

Multimodale Sentimentanalyse politischer Tweets

Ziehe, Stefan

stefan.ziehe@stud.uni-goettingen.de
GCDH, Universität Göttingen, Deutschland

Sporleder, Caroline

csporled@gwdg.de
GCDH, Universität Göttingen, Deutschland

Die automatische Analyse von Tweets mit politischem Inhalt kann Sozial- und Politikwissenschaftlern Aufschluss über die Prozesse politischer Meinungsbildung geben. Beiträge in den sozialen Medien spiegeln oft Tendenzen in der Zufriedenheit mit politischen Parteien wider und helfen umstrittene oder viel diskutierte Themen zu identifizieren. Auch das Kommunikationsverhalten verschiedener Gruppen lässt sich aus ihrer Interaktion bei Twitter analysieren (z.B. unterschiedliche Dominanz von Echokammereffekten (Colleoni et al., 2014) oder Verbreitung von Gerüchten und Fake News (Ma et al., 2018)).

Eine wichtige Rolle spielt dabei die Sentimentanalyse, die es erlaubt zu identifizieren, ob ein Akteur Zustimmung oder Ablehnung zu einem bestimmten Inhalt signalisiert. Für Textdaten lässt sich das Sentiment meist recht gut bestimmen. Tweets sind aufgrund ihrer Kürze jedoch zum einen oft schwer recht kryptisch, zum andern enthalten sie häufig weitere Materialien, insbesondere Bilder, die nennenswert zur Aussage beitragen. Z.B. ist der Text „und wieder ein neuer Morgen“ neutral formuliert, gewinnt aber eine negative Bedeutung, wenn er um eine Bild bereichert wird, das eine lange Autoschlange zeigt. Ebenso kann es sein, dass ein relativ neutrales Bild lediglich über den Text ein positives oder negatives Sentiment zugewiesen bekommt.

Die meisten bisher existierenden Sentimentanalyseverfahren beschränken sich auf die Verarbeitung entweder von Text- oder von Bilddaten. Modelle, die beide Modalitäten

berücksichtigen, sind noch vergleichsweise selten, können jedoch eine signifikant höhere Genauigkeit bei der Sentiment-Vorhersage erreichen als solche, die dies nicht tun (You et al. 2016). Fast alle multimodalen Verfahren nutzen eine Deep-Learning-Architektur. Solche Verfahren sind herkömmlichen Lernverfahren zwar oft überlegen, sie sind aber aufgrund der Vielzahl der möglichen Architekturen auch relativ schwer zu optimieren. Das Ziel dieser Arbeit ist es, verschiedene multimodale Sentimentanalyseverfahren und

-architekturen systematisch zu vergleichen und auf ihre Vor- und Nachteile hin zu untersuchen.

Das grundsätzliche Schema der Modelle orientiert sich am "Latent Multimodal Mixing" (Bruni et al. 2014); hierbei werden zunächst Text- und Bild-Features extrahiert, als Vektoren kodiert

und anschließend in einem dritten Schritt auf einen gemeinsamen (multimodalen) Vektorraum abgebildet (Fusion). Aus diesen Vektoren kann dann mit Methoden des maschinellen Lernens das Sentiment berechnet werden. Innerhalb dieses Schemas können beliebige und auch neuartige Kombinationen verschiedener Methoden zur Feature-Extraktion und Fusion verwendet werden. Hierfür gibt es unter anderem folgende Möglichkeiten:

- Texte können mit dem Doc2Vec-Verfahren auf einen latenten Vektorraum abgebildet werden. Dieses Verfahren erzielte in der Vergangenheit bereits gute Ergebnisse bei der Sentimentanalyse. (Le et al. 2014)
- Basierend auf einem existierenden Word-Embedding-Modell (z.B. GloVe (Pennington et al. 2014)) können die Word-Embeddings aller Wörter eines Textes auf verschiedene Arten zu einem Text-Embedding aggregiert werden (z.B. gewichteter Mittelwert, elementweises Minimum/Maximum). (De Boom et al. 2016)
- Für die Extraktion visueller Features können bereits existierende Deep Learning-Modelle zur Bild-Klassifikation in leicht modifizierter Form wiederverwendet werden. (Campos et al. 2017)
- Aus dem Farbhistogramm eines Bildes können statistische Features erster Ordnung berechnet werden.
- Der Fusionschritt besteht aus einer einfachen Verkettung der Text- und Bild-Vektoren. Zusätzlich kann auch eine affine Projektion auf einen latenten multimodalen Vektorraum gelernt werden. (Chen et al. 2017)

Die Datengrundlage für das Training der Modelle bilden manuell annotierte multimodale Tweets u.a. aus dem Photo Tweet Sentiment Benchmark (Borth et al. 2013), sowie das Columbia MVSO Image Sentiment Dataset (Dalmia et al. 2016). Aufgrund der unterschiedlichen Größe der Datensätze wird ein Transfer Learning-Ansatz verfolgt: Die Modelle werden zunächst auf den MVSO-Daten vortrainiert und anschließend mithilfe der Twitter-Daten feinadjustiert.

Erste Testergebnisse bestätigen, dass Modelle, die Text- und Bild-Features fusionieren, eine höhere Genauigkeit erreichen können als unimodale Modelle. Allerdings haben die bisher getesteten Modelle derzeit noch Schwierigkeiten damit, negatives Sentiment korrekt zu klassifizieren.

Bibliographie

You, Quanzeng / Luo, Jiebu / Jin, Hailin / Yang, Jianchao (2016): *Cross-modality Consistent Regression*

for Joint Visual-Textual Sentiment Analysis of Social Multimedia, in: The Ninth International Conference on Web Search and Data Mining, February 2016, San Francisco, CA, USA.

Bruni, Elia / Tran, Nam Khanh / Baroni, Marco (2014): *Multimodal distributional semantics*, in: Journal of Artificial Intelligence Research 49, 1 (Januar 2014), 1-47

Colleoni, Elanor/ Rozza, Alessandro / Arvidsson, Adam (2014): *Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data*, in: Journal of Communicatio 64 (2014) 317–332

De Boom, Cedric / Van Canneyt, Steven / Demeester, Thomas / Dhoedt, Bart (2016): *Representation learning for very short texts using weighted word embedding aggregation*, in: Pattern Recognition Letters 80, C (September 2016), 150-156.

Campos, Victor / Jou, Brendan / Giró-i-Nieto, Xavier (2017): *From Pixels to Sentiment: Fine-tuning CNNs for Visual Sentiment Prediction*, in: Image and Vision Computing 65 (September 2017), 15-22

Chen, Xingyue / Wang, Yunhong / Liu, Qingjie (2017): *Visual and Textual Sentiment Analysis Using Deep Fusion Convolutional Neural Networks*, in: IEEE International Conference on Image Processing (ICIP), Beijing, China

Dalmia, Vaidehi / Liu, Hongyi / Chang, Shih-Fu (2016): *Columbia MVSO Image Sentiment Dataset*, in: CoRR, abs/1611.04455

Borth, Damian / Ji, Rongrong / Chen, Tao / Breuel, Thomas / Chang, Shih-Fu (2013): *Large-scale Visual Sentiment Ontology and Detectors Using Adjective Noun Pairs*, in: The 21st ACM International Conference on Multimedia, October 2013, Barcelona, Spain

Le, Quoc / Mikolov, Tomas (2014): *Distributed Representations of Sentences and Documents*, in: The 31st International Conference on Machine Learning, June 2014, Beijing, China

Ma, Jing / Gao, Wei / Wong, Kam-Fai (2018) *Rumor Detection on Twitter with Tree-Structured Recursive Neural Networks*, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 1980–1989, Melbourne, Australia, July 15 - 20, 2018

Pennington, Jeffrey / Socher, Richard / Manning, Christopher (2014): *GloVe: Global Vectors for Word Representation*, in: Empirical Methods in Natural Language Processing (EMNLP), 1532-1543