

Analyse der Informationsbedürfnisse im Bereich DIY in Conversational Search

Thema:

Analyse der Informationsbedürfnisse im Bereich DIY in Conversational Search

Art:

BA

BetreuerIn:

Alexander Frummet

BearbeiterIn:

Elena Frank

ErstgutachterIn:

David Elsweiler

ZweitgutachterIn:

Niels Henze

Status:

abgeschlossen

Stichworte:

Maschinelles Lernen, NLP, Informationsbedürfnisse, Conversational Search, Prompt Engineering, Datenannotation, GPT-3, Textklassifikation, Datenanalyse

angelegt:

2024-06-19

Anmeldung:

2024-06-14

Antrittsvortrag:

2024-07-08

Abgabe:

2024-08-29

Hintergrund

ChatBots, VoiceBots, oder Sprachassistenten - mittlerweile gibt es viele verschiedene Formen der Conversational AI, welche eine intuitive und natürliche Mensch-Maschine-Interaktion, besser gesagt Mensch-Maschine-Kommunikation ermöglichen und im alltäglichen Leben integriert sind. Diese Conversational Assistants können anhand einer dialogbasierten Suche - auch Conversational Search genannt - in den verschiedensten Bereichen weiterhelfen und erleichtern somit den Alltag. Doch um sicherzustellen, dass solche Systeme die Anfragen und somit Informationsbedürfnisse der Nutzer verstehen, ist es unerlässlich diese zu untersuchen und zu analysieren. Mit der Analyse von Conversational Search beschäftigten sich bereits viele Studien. Bunt et al. (2017) erstellten eine domänenunabhängige Taxonomie, anhand derer in einem Dialog entstehende Äußerungen in verschiedene Kategorien geordnet werden können. Frummet et al. (2022) untersuchten die Informationsbedürfnisse während des Kochens und analysierten diese detailliert, um dadurch eine hierarchische Taxonomie zu erstellen. Auch Choi et al. (2022) fokussierten sich auf den Kontext Kochen, sowie do-it-yourself (DIY) und sammelten auf diese Bereiche bezogene Daten mit Hilfe einer Wizard of Oz-Studie, bei der jeder Teilnehmer einer von zwei Gruppen zugeteilt war. Die dadurch erhobenen Dialoge bestehen aus Konversationen zwischen einem Studenten, der eine spezifische

Aufgabe auszuführen hatte und einem Lehrer, der über das erforderliche Fachwissen verfügte und Ersteren während des Prozesses anleitete. Da die Datenerhebung jedoch asynchron durchgeführt wurde, bestehen die Dialoge zwar immer abwechselnd aus Äußerungen eines Teilnehmers dieser beiden Gruppen, allerdings variieren die Gesprächspartner. Somit konnte jeder Teilnehmer einer Gruppe theoretisch mit verschiedenen Teilnehmern der anderen Gruppe interagieren, was die Komplexität der Daten erhöht. Choi et al. (2022) und auch die Alexa Prize TaskBot Challenge 2021, bei welcher verschiedene Teams Conversational Assistants für die Bereiche Kochen und DIY entwickelten, verdeutlichen die Verknüpfung dieser beiden prozeduralen Bereiche, worauf auch Frummet et al. (2022, S. 25) hindeuten. Doch bisher gibt es keine detaillierte Taxonomie für DIY, welche vergleichbar mit der von Frummet et al. wäre. Durch die genannten Studien wird deutlich, dass ein tiefgreifendes Verständnis der Informationsanforderungen, vor allem aber große Datenmengen nötig sind, damit Systeme auf die speziellen Anfragen beziehungsweise Informationsbedürfnisse der Nutzer trainiert und somit Conversational Assistants für diesen Bereich entwickelt werden können. Der Prozess der dafür notwendigen Datenannotation ist sehr kosten- und zeitintensiv, wobei Large Language Models (LLMs) diesem Problem entgegenwirken sollen (Tan et al., 2024). In dieser Arbeit wird eine kosten- und zeitsparende Methodik angewandt durch die mit Hilfe der Taxonomie von Frummet et al. (2022) und einem LLM Annotationsdaten im Bereich DIY gesammelt werden. Somit wird untersucht ob mit Hilfe einer Taxonomie-Vorlage und eines LLMs die Informationsbedürfnisse im Bereich DIY vorhergesagt

Zielsetzung der Arbeit

Mit Hilfe der Taxonomie von Frummet et al. (2022) als Richtlinie soll ein DIY-Datensatz zunächst manuell annotiert und dadurch eine Taxonomie für diesen Bereich erstellt werden. Da die Annotation von Daten sehr aufwändig ist, dient demnach eine Stichprobe des Datensatzes Wizard of Tasks (Choi et al., 2022) als Datengrundlage, welche zunächst 150 Äußerungen beinhaltet. Bei einer Studie von Figueroa et al. (2012), bei der die Performance eines Klassifikationsmodells bei steigender Trainingsdatengröße untersucht wurde, wurden 100-200 gelabelte Daten als ausreichende Startgröße genannt. Hierzu werden die Level 0 und 1 der Hierarchie berücksichtigt, da diese laut Frummet et al. (2022, S. 25) auf andere Domänen mit prozeduralen Aufgaben übertragbar und demnach generalisierbar sein könnten. Um mehr annotierte Daten zu sammeln, wird ein LLM eingesetzt. Hierfür werden mit dem von OpenAI entwickelten Sprachmodell GPT-3 (Brown et al., 2020) zwei verschiedene Methoden des Prompt Engineerings getestet, welche auf „new tasks without extensive training“ (Sahoo et al., 2024, S. 2) spezialisiert sind: Zero Shot Prompting und Few Shot Prompting. GPT-3 bewährte sich bisher bei der Annotation von Daten und gilt darin als kostengünstig und zeitsparend (Ding et al., 2023; Wang et al., 2021). Die dadurch entstehenden Label der Stichprobe sollen mit den manuell erstellten verglichen werden, um letztlich mit Hilfe der besseren Methode den kompletten Datensatz zu labeln und dadurch eine große Datenmenge zu schaffen. Anschließend soll eine vergleichende Analyse der beiden Taxonomien erfolgen, demnach also deren Gemeinsamkeiten und Unterschiede gefunden werden. Hierfür werden deskriptive Statistik und explorative Datenanalyse eingesetzt, um ein Verständnis für die grundlegenden Merkmale der Daten zu erlangen, wie beispielsweise die Verteilung der einzelnen Kategorien, sowie weitere Muster zu erkennen. Zudem wird mit Hilfe der inferentiellen Statistik geprüft ob die beobachteten Ähnlichkeiten oder Unterschiede statistisch signifikant sind. Für die Annotation mit GPT-3 wird auf die OpenAI API2 zugegriffen, die Bibliothek pandas (McKinney, 2010) wird verwendet, um Excel-Dateien zu verarbeiten, in denen die Daten der Dialoge mit den zugehörigen Annotationen vorliegen. Zusätzlich werden die Bibliotheken scikit-learn (Pedregosa et al., 2011), Matplotlib (Hunter, 2007) und die darauf aufbauende Datenvisualisierungsbibliothek seaborn (Waskom, 2021) für die statistischen Analysen und deren Visualisierungen einbezogen.

Konkrete Aufgaben

- Aufbereitung der Literatur
- Manuelle Annotation der Stichprobe des Datensatzes
- Anwendung von Zero- und Few-Shot-Prompting mit GPT-3 auf der Stichprobe
- Vergleich der entstehenden Label mit manuell annotierten Daten
- Annotation des gesamten Datensatzes mit Hilfe der besseren Prompting-Methode
- Vergleichende Analyse der entstehenden Taxonomie mit der von Frummet et al. (2022) durch deskriptive und inferentielle Statistik sowie explorative Datenanalyse

Erwartete Vorkenntnisse

- Kenntnisse im Natural Language Processing
- Kenntnisse im Machine Learning
- Kenntnisse in statistischer Datenanalyse
- Programmierkenntnisse in Python/BD

Weiterführende Quellen

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). Language Models are Few-Shot Learners (arXiv:2005.14165). arXiv. <http://arxiv.org/abs/2005.14165>

Bunt, H., Petukhova, V., Traum, D., & Alexandersson, J. (2017). Dialogue Act Annotation with the ISO 24617-2 Standard. In D. A. Dahl (Hrsg.), *Multimodal Interaction with W3C Standards* (S. 109–135). Springer International Publishing. https://doi.org/10.1007/978-3-319-42816-1_6

Choi, J. I., Kuzi, S., Vedula, N., Zhao, J., Castellucci, G., Collins, M., Malmasi, S., Rokhlenko, O., & Agichtein, E. (2022). Wizard of Tasks: A Novel Conversational Dataset for Solving Real-World Tasks in Conversational Settings. In N. Calzolari, C.-R. Huang, H. Kim, J. Pustejovsky, L. Wanner, K.-S. Choi, P.-M. Ryu, H.-H. Chen, L. Donatelli, H. Ji, S. Kurohashi, P. Paggio, N. Xue, S. Kim, Y. Hahm, Z. He, T. K. Lee, E. Santus, F. Bond, & S.-H. Na (Hrsg.), *Proceedings of the 29th International Conference on Computational Linguistics* (S. 3514–3529). International Committee on Computational Linguistics. <https://aclanthology.org/2022.coling-1.310>

Ding, B., Qin, C., Liu, L., Chia, Y. K., Joty, S., Li, B., & Bing, L. (2023). Is GPT-3 a Good Data Annotator? (arXiv:2212.10450). arXiv. <http://arxiv.org/abs/2212.10450>

Figueroa, R. L., Zeng-Treitler, Q., Kandula, S., & Ngo, L. H. (2012). Predicting sample size required for classification performance. *BMC Medical Informatics and Decision Making*, 12(1), 8. <https://doi.org/10.1186/1472-6947-12-8>

Frummet, A., Elweiler, D., & Ludwig, B. (2022). “What Can I Cook with these Ingredients?”—Understanding Cooking-Related Information Needs in Conversational Search. *ACM Transactions on Information Systems*, 40(4), 1–32. <https://doi.org/10.1145/3498330>

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 9(3), 90–95. <https://doi.org/10.1109/MCSE.2007.55>

McKinney, W. (2010). Data Structures for Statistical Computing in Python. 56–61. <https://doi.org/10.25080/Majora-92bf1922-00a>

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *The Journal of Machine Learning Research*, 12, 2825–2830.

Sahoo, P., Singh, A. K., Saha, S., Jain, V., Mondal, S., & Chadha, A. (2024). A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications (arXiv:2402.07927). arXiv. <http://arxiv.org/abs/2402.07927>

Tan, Z., Beigi, A., Wang, S., Guo, R., Bhattacharjee, A., Jiang, B.,

Karami, M., Li, J., Cheng, L., & Liu, H. (2024). Large Language Models for Data Annotation: A Survey (arXiv:2402.13446). arXiv. <https://doi.org/10.48550/arXiv.2402.13446> Wang, S., Liu, Y., Xu, Y., Zhu, C., & Zeng, M. (2021). Want To Reduce Labeling Cost? GPT-3 Can Help (arXiv:2108.13487). arXiv. <http://arxiv.org/abs/2108.13487> Waskom, M. (2021). seaborn: Statistical data visualization. Journal of Open Source Software, 6(60), 3021. <https://doi.org/10.21105/joss.03021>

From:
<https://wiki.mi.uni-regensburg.de/> - MI Wiki

Permanent link:
https://wiki.mi.uni-regensburg.de/arbeiten/analyse_der_informationsbedu%CC%88rfnisse_im_bereich_diy_in_conversational_search

Last update: **18.11.2024 09:50**

