reflect fine-grained characters of entities, and have rich topic information, which can be used to build connection between entities and words.

Text Segmentation. The goal of segmentation is to divide a short text into a sequence of meaningful components. Most short text process approaches like [22] adopt the naive longest-cover method for text segmentation, that is, it prefers the longest terms in a given vocabulary. The longest-cover method does not consider the global structure of the text and cannot handle the problem of overlapping, like "new york times square". Some models [21, 38] are used to segment the text for a global optimization objective with statistics prior. But they are unable to understand the semantics of a short text, and fails in cases like "reservation hotel california". Thus, a good approach to short text segmentation must take semantics into consideration.

7 CONCLUSION

We propose a comprehensive approach for short text entity recognition and linking. Our approach is capable of detecting local topics in short texts and linking entities to knowledge base with extremely little context. We introduce concepts of entities as explicit fine-grained topics to solve the sparsity and noisy problem of short text. We incorporate our entity linking module into the text segmentation algorithm and solve the problem of unawareness of the semantic, which improve the text segmentation and entity recognition. Experiments have proved that our approach has the state-of-the-art performance in the short-text entity linking, text segmentation and entity recognition.

REFERENCES

[1] Nitish Aggarwal and Paul Buitelaar. 2014. Wikipedia-based Distributional Semantics for Entity Relatedness.
[2] Baidu. 2017. Baidu Entity Annotation. https://ai.baidu.com/tech/cognitive/entity_annotation. Accessed January 4, 2018.
[3] Denilson Barbosa. 2017. Robust Named Entity Disambiguation with Random Walks.
[4] Sumit Bhatia and Anshu Jain. 2016. Context Sensitive Entity Linking of Search Queries in Enterprise Knowledge Graphs. In *ESWC*.
[5] Roi Blanco, Giuseppe Ottaviano, and Edgar Meij. 2015. Fast and Space-Efficient Entity Linking for Queries. In *WSDM*.
[6] Xiao Cheng and Dan Roth. 2013. Relational Inference for Wikification. In *EMNLP*.
[7] Leon Derczynski, Diana Maynard, Guiseppe Rizzo, Marieke van Erp, Genevieve Gorrell, Raphaël Troncy, Johann Petrak, and Kalina Bontcheva. 2015. Analysis of named entity recognition and linking for tweets. *Inf. Process. Manage.* 51 (2015), 32–49.
[8] Thomas Emerson. 2005. The Second International Chinese Word Segmentation Bakeoff. In *SIGHAN@IJCNLP 2005*.
[9] Yansong Feng, Zhe Han, and Kun Zhang. 2015. Overview of the NLPCC 2015 Shared : Entity Recognition and Linking in Search Queries. In *NLPCC*.
[10] Paolo Ferragina and Ugo Scaiella. 2010. TAGME: on-the-fly annotation of short text fragments (by wikipedia entities). In *CIKM*.
[11] Octavian-Eugen Ganea and Thomas Hofmann. 2017. Deep Joint Entity Disambiguation with Local Neural Attention. In *EMNLP*.
[12] Stephen Guo, Ming-Wei Chang, and Emre Kiciman. 2013. To Link or Not to Link? A Study on End-to-End Tweet Entity Linking. In *HLT-NAACL*.
[13] Nitish Gupta, Sameer Singh, and Dan Roth. 2017. Entity Linking via Joint Encoding of Types, Descriptions, and Context. In *EMNLP*.
[14] Xianpei Han and Le Sun. 2012. An Entity-Topic Model for Entity Linking. In *EMNLP-CoNLL*.
[15] Zhengyan He, Shujie Liu, Mu Li, Ming Zhou, Longkai Zhang, and Houfeng Wang. 2013. Learning Entity Representation for Entity Disambiguation. In *ACL*.
[16] Johannes Hoffart, Stephan Seufert, Dat Ba Nguyen, Martin Theobald, and Gerhard Weikum. 2012. KORE: keyphrase overlap relatedness for entity disambiguation. In *CIKM*.
[17] Johannes Hoffart, Fabian M. Suchanek, Klaus Berberich, and Gerhard Weikum. 2013. YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. *Artif. Intell.* 194 (2013), 28–61.
[18] Johannes Hoffart, Mohamed Amir Yosef, Ilaria Bordino, Hagen Fürstenau, Manfred Pinkal, Marc Spaniol, Bilyana Taneva, Stefan Thater, and Gerhard Weikum. 2011. Robust Disambiguation of Named Entities in Text. In *EMNLP*.
[19] Wen Hua, Zhongyuan Wang, Haixun Wang, Kai Zheng, and Xiaofang Zhou. 2015. Short text understanding through lexical-semantic analysis. *2015 IEEE 31st International Conference on Data Engineering* (2015), 495–506.
[20] isnowfy. 2013-2014. SnowNLP. https://github.com/isnowfy/snownlp. Accessed May 4, 2018.
[21] Sun Junyi. 2013. jieba segmentation. https://github.com/fxsjy/jieba. Accessed May 4, 2018.
[22] Dongwoo Kim, Haixun Wang, and Alice H. Oh. 2013. Context-Dependent Conceptualization. In *IJCAI*.
[23] Nevena Lazic, Amarnag Subramanya, Michael Ringgaard, and Fernando Pereira. 2015. Plato: A Selective Context Model for Entity Resolution. *TACL* 3 (2015), 503–515.
[24] Xiaohua Liu, Yitong Li, Haocheng Wu, Ming Zhou, Furu Wei, and Yi Lu. 2013. Entity Linking for Tweets. In *ACL*.
[25] Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. 2014. The Stanford CoreNLP Natural Language Processing Toolkit. In *ACL*.
[26] Tomas Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In *NIPS*.
[27] David N. Milne and Ian H. Witten. 2008. Learning to link with wikipedia. In *CIKM*.
[28] Erwan Moreau, François Yvon, and Olivier Cappé. 2008. Robust Similarity Measures for Named Entities Matching. In *COLING*.
[29] Andrea Moro, Alessandro Raganato, and Roberto Navigli. 2014. Entity Linking meets Word Sense Disambiguation: a Unified Approach. *TACL* 2 (2014), 231–244.
[30] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global Vectors for Word Representation. In *EMNLP*.
[31] Jonathan Raiman and Olivier Raiman. 2018. DeepType: Multilingual Entity Linking by Neural Type System Evolution. *CoRR* abs/1802.01021 (2018).
[32] Lev-Arie Ratinov, Dan Roth, Doug Downey, and Mike Anderson. 2011. Local and Global Algorithms for Disambiguation to Wikipedia. In *ACL*.
[33] Stephen E Robertson, Steve Walker, Susan Jones, Micheline M Hancock-Beaulieu, Mike Gatford, et al. 1995. Okapi at TREC-3. *Nist Special Publication Sp* 109 (1995), 109.
[34] Denis Savenkov and Eugene Agichtein. 2016. When a Knowledge Base Is Not Enough: Question Answering over Knowledge Bases with External Text Data. In *SIGIR*.
[35] Valentin I. Spitkovsky and Angel X. Chang. 2012. A Cross-Lingual Dictionary for English Wikipedia Concepts. In *LREC*.
[36] Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. Yago: a core of semantic knowledge. In *WWW*.
[37] Shuyan Tech. 2018. CNProbase concept api. http://shuyantech.com/api/cnprobase/concept. Accessed May 22, 2018.
[38] Kun Wang, Chengqing Zong, and Keh-Yih Su. 2009. Which is More Suitable for Chinese Word Segmentation, the Generative Model or the Discriminative One?. In *PACLIC*.
[39] Zhongyuan Wang, Haixun Wang, and Zhirui Hu. 2014. Head, modifier, and constraint detection in short texts. *2014 IEEE 30th International Conference on Data Engineering* (2014), 280–291.
[40] Zhongyuan Wang, Kejun Zhao, Haixun Wang, Xiaofeng Meng, and Ji-Rong Wen. 2015. Query Understanding through Knowledge-Based Conceptualization. In *IJCAI*.
[41] Bo Xu, Yong Xu, Jiaqing Liang, Chenhao Xie, Bin Liang, Wanyun Cui, and Yanghua Xiao. 2017. CN-DBpedia: A Never-Ending Chinese Knowledge Extraction System. In *IEA/AIE*.
[42] Ikuya Yamada, Hiroyuki Shindo, Hideaki Takeda, and Yoshiyasu Takefuji. 2016. Joint Learning of the Embedding of Words and Entities for Named Entity Disambiguation. In *CoNLL*.