

Gemeinsame Embeddings von Text und Graphen

Wissenschaftliche Fragestellung und Zielsetzung der Arbeit

Durch die wachsende Menge an Daten und die Abhängigkeit von ihnen wird die effiziente Verwaltung dieser Daten immer wichtiger. Eine Möglichkeit, mit großen Mengen komplexer Daten umzugehen, ist das semantische Datenmanagement. Ein wesentlicher Bestandteil des semantischen Datenmanagements ist die Erstellung semantischer Modelle, d. h. von Wissensrepräsentationen, die zum Verständnis und zur Interpretation des gegebenen



Datensatzes erforderlich sind. Methoden zur automatischen Erstellung erster Versionen semantischer Modelle konzentrieren sich auf Beschriftungen oder Werte innerhalb des Datensatzes. Eine andere Möglichkeit, ein semantisches Modell automatisch zu erstellen, die in diesem Zusammenhang noch kaum erforscht ist, ist die Verwendung von beschreibenden Texten der Datenquelle. Um Algorithmen des maschinellen Lernens aus der Welt der Semantik und der NLP-Welt zu kombinieren, ist es notwendig, eine gemeinsame Repräsentation von semantischen Modellen und textueller Beschreibung in Form von Einbettungen zu finden.

Vorgehensweise und Methodik

In dieser Arbeit soll die Möglichkeit des Trainings von Einbettungen, die Informationen aus Texten und Graphen kombinieren, evaluiert und praktisch umgesetzt werden. Anhand vorhandener Datensätze sollen die trainierten Einbettungen an ersten Machine-Learning-Aufgaben getestet werden. Zur wissenschaftlichen Methodik gehört zunächst die Einarbeitung in den aktuellen Stand der Technik mit einem Fokus auf die Domänen Graph-Embeddings sowie Text-Embeddings. Weiter sollen geeignete Daten zum Training sowie zur späteren Evaluierung entsprechender Embeddings identifiziert werden. In verschiedenen Experimenten werden die Embeddings trainiert und optimiert. Abschließend erfolgt ein Transfer zur Evaluierung eines bereits bestehenden Datensatzes des TMDT Lehrstuhls.

Benötigte Kenntnisse

- Programmierkenntnisse in Python
- Erfahrungen im Bereich des Maschinellen Lernens
- Grundlegende Kenntnisse in der Natürlichen Sprachverarbeitung (NLP)

Art der Arbeit

Bachelor-/Masterarbeit

Ansprechpartner

Andreas Burgdorf, M.Sc. | Tel.: +49 202 439 1047 | E-Mail: burgdorf@uni-wuppertal.de