

Was macht positiv wahrgenommene Kommunikation in Entwicklerteams aus? Eine statistische Analyse zur Identifikation von relevanten Metriken

Hintergrund

Die Sentimentanalyse gewinnt in Softwareprojekten eine immer größere Bedeutung, da sie eine Stimmungsanalyse der Kommunikation ermöglicht. Dabei wird die Sentimentanalyse auf verschiedene Arten schriftlicher Kommunikation angewandt: App-Reviews, Chatverläufe, Kommentare an Issues, ... Die Aussagekraft und die Güte der Ergebnisse hängt jedoch stark von den zugrundeliegenden Trainingsdaten der Klassifikationsalgorithmen ab. Dazu werden große Mengen an Aussagen von Personen als positiv, negativ und neutral klassifiziert. Der Algorithmus errechnet dann basierend auf den vorliegenden Daten Regeln, die ihm bei der Klassifikation neuer Sätze helfen.

Aufgabe

Im Rahmen dieser Arbeit soll analysiert werden, welche Eigenschaften von Sätzen neben den verwendeten Wörtern einen signifikanten Einfluss auf die Polarität des Satzes haben. Zu diesen (quantitativen) Eigenschaften gehören zum Beispiel die Länge des Satzes oder des Wortes, die Anzahl an Tippfehlern, und die Anzahl an Wörtern. Im Rahmen dieser Masterarbeit sollen weitere quantitative Eigenschaften hergeleitet werden, die dann auf ihren Einfluss auf die Klassifikation hin untersucht werden.

Diese Arbeit gliedert sich in die folgenden Schritte:

- 1) Einarbeiten in die Sentimentanalyse und die für diese Arbeit benötigten Methoden der Statistik
- 2) Auswahl und Vorbereitung eines geeigneten Datensatzes inkl. Extraktion der quantitativen Metriken
- 3) Einzelne Regressionsanalysen, um den Zusammenhang zwischen einzelnen Metriken und der Polarität zu untersuchen
- 4) Konstruktion eines logistischen Regressionsmodells, um auch Wechselwirkungen zwischen den Metriken zu berücksichtigen
- 5) (optional) Trainieren eines Klassifizierers nur unter Verwendung der als signifikant identifizierten Metriken, um zu überprüfen, wie aussagekräftig die alleinige Betrachtung dieser Metriken ist
- 6) Diskussion und Interpretation der Ergebnisse

Hinweis: Diese Arbeit ist so konzipiert, dass sie vollständig aus dem Home Office bearbeitet werden kann. Eine Verlängerung der Bearbeitungszeit wegen Corona ist nicht vorgesehen.

Organisatorisches

Betreuer: Jil Klünder
Prüfer: Kurt Schneider und Jil Klünder
Beginn: ab sofort möglich