CS 5/7322 Spring 2023

Project

The goal of the project is to allow student to explore various problems in Natural Language Processing.

I will provide a list of projects below. Each project will comes with a brief description, and some idea of what is expected. It will also come with a few papers, in which you should try to read before the first meeting. You should rank the projects in order of preference and e-mail me back your options by noon, 2/28 (Wed). The project assignment will be announced on Wednesday afternoon. I expect a maximum of 2 groups to be assigned to each project. And groups assigned to the same project may also work on different aspects of the same project.

For each project there is the following steps:

- I will meet each project group on a weekly starting 3/6 (Mon) The meeting will roughly be 10-15 minutes. I will work with each group for a time slot, but it will be either Monday or Friday. For each meeting (starting the second one) there will be milestones I expect each group to finish by then. Overall progress through the milestones will count towards 25% of the project grade.
- Each group will need to present its work on 5/9 (Tue) between 3-7pm. Each group will have around 10-12 minutes for its presentation. More details will be provided later. This will count towards 15% of the grade
- The final deliverables for each project need to be uploaded to Canvas (as a zip file) by noon 5/10 (Wed). This will count towards 60% of the grade.

List of projects

1. Argument Mining

   The overarching goal for this project is to detect arguments that is being made in written text, where for this project is limited mostly to research papers. For instance, an argument is made up of a claim ("the earth is round"), and a set of evidence that support the claim ("ships that travel west eventually come back to the original place"), together with logical argument steps that justify the use of such evidence. The goal of this project is to develop methods to automate that process.

   Papers:

   - John Lawrence and Chris Reed. 2019. Argument Mining: A Survey. Computational Linguistics, 45(4):765–818.

   - Keshav Singh, Paul Reisert, Naoya Inoue, Pride Kavumba, and Kentaro Inui. 2019. Improving Evidence Detection by Leveraging Warrants. In *Proceedings of the Second Workshop on Fact Extraction and VERification (FEVER)*, pages 57–62, Hong Kong, China. Association for Computational Linguistics.

- Eva Maria Vecchi, Neele Falk, Iman Jundi, and Gabriella Lapesa. 2021. Towards Argument Mining for Social Good: A Survey. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1338–1352, Online. Association for Computational Linguistics

2. Classifying twitter's user opinion/stance

   Social media provides a platform for people to display their stands on certain topics. For this project, we will be given a topic/issue (by key terms, or hashtags), the goal is to detect and classify tweets (and maybe users) by their stance on (aspects of) the issue

   Papers:

   - Momchil Hardalov, Arnav Arora, Preslav Nakov, and Isabelle Augenstein. 2022. A Survey on Stance Detection for Mis- and Disinformation Identification. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 1259–1277, Seattle, United States. Association for Computational Linguistics.
   - Parush Gera and Tempestt Neal. 2022. A Comparative Analysis of Stance Detection Approaches and Datasets. In *Proceedings of the 3rd Workshop on Evaluation and Comparison of NLP Systems*, pages 58–69, Online. Association for Computational Linguistics.
   - Kyle Glandt, Sarthak Khanal, Yingjie Li, Doina Caragea, and Cornelia Caragea. 2021. Stance Detection in COVID-19 Tweets. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1596–1611, Online. Association for Computational Linguistics.

3. Overcoming LDA instability

   One challenge of using topic models like LDA is the instability of it. For instance, running LDA on the same set of documents for multiple times likely will return different topics. This project explore various methods that will enable a coherent results of topics to be output.

   Papers:

   - Yi Yang, Shimei Pan, Jie Lu, Mercan Topkara, and Yangqiu Song. 2016. The Stability and Usability of Statistical Topic Models. ACM Trans. Interact. Intell. Syst. 6, 2, Article 14 (August 2016), 23 pages. https://doi.org/10.1145/2954002
   - Yi Yang, Shimei Pan, Yangqiu Song, Jie Lu, and Mercan Topkara. 2016. Improving topic model stability for effective document exploration. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI'16). AAAI Press, 4223–4227.

- Mika V. Montyla, Maelick Claes and Umar Faroaq, Measuring LDA topic stability from clusters of replicated runs, *In Proceedings of the 12th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM '18)*, 2018.

4. Detecting trends in topic modelling

   We have obtained a set of meeting records for a city's legislature, including debates and bills. We would like to apply topic modelling to determine topics that are of interests, as well as divide the data up in multiple ways (e.g. time, political affiliation etc.) for model comparison. We may also apply different kinds of topic modeling techniques.

   References

- David M. Blei and John D. Lafferty. 2006. Dynamic topic models. In Proceedings of the 23rd international conference on Machine learning (ICML '06). Association for Computing Machinery, New York, NY, USA, 113–120. DOI:https://doi.org/10.1145/1143844.1143859
- Rem Hida, Naoya Takeishi, Takehisa Yairi, Koichi Hori: Dynamic and Static Topic Model for Analyzing Time-Series Document Collections. ACL (2) 2018: *516-520*
- Liangjie Hong, Byron Dom, Siva Gurumurthy, and Kostas Tsioutsiouliklis. 2011. A time-dependent topic model for multiple text streams. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '11). Association for Computing Machinery, New York, NY, USA, 832–840. https://doi.org/10.1145/2020408.2020551
- King Ip Lin and Sabrina Peng. 2022. Enhancing Digital History – Event discovery via Topic Modeling and Change Detection. In *Proceedings of the 2nd International Workshop on Natural Language Processing for Digital Humanities*, pages 69–78, Taipei, Taiwan. Association for Computational Linguistics.

5. Text to SQL comparison

   Recently there has been development of tools that given a database scheme and a query written in natural language, convert it to SQL. For this project you are to explore such systems and compare their performances. We may also try to adapt them by slight changes of the algorithms.

   References

Kalajdjieski, Jovan, Martina Toshevska, and Frosina Stojanovska. "Recent Advances in SQL Query Generation: A Survey." *arXiv preprint arXiv:2005.07667* (2020).

Catherine Finegan-Dollak, Jonathan K. Kummerfeld, Li Zhang, Karthik Ramanathan, Sesh Sadasivam, Rui Zhang, and Dragomir Radev. 2018. Improving Text-to-SQL Evaluation Methodology. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 351–360, Melbourne, Australia. Association for Computational Linguistics.
Spider: https://yale-lily.github.io/spider

- Ruiqi Zhong, Tao Yu, and Dan Klein. 2020. Semantic Evaluation for Text-to-SQL with Distilled Test Suites. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 396–411, Online. Association for Computational Linguistics.

6. Common sense reasoning methods

   AI has been very good in reasoning within a very specific domain. However, so far AI is having problems reasoning with "common sense", simple but wildly applicable facts that is true over domains (e.g. when you are in a dark room, you cannot see clearly). There is a recent focus on how to enable common sense reasoning for computer systems. The goal of this project is to survey the state of the art and test out various systems that approach the task.

   Papers:

   Davis, Ernest. "Benchmarks for Automated Commonsense Reasoning: A Survey." *arXiv preprint arXiv:2302.04752* (2023).

   Ernest Davis and Gary Marcus. 2015. Commonsense reasoning and commonsense knowledge in artificial intelligence. Commun. ACM 58, 9 (September 2015), 92–103. https://doi.org/10.1145/2701413

   Nazneen Fatema Rajani, Bryan McCann, Caiming Xiong, and Richard Socher. 2019. Explain Yourself! Leveraging Language Models for Commonsense Reasoning. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4932–4942, Florence, Italy. Association for Computational Linguistics.

7. Learning the 5W1H of events via semantic role labeling

   Semantic role labeling is a technique to discover various aspects (when? Where? Why? Who? What? How?) of events in text. The goal of this project is to apply various machine learning method to develop tools to apply role labelling to answer such questions.

   Papers

- Ana-Maria Giuglea and Alessandro Moschitti. 2006. Semantic Role Labeling via FrameNet, VerbNet and PropBank. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 929–936, Sydney, Australia. Association for Computational Linguistics.
- Daniil Larionov, Artem Shelmanov, Elena Chistova, and Ivan Smirnov. 2019. Semantic Role Labeling with Pretrained Language Models for Known and Unknown Predicates. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, pages 619–628, Varna, Bulgaria. INCOMA Ltd..
- Sangdo Han, Kyusong Lee, Donghyeon Lee, and Gary Geunbae Lee. 2013. Counseling Dialog System with 5W1H Extraction. In *Proceedings of the SIGDIAL 2013 Conference*, pages 349–353, Metz, France. Association for Computational Linguistics.

8. Hypernym discovery (SemEval 2018, Task 9)

Hypernymy is the capability to generalize for specific case to a general case. E.g. “Red”, :”Green” and “Blue” are all colors. The goal of this project is to examine (and maybe develop tools that allow one to discover hypernym information over a set of corpus

Papers:

Jose Camacho-Collados. 2017. *Why we have switched from building full-fledged taxonomies to simply detecting hypernymy relations*. arXiv preprint arXiv:1703.04178.

William Held and Nizar Habash. 2019. The Effectiveness of Simple Hybrid Systems for Hypernym Discovery. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3362–3367, Florence, Italy. Association for Computational Linguistics.

Sergey Afonin. 2010. On Automated Hyperonym Hierarchy Construction Using an Internet Search Engine. In *Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation*, pages 341–348, Tohoku University, Sendai, Japan. Institute of Digital Enhancement of Cognitive Processing, Waseda University.