IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 19, Issue 3, Ver. V (May - June 2017), PP 14-15
www.iosrjournals.org

SSL-QA: Analysis of Semi-Supervised Learning for Question-Answering

Parth Patel[1], Jignesh Prajapati[2]
[1, 2] *Assistant Professor, Vadodara Institute of Engineering, India*

Abstract*: Open domain natural language question answering (QA) is a process of automatically finding answers to questions searching collections of text files. Question answering (QA) is a long-standing challenge in NLP, and the community has introduced several paradigms and datasets for the task over the past few years. These patterns differ from each other in the type of questions and answers and the size of the training data, from a few hundreds to millions of examples. Context-aware QA paradigm and two most notable types of supervisions are coarse sentence-level and fine-grained span-level. In this paper we analyse different intensive researches in semi-supervised learning for question-answering.*

I. Introduction

There are many researches done in this area such as a graph-based semi-supervised learning for the question answering (QA) task for ranking candidate sentences. Using textual entailment analysis, we obtain entailment scores between a natural language question posed by the user and the candidate sentences returned from search engine [1]. The textual deduction between two sentences is assessed via features representing high-level attributes of the entailment problem such as sentence structure matching, question-type named-entity matching based on a question-classifier, etc. A SSL to demonstrate that utilization of more unlabeled data points can improve the answer-ranking task of QA. The graph for labeled and unlabeled data using match scores of textual entailment features as similarity weights between data points [1]. A summarization method applied on the graph to make the computations feasible on large datasets.

II. A Semi-Supervised Learning Approach TO Why-Question Answering

Why-question answering method is a semi-supervised learning method for improving to generate training data from causal relations in texts such as "[Tsunamis are generated]$_{effect}$ because [the ocean's water mass is dis-placed by an earthquake]$_{cause[2]}$." A naive method for the generation would be to make a question-answer pair by simply converting the effect part of the causal relations into a why-question, like "Why are tsunamis generated?" from the above example, and using the source text of the causal relations as an answer. However, in our preliminary experiments, this naive method actually failed to improve the why-QA performance[2]. The main reason was that the machine-generated questions were often incomprehensible like "Why does (it) happen?", and that the system suffered from over fitting to the results of our automatic causality recognizer. Hence, a novel method that effectively filters out incomprehensible questions and retrieves from texts answers that are likely to be paraphrases of a given causal relation. Through a series of experiments, this approach significantly improved the precision of the top answer by 8% over the current state-of-the-art system for Japanese why-QA. A novel approach to why-QA, which is semi-supervised learning that, exploits automatically generated training data for why-QA, a non-trivial task, to generate training data for why-QA using causal relations in texts[2]. This method generates comprehensible questions from causal relations and retrieves from web texts answers to the questions, which are likely to be paraphrases of a given causal relation, using our baseline why-QA system and vocabulary overlap between answers and causal relations. These paraphrases of a given causal relation in the retrieved answers allow why-QA systems to learn a wide range of causality expression patterns and to recognize such causality expressions as candidates for answers to why-questions. This method achieved 8% improvement in precision at the top answer over the current state-of-the-art system for Japanese why-QA, which was actually used as a starting point for our semi-supervised learning.

QA through Transfer Learning from Large Fine-grained Supervision Data

In this approach, a task of Question-Answering can crucially benefit from the transfer learning of models trained on a different large, fine-grained QA dataset. We achieve the state of the art in two well-studied QA datasets, WikiQA and SemEval-2016 (Task 3A), through a basic transfer learning technique from SQuAD also finer supervision provides effective advice for learning lexical and syntactic information than coarser supervision, through quantitative results and visual analysis[3]. A similar transfer learning procedure achieves the state of the art on an entailment task. To show state-of-the-art results on WikiQA and SemEval-2016 (Task 3A) as well as an entailment task, SICK, outperforming previous results by 8%, 1%, and 2%, respectively[3].

Question answering with sentence-level super-vision can greatly benefit from standard transfer learning of a question answering model trained on a large, span-level supervision. So, such transfer learning can be applicable in other NLP tasks such as textual entailment.

Semi-Supervised QA with Generative Domain-Adaptive Nets

In this approach, the problem of semi-supervised question answering deploying unlabeled text to boost the performance of question answering models, a novel training framework, the Generative Domain-Adaptive Nets[4]. On the basis of reinforcement learning, develop novel domain adaptation algorithms, to reduce the discrepancy between the model-generated data distribution and the human-generated data distribution and this model to generate questions based on the unlabeled text, and combine model-generated questions with human-generated questions for training question answering models[4]. In serious and challenging problem, semi-supervised question answering, neural framework called Generative Domain-Adaptive Nets, which incorporate domain adaptation techniques in combination with generative models for semi-supervised learning. Empirically, this approach leads to substantial improvements over supervised learning models and outperforms several strong base-lines including GANs and dual.

III. Conclusions AND Discussions

In this paper, we study different semi supervised learning for question answering and analysis for the same. In a graph-based SSL algorithm to improve the performance of QA task by exploiting unlabeled entailment relations between affirmed question and candidate sentence pairs. A semantic and syntactic feature for textual entailment analysis has individually shown to improve the performance of the QA compared to the baseline. New graph representation for SSL that can represent textual entailment relations while embedding different question structures. Summarization on graph-based SSL can improve the QA task performance when more unlabeled data is used to learn the classifier model. We show that semi-supervise learning algorithm for why question-answering, QA through transfer learning, QA with generative domain adaptive nets.

References

[1]. Asli Celikyilmaz, Marcus Thint, Zhiheng Huang, A Graph-based Semi-Supervised Learning for Question-Answering. Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP- 2009.
[2]. Jong-Hoon Oh, Kentaro Torisawa, Chikara Hashimoto, Ryu Iida, Masahiro Tanaka, Julien Kloetzer, A Semi-Supervised Learning Approach to Why-Question Answering. Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence-2016.
[3]. Zhilin Yang, Junjie Hu, Ruslan Salakhutdinov, William W. Cohen, Semi-Supervised QA with Generative Domain-Adaptive Nets, arXiv -2017.
[4]. Jinxiu Chen, Donghong Ji, C. Lim Tan, and Zhengyu Niu. 2006. Relation extraction using label propaga-tion based semi-supervised learning. In Proceedings of the ACL-2006.
[5]. Charles L.A. Clarke, Gordon V. Cormack, R. Thomas Lynam, and Egidio L. Terra. 2006. Question an-swering by passage selection. In In: Advances in open domain question answering, Strzalkowski, and Harabagiu (Eds.), pages 259–283. Springer.
[6]. Oliver Delalleau, Yoshua Bengio, and Nicolas Le Roux. 2006. Large-scale algorithms. In In: Semi-Supervised Learning, pages 333–341. MIT Press.
[7]. Sandra Harabagiu and Andrew Hickl. 2006. Methods for using textual entailment in open-domain ques-tion answering. In In Proc. of ACL-2006, pages 905–912.
[8]. Zhiheng Huang, Marcus Thint, and Zengchang Qin. 2008. Question classification using headwords and their hypernyms. In Proceedings of the Conference on Empirical Methods in Natural Language Pro-cessing (EMNLP-08), pages 927–936.
[9]. Dan Klein and Christopher D. Manning. 2003. Ac-curate unlexicalized parsing. In Proceedings of the 41st Meeting of the ACL-2003, pages 423–430.
[10]. Michael Lesk. 1988. They said true things, but called them by wrong names - vocabulary problems in re-trieval systems. In In Proc. 4th Annual Conference of the University of Waterloo Centre for the New OED.
[11]. Rong Liu, Jianzhong Zhou, and Ming Liu. 2006. A graph-based semi-supervised learning algorithm for web page classification. In Proc. Sixth Int. Conf. on Intelligent Systems Design and Applications.
[12]. George Miller. 1995. Wordnet: A lexical database for English. In Communications of the ACL-1995.
[13]. Zheng-Yu Niu, Dong-Hong Ji, and Chew-Lim Tan. 2005. Word sense disambiguation using labeled propagation based semi-supervised learning. In
[14]. Proceedings of the ACL-2005.
[15]. Jahna Otterbacher, Gunes Erkan, and R. Radev Dragomir. 2009. Biased lexrank: passage retrieval using random walks with question-based priors. In-formation Processing and Management, 45:42–54.
[16]. Eric W. Prager, John M.and Brown, Dragomir Radev, and Krzysztof Czuba. 2000. One search engine or two for question-answering. In Proc. 9th Text RE-trieval conference.