich mit der automatisierten Identifizierung und Klassifizierung von Meinungen in Texten beschäftigt.

Insbesondere in Bereichen wie Kundenbewertungen, sozialen Medien und Produktrezensionen spielt die Sentiment Analyse eine entscheidende Rolle bei der Extraktion von nutzbaren Informationen aus großen Datenmengen.

Die Aspekt-basierte Sentiment Analyse erweitert dieses Konzept, indem sie nicht nur das globale Sentiment eines Textes misst, sondern auch die spezifischen Aspekte oder Themen, die innerhalb des Textes diskutiert werden. Dies ermöglicht eine feinere Granularität und tiefere Einsicht in die Bewertungen, was besonders in Domänen mit komplexen Inhalten von Vorteil ist. In Low-Resource Sprachen wie dem Deutschen sind für bestimmte Anwendungsfälle oder -domänen oft keine Trainingsdaten vorhanden. Zudem stellt die Erstellung von qualitativ hochwertigen Trainingsdaten eine Herausforderung dar und ist oft zeitintensiv und somit kostspielig.

Zielsetzung der Arbeit

Das Hauptziel dieser Arbeit besteht darin, verschiedene Arten der Annotation, nämlich Annotationen durch Experten, Studenten und Crowd-Sourcing, zu evaluieren und deren Auswirkungen auf die Datengrundlage für Machine Learning Methoden zur Aspekt-basierten Sentiment Analyse zu untersuchen. Im Fokus steht die Analyse möglicher Qualitätsunterschiede zwischen den Annotationen und deren Konsequenzen für die Leistungsfähigkeit von Machine Learning Modellen.

Konkrete Aufgaben

- Vertiefte Auseinandersetzung mit existierenden Methoden der Sentiment Analyse und Aspekt-basierten Sentiment Analyse.
- Erstellung eines Ground Truth Datensatzes durch Expertenbewertungen mit Majority Decision in der deutschen Sprache.
- Durchführung von Annotationsstudien durch verschiedene Gruppen (Experten, Studenten, Crowd-Sourcing) zur Erzeugung von Trainingsdaten.
- Vergleich der Annotationsqualität mit dem Ground Truth Datensatz und detaillierte Analyse möglicher Qualitätsunterschiede.
- Implementation und Evaluation eines Ansatzes zur Aspekt-basierten Sentiment Analyse unter Verwendung der verschiedenen Trainingsdaten.

Erwartete Vorkenntnisse

Empfohlen: - Python / NLP / Data Science - Annotationsstudien

Weiterführende Quellen

Nach Absprache mit dem Betreuer.

From:

<https://wiki.mi.uni-regensburg.de/> - **MI Wiki**

Permanent link:

<https://wiki.mi.uni-regensburg.de/arbeiten/absa-annotation>

Last update: **14.04.2024 15:58**

