

DEUTSCH IM WANDEL

Dienstag, 11. März 2025, 15:45 Uhr

Automatische Bedeutungswandelerkennung in der Computerlinguistik: Ein Überblick

Dominik Schlechtweg (Stuttgart)

Die menschliche Sprache verändert sich im Laufe der Zeit. Dieser Wandel vollzieht sich auf verschiedenen sprachlichen Ebenen wie Grammatik, Laute oder Bedeutung. Die Untersuchung von Bedeutungsänderungen auf der Wortebene wird oft als 'lexikalischer Bedeutungswandel' bezeichnet und wird traditionell entweder aus einer onomasiologischen Perspektive mit der Frage angegangen, durch welche Wörter eine Bedeutung ausgedrückt werden kann, oder aus einer semasiologischen Perspektive mit der Frage, welche Bedeutungen ein Wort im Laufe der Zeit ausdrücken kann. In den letzten Jahren hat sich die Aufgabe der automatischen Erkennung von semasiologischem Bedeutungswandel aus Textdaten als ein eigenes Gebiet der Computerlinguistik unter dem Namen 'Bedeutungswandelerkennung' etabliert. Zwei Hauptfaktoren haben zu dieser Entwicklung beigetragen: (i) Der 'digital turn' in den Geisteswissenschaften hat große Mengen historischer Texte in digitaler Form verfügbar gemacht. (ii) Es wurden neue Computermodelle eingeführt, die semantische Aspekte von Wörtern allein aus Texten effizient erlernen.

Eine der Hauptmotivationen für die Arbeit an der Bedeutungswandelerkennung sind ihre Anwendungen in der historischen Semantik und der historischen Lexikographie, wo sich Forscher unter anderem mit der Klassifizierung von Wörtern in Kategorien des Bedeutungswandels beschäftigen. Automatische Methoden haben den Vorteil, dass sie Bedeutungswandel für große Datenmengen in kurzer Zeit vorhersagen und so den menschlichen Aufwand in den genannten Bereichen erheblich verringern können, während sie in der Lage sind, mehr Daten zu scannen und somit mehr semantische Veränderungen aufzudecken, die gleichzeitig weniger durch die von Forschern verwendeten Ad-hoc-Stichprobenkriterien beeinflusst werden. Andererseits können automatische Methoden auch schädlich sein, wenn ihre Vorhersagen fehlerhaft sind, d. h. sie können zahlreiche semantische Veränderungen übersehen oder Bedeutungswandel bei Wörtern erkennen, die keinen durchlaufen. Die auf diese Weise gewonnenen Ergebnisse könnten dann Forscher dazu verleiten, empirisch fehlerhafte Verallgemeinerungen über Bedeutungswandel vorzunehmen. Daher sollte man automatischen Methoden zur Erkennung von Bedeutungswandel erst dann vertrauen, wenn sie gründlich evaluiert wurden und ihre Vorhersagen nachweislich einen akzeptablen Grad an Korrektheit erreicht haben.

Der Vortrag wird einen Überblick über den aktuellen Stand der automatischen Bedeutungswandelerkennung geben unter Einbezug der am häufigsten verwendeten Datensätze, Evaluationsmethoden, Modelle und Anwendungen. Dabei werde ich insbesondere auf Vor- und Nachteile des herrschenden Evaluationsparadigmas in der Bedeutungswandelerkennung eingehen, welches auf paarweise annotierten Wortverwendungen aus zwei Zeitperioden beruht, und welches ich größtenteils in eigenen Arbeiten entwickelt habe. Ich werde aus diesem Paradigma abgeleitete Entwicklungen im Bereich der Modellierung durch Modelle semantischer Nähe beschreiben, welche auf menschlich annotierten Wortverwendungspaaren optimiert werden. Darüber hinaus werde ich auf die kürzliche Hinwendung des Forschungsbereichs zu wörterbuchabhängigen Evaluationsmethoden und Veränderungen durch die Einführung von großen Sprachmodellen (LLMs) eingehen. Zuletzt werde ich mögliche Anwendungen im Bereich der Lexikographie anhand von eigenen Forschungsarbeiten darstellen. Dazu gehören

unter anderem die automatische Erkennung unbekannter Wortbedeutungen durch optimierte Modelle der semantischen Nähe zwischen Wortverwendungen und die automatische Generierung von Bedeutungsdefinitionen für Wörterbucheinträge durch große Sprachmodelle.