

Semantische Klassifikation von Adverbien

Joachim Bingel
Institut für Computerlinguistik
Universität Heidelberg
bingel@cl.uni-heidelberg.de

Im Rahmen einer Bachelorarbeit wurde eine automatisierte typen- und tokenbasierte semantische Klassifizierung von Adverbien in acht semantische Klassen mit Hilfe von Vektorraummodellen und maschinellem Lernen vorgenommen.

Keine dem Autor bekannte semantische NLP-Ressource bietet eine Kategorisierung von Adverbien hinsichtlich ihrer temporalen, räumlichen oder anderen semantischen Beiträge. Dabei liegt der Nutzen solcher semantischer Informationen gerade für Aufgaben innerhalb der Informationsextraktion, aber auch jeglicher anderer komplexen NLP-Systeme, auf der Hand. Diese Arbeit ist ein Beitrag dazu, dieses Defizit zu überwinden.

Zunächst wird ein Datensatz von 13,960 Adverbtokens (224 Typen) im Korpus manuell nach Klassenzugehörigkeit annotiert, was als Grundlage für das datengetriebene Lernen und die Evaluation dient.

Das Klassifikationsproblem wird auf zwei verschiedene Weisen angegangen. Der erste Ansatz fußt auf der Distributionellen Hypothese, welche besagt, dass semantisch ähnliche Wörter in ähnlichen Kontexten auftreten. Daraus abgeleitet ist die Annahme, dass Adverbien, die derselben semantischen *Klasse* angehören, ebenfalls in ähnlichen Kontexten auftreten. Aufgrund dieser Annahme wird das Klassifikationsproblem mit Hilfe eines Vektorraummodells (VSM) gelöst.

Der zweite Ansatz ist auf der Beobachtung begründet, dass die verschiedenen semantischen Klassen unterschiedliches syntaktisches Verhalten aufweisen. Aus einer Baumbank werden so für jedes Adverbvorkommen acht syntaktische Features extrahiert und als Merkmale in einem maschinellen Lernverfahren verwendet, welches aus diesen syntaktischen Eigenschaften eines Adverbs die jeweilige semantische Klasse ableitet.

Die beiden Ansätze erreichen eine korrekte Klassifikation von 69,4% bzw. 68,2% der Testdaten.

Referenzen

- Conlon, S. P.-N. and Evens, M. (1992). *Can computers handle adverbs?* In *Proceedings of the 14th International Conference on Computational Linguistics (COLING)*, Nantes, France.
- Jacobson, S. (1964). *Adverbial positions in English*. PhD thesis, Uppsala Universitet.

- Hartung, M. and Frank, A. (2010). *A Structured Vector Space Model for Hidden Attribute Meaning in Adjective-Noun Phrases*. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC)*, Valletta, Malta.
- Padó, S. and Lapata, M. (2007). *Dependency-based construction of semantic space models*. In *Computational Linguistics*, 33(2):161–199.
- Turney, P. D. and Pantel, P. (2010). *From frequency to meaning: Vector space models of semantics*. In *Journal of Artificial Intelligence Research*, 37:141–188.