Big Data Praktikum: Attribute extraction from eCommerce product descriptions

Gruppe: Gregor Pfänder, Thilo Brummerloh

4. Mai 2021

1 Thema

Produktseiten von Onlineshops enthalten oft viele unzureichend strukturierte Produktbeschreibungen oder sogar gar keine Beschreibungen. Zum Preisvergleich des gleichen Produkts auf verschiedenen Webseiten muss allerdings bekannt sein um welche Ausprägung eines Produkt es sich handelt. So können Preise nicht nur von Produkten, sondern auch ihren Unterausprägungen, wie dem Speicherplatz oder der Farbe, verglichen werden.

2 Fragestellung

Es sollte möglich sein mithilfe von einem Computerprogramm diese Arbeit automatisch durchzuführen. Sollte ein Produkt auf einer WEbseite keine Produktinformationen haben entsteht ein schwierigeres Problem der Identifikation. Wenn allerdings Produktspezifikationen innerhalb eines Freitextes vorliegen, sollte es möglich sein daraus Wörter und Wortgruppen zu erkennen und einer Spezifikation zuzuweisen. Es soll eine Methode zur Entity Resolution ausgewählt werden. Damit

3 Daten

Je nach Methodenwahl werden unterschiedliche Daten benötigt. Zur Entity Resolution wäre nur ein Menge von ungeordneten Produktbeschreibungen nötig. Wenn allerdings zusätzlich ein Neural Network trainiert werden soll ist es auch nötig einen bereits gelabelten Datensatz zum training zu haben.

4 Methoden

4.1 Auswahl der Methode

Machine Learning oder irgendeine Regelbasierte Methode? Machine Learning

Supervised Learning, semi-supervised Learning[1], oder Unsupervised Learning? Semi, oder Unsupervised Learning

Klassische RNNs oder LSTMs? LSTM optimalerweise[2]S. 3

4.2 Trainieren des Modells

Literaturverzeichnis

Ghani, Rayid u. a. (2006). "Text mining for product attribute extraction". In: *ACM SIGKDD Explorations Newsletter* 8.1, S. 41–48. ISSN: 1931-0145. DOI: 10.1145/1147234.1147241.

Majumder, Bodhisattwa Prasad u. a. (o. D.). *Deep Recurrent Neural Networks for Product Attribute Extraction in eCommerce*. URL: http://arxiv.org/pdf/1803.11284v1.

[1] Ghani u. a. 2006.

[2] Majumder u. a. o. D.