
Natural Language Processing and Machine Learning Applications in the Social Sciences

Christoph Johannes Weißer

Abstract dissertation

This dissertation presents five independent research papers. The first three papers develop natural language processing methodologies. In the fourth and fifth paper, natural language processing and machine learning models are applied to address research questions in the social sciences. More precisely, the first paper develops a framework to evaluate topic models on short text data. The second paper provides a framework to detect specific and coherent topics in social media text data. The third paper proposes a classification rule for unsupervised document classification. The fourth paper uses natural language processing models to generate features that are subsequently incorporated in conflict prediction models. The fifth paper provides a big data application by using machine learning models for customer segmentation.

Zusammenfassung der Dissertation

Die Dissertation besteht aus fünf unabhängigen Forschungspapieren. Während die ersten drei Beiträge Natural Language Processing Methoden entwickeln, werden im vierten und fünften Beitrag Natural Language Processing und Machine Learning Modelle angewendet, um sozialwissenschaftliche Forschungsfragen zu untersuchen. Das erste Forschungspapier entwickelt ein Framework, mit dem Topicmodelle auf der Grundlage kurzer Texte evaluiert werden können. In der zweiten Studie wird ein Framework vorgestellt, mit dem spezifische und kohärente Topics in Textdaten aus sozialen Netzwerken geschätzt werden können. Im dritten Beitrag wird eine Klassifizierungsregel für die unüberwachte Klassifizierung von Dokumenten vorgeschlagen. Das vierte Forschungspapier verwendet Natural Language Processing Modelle, um Variablen für Konfliktvorhersagemodelle zu generieren. Im fünften Beitrag wird eine Big-Data-Anwendung vorgestellt, bei der Machine Learning Modelle zur Kundensegmentierung eingesetzt werden.