IMSM
Institute for Interactive Marketing & Social Media

AUSSCHREIBUNG

Bachelorarbeit

STICHWÖRTER

- Soziale Medien
- Emoji-Analyse
- Word2Vec
- Empirische Analyse

THEMA: EIN EMOJI SAGT MEHR ALS TAUSEND WORTE 🤓

Social Media Plattformen sind mittlerweile eine der beliebtesten Datenquellen für die Ermittlung der Reputation einer Marke (Ott 2015), Erkennung von Shit-Storms (Alexander 2014) oder dem Vorhersagen von Verkaufszahlen (Dellarocas 2007). Jedoch haben sich sowohl die geposteten Inhalte als auch die Sprache mit der Zeit erheblich verändert (Pavalanathan 2016). So beinhaltet bereits der Großteil der Instagram-Kommentare sogenannte Emojis (wie 😂, 🍔 oder 🎃). Diese im Unicode-Standard spezifizierten Sonderzeichen verdrängen zusehends herkömmliche Wörter, bieten aber einige Vorteile: So sind diese sehr ausdrucksstark, von der Sprache weitgehend unabhängig und bestehen aus einer im Verhältnis zu herkömmlichen Textanalyse sehr kleinen Menge an wohldefinierten Zeichen (Novak 2015).

Um Informationen aus Emojis extrahieren zu können, bedarf es jedoch an aktuellen Text Mining Algorithmen. In der Analyse von Textdaten hat sich vor allem das word2vec-Modell und dessen Erweiterungen etabliert (z.B. Minarro-Giménez et.al 2014, Xue et.al 2014 oder Joshi et.al 2016). Bei diesen werden Worte in einen Vektorraum transformiert, wodurch man sowohl die Ähnlichkeit zwischen Wörtern ermitteln, als auch mit den Wortrepräsentationen direkt rechnen (z.B. Königin ≅ König – Mann + Frau) kann (Mikolov 2013). In wissenschaftlichen Arbeiten wurde dieser Ansatz bereits für Emoji-Beschreibungen (Eisner et.al 2016), nur Emojis (Barbieri 2016), als auch eine Mischung zwischen Emojis und Wörtern (Barbieri 2016) adaptiert.

Da mittlerweile Emojis in satzähnlichen Konstruktionen (Emoji-Sequenzen wie 😍😘😭😭) verwendet werden, soll im Rahmen dieser Bachelorarbeit ein Modell erstellt werden, welches ausschließlich auf diesen Emoji-Sequenzen basiert. Hier soll jedoch im Gegensatz zu bisherigen Forschungsarbeiten das word2vec Modell speziell für Emojis angepasst werden um auf die Eigenheiten dieser Zeichen entsprechend Rücksicht zu nehmen. Hierzu sollen aus einem

gegebenen Instagram Datensatz, der über 5 Millionen Kommentare enthält, Emoji-Sequenzen extrahiert und damit ein Vektorraum generiert werden. Die Qualität der Resultate soll in Folge in einem passenden Anwendungsbeispiel (z.B. Sentiment-Analyse) oder im direkten Vergleich mit Resultaten ähnlicher Forschungsarbeiten überprüft werden.

Da größere Datenmengen analysiert und empirisch ausgewertet werden sollen, ist für die Bearbeitung des Themas ein gewisses technisch-mathematisches Verständnis Voraussetzung (Idealerweise hat man die Data Science SBWL absolviert oder einschlägige Praktika vorzuweisen).

LITERATUR:

- **Mikolov T., Yih W. & Zweig G. (2013):** "Linguistic regularities in continuous space word representations." hlt-Naacl. Vol. 13.
- **Novak P. et al. (2015):** "Sentiment of emojis." PloS one 10.12. e0144296.
- **Eisner B. et al. (2016):** "emoji2vec: Learning emoji representations from their description." arXiv preprint arXiv:1609.08359.
- **Barbieri F., Ronzano F., & Saggion H. (2016):** "What does this Emoji Mean? A Vector Space Skip-Gram Model for Twitter Emojis." LREC.
- **Ott L & Theunissen P. (2015):** "Reputations at risk: Engagement during social media crises." Public Relations Review. 41.1. 97-102.
- **Dellarocas, C., Zhang, X. M., & Awad, N. F. (2007):** Exploring the value of online product reviews in forecasting sales: The case of motion pictures. Journal of Interactive marketing, 21(4), 23-45.
- **Alexander, D. E. (2014):** Social media in disaster risk reduction and crisis management. Science and Engineering Ethics, 20(3), 717-733.
- **Pavalanathan, U., & Eisenstein, J. (2016):** More emojis, less:) The competition for paralinguistic function in microblog writing. First Monday, 21(11).
- **Minarro-Giménez, J. A., Marin-Alonso, O., & Samwald, M. (2014):** Exploring the application of deep learning techniques on medical text corpora. Studies in health technology and informatics, 205, 584-588.
- **Xue, B., Fu, C., & Shaobin, Z. (2014, June):** A study on sentiment computing and classification of sina weibo with word2vec. In Big Data (BigData Congress), 2014 IEEE International Congress on (pp. 358-363).
- **Joshi, A., Tripathi, V., Patel, K., Bhattacharyya, P., & Carman, M. (2016):** Are Word Embedding-based Features Useful for Sarcasm Detection?. arXiv preprint arXiv:1610.00883.

BETREUER:

- DI Christian Hotz-Behofsits www.wu.ac.at/imsm/jobs/team/christian-hotz-behofsits/
- Univ.Prof.Dr. Nadia Abou Nabout https://www.wu.ac.at/imsm/jobs/team/abounabout

BEWERBUNGEN

Die Bewerbungen mit Lebenslauf und einer aktuellen Notenübersicht reichten Sie bitte an Christian Hotz-Behofsits (christian.hotz-behofsits@wu.ac.at).