

Evaluating Pretrained Transformer Models for Entity Linking in Task-Oriented Dialog

Sai Muralidhar Jayanthi, Varsha Embar, Karthik Raghunathan

MindMeld, Cisco Systems

{saijayan, vembar, ktick}@cisco.com

A Appendices

A.1 Dataset Curation

To conduct a comprehensive evaluation, we curate datasets for entity linking spanning the following matching criterion– semantic, syntactic, short-forms, numerals and phonetics. We utilize both in-house as well as publicly available datasets in our evaluation.

- MM_BP: MindMeld Blueprints¹ spanning the domains of food ordering, grocery shopping, video streaming, human resources assistant and banking assistant chatbots. These datasets reflect real-world use-cases with custom entities which are manually annotated by human experts.
- MM_BP-WORD and MM_BP-CHAR: We also curate misspelled versions of the above data to understand the robustness of these models against spelling errors which are common user errors when interacting with a chatbot. We use Neuspell’s (Jayanthi et al., 2020) word- and character-noising models with default parameter settings to inject misspelling in our test queries. As only a subset of all test queries could be noised, for the misspelling datasets, we benchmark the different models with only the noise-able subset of test queries before and after noising.
- COMPLWQ and MKQA: To broaden the scope of evaluation to a larger scale, we repurpose Question-Answering (QA) datasets such as ComplexWebQuestions (Talmor and Berant, 2018) and MKQA (Longpre et al., 2020) for this task. For each answer, we select a subset of answer span aliases with some filtering heuristics and use them as our test queries.

- ACRI: To construct abbreviations and short-forms type matching, we use the Acronym Identification (ACRI) dataset (Veyseh et al., 2020). The entities obtained from these datasets closely resemble those from free-form and spoken language texts.
- ASR-MIS: Most dialog systems utilize 3rd party speech recognition systems (ASR) that often mistranscribe uncommon words or entities. In such cases, these EL systems need to correct for errors that are beyond textual variations, by utilizing phonetics. To test the phonetic matching capabilities of the models, we include an in-house dataset of ASR mis-transcriptions for person names. To construct this dataset, we randomly sampled 3K names from the directory of an organization and use Google Speech-to-Text² to collect 10-best ASR transcripts for each sample.

To create the manually annotated sets – SEMANTIC, SYNTACTIC, SHORT-FORMS, NUMERALS and PHONETICS – we first split the pool of all curated datasets into 70-30% ratio at random. We then retain the 70% split for training purposes and use the rest 30% to sub-sample queries for manual annotation. Retaining a training split helps to study the impact of using aliases in entity linking process on each model (Figure 2).

Due to our semi-automated curation strategies, a small portion of test queries might end up with incorrect labels. We plan to address this artifact in future work.

¹<https://github.com/CiscoDevNet/mindmeld-blueprints>

²<https://cloud.google.com/speech-to-text/>

A.2 Experimental Setup

MindMeld’s Entity Resolution: To conduct all our experiments related to Entity Linking, we use the publicly available open-source library MindMeld³. MindMeld is a Conversational AI platform for building production-quality conversational applications. It is a Python-based machine learning framework which encompasses all of the algorithms and utilities required for this purpose.

Due to our deliberate choice to evaluate Entity Linking step decoupled from Entity Recognition step (to better understand the efficacies of PTMs across different matching criterion), MindMeld’s Entity Resolution (ER) module⁴ is a perfect choice for our experimental setup.

The MindMeld ER module provides an easy interface to index Knowledge Bases (KB) with an option of using frameworks like ElasticSearch⁵ for quick retrieval. Various models to be used to extract representations (n-grams or dense representations) can be configured in the config file along with the similarity metric to be used (like BM25, cosine similarity).

A close alternate choice is DeepPavlov’s (Burtsev et al., 2018) Entity Resolution module⁶.

Aliases in Knowledge Bases: In experiments wherein KBs contain aliases, all the aliases’ vector representations are treated independently and matching to any alias implies matching to its corresponding canonical title. Meaning, the aliases’ representations are only used to obtain the matching scores but once obtained, a KB title is assigned the maximum score amongst the title’s score as well as its aliases’.

References

Mikhail Burtsev, Alexander Seliverstov, Rafael Airapetyan, Mikhail Arkhipov, Dilyara Baymurzina, Nickolay Bushkov, Olga Gureenkova, Taras Khakhulin, Yuri Kuratov, Denis Kuznetsov, Alexey Litinsky, Varvara Logacheva, Alexey Lymar, Valentin Malykh, Maxim Petrov, Vadim Polulyakh, Leonid Pugachev, Alexey Sorokin, Maria Vikhрева, and Marat Zaynutdinov. 2018. *DeepPavlov: Open-source library for dialogue systems*. In *Proceedings of ACL 2018, System Demonstrations*,

pages 122–127, Melbourne, Australia. Association for Computational Linguistics.

Sai Muralidhar Jayanthi, Danish Pruthi, and Graham Neubig. 2020. *NeuSpell: A neural spelling correction toolkit*. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 158–164, Online. Association for Computational Linguistics.

Shayne Longpre, Yi Lu, and Joachim Daiber. 2020. *Mkqa: A linguistically diverse benchmark for multilingual open domain question answering*.

A. Talmor and J. Berant. 2018. The web as a knowledge-base for answering complex questions. In *North American Association for Computational Linguistics (NAACL)*.

Amir Pouran Ben Veyseh, Franck Dernoncourt, Quan Hung Tran, and Thien Huu Nguyen. 2020. *What Does This Acronym Mean? Introducing a New Dataset for Acronym Identification and Disambiguation*. In *Proceedings of COLING*.

³<https://github.com/cisco/mindmeld>

⁴https://www.mindmeld.com/docs/entity_resolver.html

⁵<https://www.elastic.co/elasticsearch/>

⁶http://docs.deeppavlov.ai/models/entity_linking.html