



Bachelor-, Master- und Doktorandenseminar
des Instituts für Informatik

Vergleich und Verbindung manueller und automatischer Verfahren zur Inhaltsanalyse

Britta Rannenberg, TU Clausthal

Diese Arbeit befasst sich mit der manuellen und der automatischen Inhaltsanalyse von Texten und deren Verbindungsmöglichkeiten. Bei der manuellen Analyse werden durch einen Menschen einzelne Textpassagen bestimmten Merkmalen, sogenannten Kategorien, zugeordnet. Diese Kategorien können grammatikalische Strukturen, wie z.B. Verben und Adjektive, aber auch Begriffe wie z.B. Arbeit, sein. Für jede Textstelle wird erfasst, ob sie der Kategorie zugeordnet wurde oder nicht.

Die manuelle Inhaltsanalyse ist zeitaufwendig, daher wurden Programme entwickelt, die dieses automatisch machen. Für eine automatische Analyse werden z.B. Empath und LIWC genutzt. Diese haben ein umfangreiches, vorgegebenes Kategoriensystem, mit dem die Texte analysiert werden.

Das Ergebnis der manuellen Inhaltsanalyse kann durch eine binäre logistische Regression dargestellt werden. Bei der binären logistischen Regression kann die abhängige Variable, die hier die manuelle Analyse ist, nur zwei Werte annehmen. Mit den unabhängigen Variablen, die hier das Ergebnis der automatischen Analyse sind, werden Wahrscheinlichkeiten berechnet. Mit diesen Wahrscheinlichkeiten lässt sich vorhersagen, ob eine Textstelle der manuellen Kategorie zugeordnet wird oder nicht.

Für jede manuelle Kategorie und jedes automatische Analyseprogramm muss ein eigenes Modell erstellt werden. Die korrekt vorhergesagten Werte in den einzelnen Kategorien sind unterschiedlich, daher muss immer abgewägt werden, wie hoch der prozentuale Anteil der korrekt vorhergesagten Werte sein muss. Es ist nicht möglich für alle manuellen Kategorien ein Regressionsmodell zu finden, da teilweise die Datengrundlage die Modellbildung nicht zulässt.

Mit der automatischen Inhaltsanalyse können Texte analysiert werden und anschließend kann mit der Regressionsmodell überprüft werden, ob dieser Text das in der manuellen Analyse untersuchte Merkmal beinhaltet. Die Analyse kann z.B. in Chats und Foren eingesetzt werden und dort als Schutzmaßnahme, zur Unterstützung der Nutzer oder für kommerzielle Zwecke dienen.

Dienstag, den 02.05.2017, 11 Uhr s.t. im
Besprechungsraum 106, IfI, Julius-Albert-Straße 4