

1 Einleitung

Bei dieser Aufgabe sollen Vektorrepräsentationen von Knoten innerhalb eines Graph erstellt werden. Diese Darstellung bietet die Möglichkeit als Eingabe für ML-basierte Verfahren zu fungieren oder effizient Ähnlichkeiten zwischen Knoten zu berechnen. Die Grundidee, die der Generierung der Embeddings zu Grunde liegt, ist die Erstellung von Knotensequenzen mithilfe eines Random Walks sowie dem Word2Vec Model.

2 Aufgabenstellung

Das Ziel dieser Arbeit ist die Erstellung von Embeddings für einen beliebigen Graphen unter Verwendung von DeepLearning4j und den bereits implementierten DeepWalk[2] Algorithmus. Die generierten Embeddings sollen verwendet werden, um eine Webapplikation zu realisieren, die die Ähnlichkeiten zwischen Knoten darstellt. Die Darstellung soll dabei verschiedene Aspekte betrachten, wie z.B. Darstellung aller Knoten im Vektorraum, Ähnlichkeit zwischen Knoten, Top K Knoten bzgl. eines Knotens. Als Graph soll ein Social Network Graph verwendet werden. Weiterhin soll die Ähnlichkeit zwischen Nutzern mit den Themen verglichen werden für die sie sich interessieren. Diesbezüglich sollen die Features für jeden Nutzer verwendet werden und mithilfe des ParagraphVector[1] Modell ebenfalls als Embedding dargestellt werden. Die Eingabe für das Modell sind Dokumente. In dem Fall eine Aneinanderreihung der Tags, so dass für jeden Nutzer ein Embedding generiert werden kann-

1. 1. Testat

- Import des Graphen in Graphstruktur von DeepLearning4J sowie Preprocessing der Features für die Verwendung von ParagraphVector
- Generierung der Embeddings für verschiedene Walk-Längen und Window-Größen sowie Paragraph Vektoren
- Persistierung

2. 2. Testat

- Erstellung einer Webapplikation mithilfe von D3¹
- Darstellung globaler Ansichten der Nutzer im dreidimensionalen Vektorraum bzgl. der Graph-Embeddings sowie der Paragraph-Embeddings.
 - Auswahl Knoten in einer Ansicht führt zur Selektion seines Embedding in der anderen Ansicht

¹<https://d3js.org/>

- Auswahl von zwei Nutzer und Berechnung der Ähnlichkeit für beide Embedding Arten
- Anzeige der Top-K ähnlichsten Nutzer bei Eingabe eines bestimmten Nutzers für eine Embedding Art. Berechnung der Ähnlichkeit für die andere Embedding-Art. Das Ziel ist die Überprüfung inwieweit Freundschaftsbeziehung und Interessen korrelieren und umgekehrt. Somit wäre es z.B. möglich, durch ähnliche Interessen neue Beziehungen vorzuschlagen. Ein weiteres Anwendungsbeispiel wären politische Diskussionen. Extreme Gruppierungen sind in der Regel mit Menschen befreundet, die einer ähnlichen Überzeugung sind. Dennoch lassen sich bei beiden Gruppen Kommentare oder Beiträge finden, die gleiche Probleme oder Ansichten identifiziert haben. Ziel wäre es, diesen Konsens darzustellen, um lösungsorientierte Diskussionen zu fördern und gegenseitige Provokationen zu vermeiden.

3 Datensatz

Als Graph können sie einen Twittergraphen verwenden. <https://snap.stanford.edu/data/egonets-Twitter.html>

Literatur

- [1] Q. V. Le and T. Mikolov. Distributed Representations of Sentences and Documents. 32, 2014.
- [2] B. Perozzi and S. Skiena. DeepWalk : Online Learning of Social Representations Categories and Subject Descriptors.