

Automated Product Classification for a Recommendation-System

Thema:

Automated Product Classification for a Recommendation-System

Art:

MA

BetreuerIn:

Christian Wolff

BearbeiterIn:

Kay Brinkmann

ErstgutachterIn:

Christian Wolff

Status:

abgeschlossen

Stichworte:

semi-supervised learning, recommender system, nlp, text classification, product classification, Klassifikation, Produktklassifikation, data augmentation, information retrieval

angelegt:

2021-11-15

Anmeldung:

2022-03-22

Antrittsvortrag:

2021-11-29

Abschlussvortrag:

2022-05-09

Hintergrund

Die Arbeit wird in Verbindung mit einer Werkstudentenstelle bei 8SELECT erstellt. Das Unternehmen bietet ein Recommender-System für E-Shops an, in dem zusammenpassende Produkte in Form von Produkt-Sets mehrschichtig vorgestellt werden. Um sinnvoll Produktvorschläge anzubieten, werden die Produkte eines Shops anhand ihrer Eigenschaften, wie dem Namen oder der Produktbeschreibung attribuiert (z.B. als Kategorie „Hosen“ oder „T-Shirts“). Das Manuelle attribuiere von Produkten ist sehr umständlich und kostet viel Arbeit, insbesondere wenn es darum geht, neue Produktbereiche zu erschließen, für die es noch keine bestehenden Klassifizierungen gibt. Es ist deshalb von Vorteil, diesen Prozess so weit wie möglich zu automatisieren.

Zielsetzung der Arbeit

Die Produkte eines e-shops sollen anhand von den in Produktfeeds enthaltenen Daten automatisch auf Attribute klassifiziert werden. Die Klassifizierung soll dabei anhand von möglichst wenig Trainingsdaten durchgeführt werden, um die nötige manuelle Vorarbeit so weit wie möglich zu verringern. Als 2. Schritt sollen dann auch die Attribuiere von verschiedenen Shops normalisiert werden, indem ähnliche bzw. gleiche Klassen aus den verschiedenen Shops zusammengeführt

werden. Als Ergebnis steht dann eine Shop-übergreifende Klassifizierung (also Attribuierung) und ein Shop-übergreifender Klassifizierer für einen bestimmten Bereich (z.B. DIY).

Konkrete Aufgaben

- sinnvolle Trainings- und Testdaten finden
- Eigenschaften bestimmen, anhand derer die Produkte klassifiziert werden. Z.B.:
 - Text Similarity (z.B. Name, Produktbeschreibung)
 - Text Matching (z.B. Marke, Serie)
 - Ähnliche Werte (z.B. Preis, Gewicht)
 - Welche Eigenschaften sind befüllt/ nicht befüllt
 - Bilderkennung
- verschiedene Verfahren für die Klassifizierung testen und anhand ihrer Präzision vergleichen
 - verschiedene Arten an Klassifizierern
 - Verfahren zum Klassifizieren aufgrund von wenig Trainingsdaten (z.B.: Semi-supervised learning, Data Augmentation)
- Erstellen eines Klassifizierungssystems
- Klassifizieren der Produkte
- Zusammenführen von Klassifizierungen/ Klassifizierern

Erwartete Vorkenntnisse

- NLP
- IR
- ML / DL ...

Weiterführende Quellen

Kiang, M. Y., Ye, Q., Hao, Y., Chen, M., & Li, Y. (2011). A service-oriented analysis of online product classification methods. *Decision Support Systems*, 52(1), 28–39.

<https://doi.org/10.1016/j.dss.2011.05.001>

Ghani, R., Probst, K., Liu, Y., Crema, M., & Fano, A. (2006). Text mining for product attribute extraction. *ACM SIGKDD Explorations Newsletter*, 8(1), 41–48.

<https://doi.org/10.1145/1147234.1147241>

Xu, H., Liu, B., Shu, L., & Yu, P. (2019). Open-world Learning and Application to Product Classification. *The World Wide Web Conference on - WWW '19*, 3413–3419.

<https://doi.org/10.1145/3308558.3313644>

Abels, S., & Hahn, A. (2005). Automatic Classification and Re-Classification of Product Data in e-Business. 2005 Symposium on Applications and the Internet Workshops (SAINT 2005 Workshops), 350–353. <https://doi.org/10.1109/SAINTW.2005.1620046>

Li, B., & Han, L. (2013). Distance Weighted Cosine Similarity Measure for Text Classification. In H. Yin, K. Tang, Y. Gao, F. Klawonn, M. Lee, T. Weise, B. Li, & X. Yao (Hrsg.), *Intelligent Data Engineering and Automated Learning - IDEAL 2013* (Bd. 8206, S. 611–618). Springer Berlin Heidelberg.

https://doi.org/10.1007/978-3-642-41278-3_74

Ristoski, P., Petrovski, P., Mika, P., & Paulheim, H. (2018). A machine learning approach for product matching and categorization: Use case: Enriching product ads with semantic structured data. *Semantic Web*, 9(5), 707–728. <https://doi.org/10.3233/SW-180300>

From:

<https://wiki.mi.ur.de/> - **MI Wiki**

Permanent link:

https://wiki.mi.ur.de/arbeiten/automatisierte_klassifizierung_von_produkten_fuer_recommender-systeme

Last update: **19.12.2022 10:44**

