

LEAN-LIFE: A Label-Efficient Annotation Framework Towards Learning from Explanation

Dong-Ho Lee^{1*} Rahul Khanna^{1*} Bill Yuchen Lin¹ Seyeon Lee¹ Qinyuan Ye¹
Elizabeth Boschee² Leonardo Neves³ Xiang Ren^{1,2}

¹Department of Computer Science, University of Southern California

²Information Science Institute, University of Southern California ³Snap Inc.

{dongho.lee, rahulkha, yuchen.lin, seyeonle, qinyuany}@usc.edu,

boschee@isi.edu, lneves@snap.com, xiangren@usc.edu

Abstract

Successfully training a deep neural network demands a huge corpus of labeled data. However, each label only provides limited information to learn from and collecting the requisite number of labels involves massive human effort. In this work, we introduce LEAN-LIFE¹, a web-based, Label-Efficient Annotation framework for sequence labeling and classification tasks, with an easy-to-use UI that not only allows an annotator to provide the needed labels for a task, but also enables Learning From Explanations for each labeling decision. Such explanations enable us to generate useful additional labeled data from unlabeled instances, bolstering the pool of available training data. On three popular NLP tasks (named entity recognition, relation extraction, sentiment analysis), we find that using this enhanced supervision allows our models to surpass competitive baseline F1 scores by more than 5-10 percentage points, while using 2X times fewer labeled instances. Our framework is the first to utilize this enhanced supervision technique and does so for three important tasks—thus providing improved annotation recommendations to users and an ability to build datasets of *(data, label, explanation)* triples instead of the regular *(data, label)* pair.

1 Introduction

Deep neural networks have achieved state-of-the-art performance on a wide range of sequence labeling and classification tasks such as named entity recognition (NER) (Lample et al., 2016; Ma and Hovy, 2016), relation extraction (RE) (Zeng et al., 2015; Zhang et al., 2017; Ye et al., 2019), and sentiment analysis (SA) (Wang et al., 2016). However, they only yield such performance levels

*Both authors contributed equally.

¹The source code is publicly available at <http://inklab.usc.edu/leanlife/>.

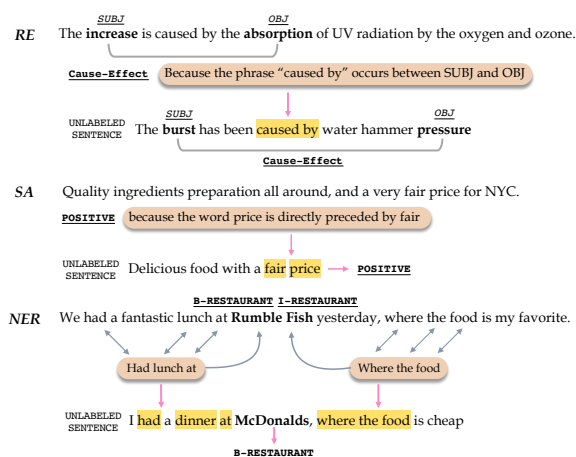


Figure 1: **Leveraging Labeling Explanations:** 1) **RE:** the explanation “the phrase ‘caused by’ occurs between SUBJ and OBJ” can aid in weakly labeling unlabeled instances like “The burst has been caused by water hammer pressure” with the label “cause-effect”; 2) **NER:** Trigger spans near the labeled restaurant such as “had lunch at” and “where the food” can aid in weakly labeling unlabeled instances like “I had a dinner at McDonalds, where the food is cheap”.

in supervised learning scenarios, and in particular when human-annotated data is abundant. As we seek to apply NLP models to larger variety of domains, such as product reviews (Luo et al., 2018), social media messages (Lin et al., 2017), while reducing human annotation efforts, better annotation frameworks with label-efficient learning techniques are crucial to our progress.

Annotation frameworks have been explored by several previous works (Stenatorp et al., 2012; Bontcheva et al., 2014; Morton and LaCivita, 2003; de Castilho et al., 2016; Yang et al., 2018a). These existing open-source sequence annotation tools mainly focus on optimizing user-friendly user interfaces, such as providing shortcut key functionality to allow for faster tagging. The frameworks also attempt to provide annotation recommendation to reduce human annotation ef-

forts. However, these recommendations are provided by a pre-trained model or via dictionary look-ups. This methodology of providing recommendations often proves to be unhelpful when little annotated data exists for pre-training, as is usually the case for natural language tasks being applied to domain-specific or user-provided corpora.

To resolve this issue, AlpacaTag, an annotation framework for sequence labeling (Lin et al., 2019) attempts to provide annotation recommendations from a learned sequence labeling model that is incrementally updated by batches of incoming human annotations. Its model training follows an active learning strategy (Shen et al., 2017), which is shown to be a label-efficient, thus it attempts to minimize human annotation efforts. AlpacaTag selects the most informative batches of documents for humans to annotate and thus achieves a more cost-effective way of using human efforts. While active learning allows the model to achieve higher performance earlier in the learning process, model performance could be improved if additional supervision existed. It is imperative that provided annotation recommendations be as accurate as possible, as inaccurate annotation recommendations from the framework can push users towards generating noisy data, hindering instead of aiding the model training process.

Our effort to prevent this problem is centered around allowing annotators to provide additional supervision by capturing labeling explanations, while still taking advantage of the cost-effectiveness of active learning. Specifically, as shown in Fig. 1, we allow annotators to provide explanations for their decisions in natural language or by selecting *triggers*—nearby phrases that provide helpful context for their decisions. These enhanced annotations allow for model training over both user-provided labels, as well as weakly labeled data created by parsing explanations into high precision labeling rules. We therefore make attempts to ameliorate the erroneous recommendation problem by a performance-boosting training strategy that incorporates both labeled and unlabeled data.

Our work is also similar to recent attempts that exploit explanations for an improved training process (Srivastava et al., 2017; Hancock et al., 2018; Zhou et al., 2020; Wang* et al., 2020), but with two main differences. First, we embed this improved process in a practical application and sec-

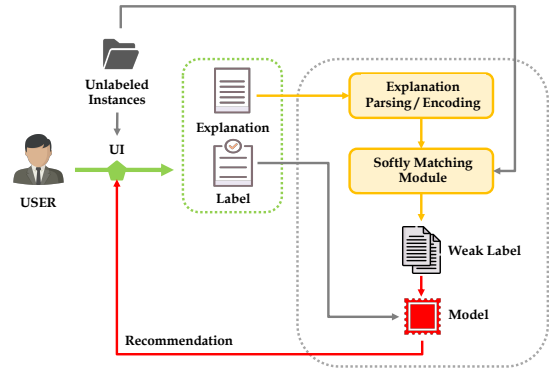


Figure 2: **System Architecture.**

ond, we design task specific architectures to incorporate the now captured explanations into training.

To the best of our knowledge, there is no existing open-source, easy-to-use, recommendation-providing, online-learning annotation framework that can also capture explanations. LEAN-LIFE is the first framework to capture and leverage explanations for improved model training and performance, while still inheriting the advantages of existing tools. We summarize our contributions as:

- **Improved Model Training:** Our recommendation models use a performance improving training process that leverages explanations to weakly label unlabeled instances. Our models improve on competitive baseline F-1 scores by more than 5-10 percentage points, while using 2X less data.
- **Multiple Supported Tasks:** Our framework supports both sequence labeling (as in NER) and sequence classification (as in RE, SA).
- **Explanation Dataset Creation:** We make it easy to build a new type of dataset, one that consists of triples of: text, labels and labeling explanations. The exporting of this captured data is available in two common data formats, CSV and JSON.

2 System Overview

As shown in Fig. 2, our framework consists of two main components, a user-friendly web-UI that can capture labels and explanations for labeling decisions, and a weak supervision framework that parses explanations for the creation of weakly labeled data. The framework then uses this weakly labeled data in conjunction with user-provided labels to train models for improved annotation recommendations. Our UI shows annotators unlabeled instances (can be sampled using active learning), along with annotation recommendations in an effort to reduce annotation costs. We use PyTorch to build our models and implement an API

for communication between the web-UI and our weak supervision framework. The learned parameters of our framework are updated in an online fashion, thus improving in near real time. We will first touch on the annotation UI (§3) and then go into our weak supervision framework (§4).

3 UI for Capturing Human Explanation

The emphasis of our front-end design is to simplify the capture of both label and explanation for each labeling decision, while reducing annotation effort via accessible annotation recommendation. Our framework supports two forms of explanations, *Triggers* and *Natural Language*. A Trigger is a group of words in the sentence being annotated that aided the annotator’s labeling decision, while Natural Language is a written explanation of the labeling decision. This section presents first the UI for capturing triggers (§3.1) and then the UI for capturing natural language explanations (§3.2).

3.1 Capturing Triggers

Fig. 3 illustrates how our framework can capture both a named entity (NE) label and triggers for the sentence “We had a fantastic lunch at Rumble Fish yesterday where the food is my favorite”. The user is first presented with a piece of text to annotate (Annotating Section), the available labels that may be applied to sub-sequences (spans) of text (in the blue header) and recommendations of what spans of text should be considered as NE mentions (Named Entity Recommendation Section). The user may choose to select a span of text to label, or they may click on one of the recommended spans below (Fig. 2a). If the user clicks on a recommended span, a small pop-up displaying the available labels appear with the recommended label circled in red (Fig. 2a). Once the user selects a label for a span of text by either clicking on the desired label button or via a predefined shortcut key (ex: for Restaurant the shortcut key is **r**), a pop-up appears (Fig. 2b), asking the user to select helpful spans (triggers) from the text that provide useful context in deciding the label for the NEM—multiple triggers may be selected. The user may cancel their decision to label a span of text with a label by clicking the *x* button in the pop-up, but if the user wants to proceed and has selected at least one trigger, they finish the labeling by hitting done. Then, their label is visualized in the Annotating Section by highlighting the NEM.

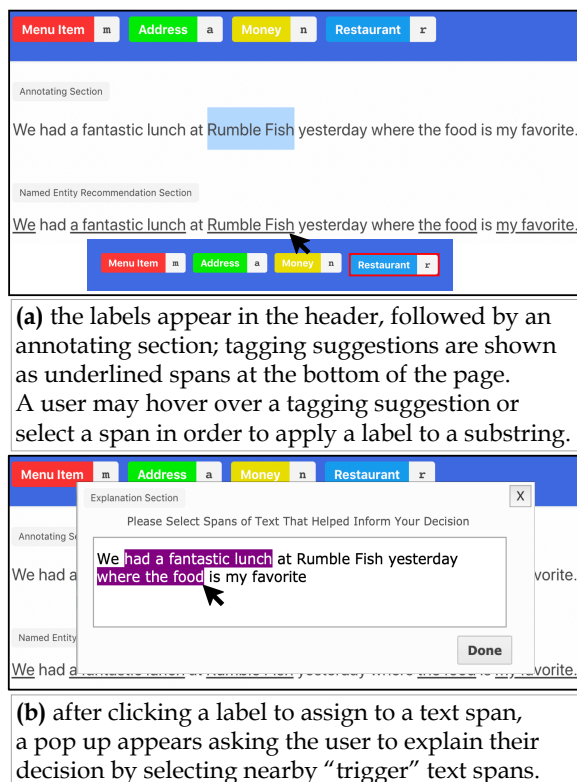


Figure 3: The workflow to annotate a NE label and trigger span. (“Rumble Fish” as Restaurant).

3.2 Capturing Natural Language

Fig. 4 illustrates how for the sentence “Tahawwur Hussain Rana who was born in Pakistan but is a Canadian citizen” our framework can capture both a relation label between NEs and the subsequent natural language explanation. First, the user is tasked to find the NEs in the sentence. After labeling at least two non-consecutive spans of text as NEs, the user may check off the boxes that appear above the labeled NEs. Once two boxes have been checked off, the labels in the blue header are replaced with the labels for relations. The click-order of the checked boxes is displayed and is considered the order of the relation. Also, we display a recommend label to the user in the header section with a circle (Fig. 2a). After clicking on a label, a pop-up appears asking the user to indicate semantic and syntactic reasons as to why the labeling decision is true. Since the natural language explanations are assumed to be made up of predefined predicates, as the user types we incrementally provide predicates to aid the construction of an explanation (Fig. 2b). In this way, we nudge users towards writing explanations the semantic parser is able to break down, allowing our framework to extract a useful logical form from the explanation.

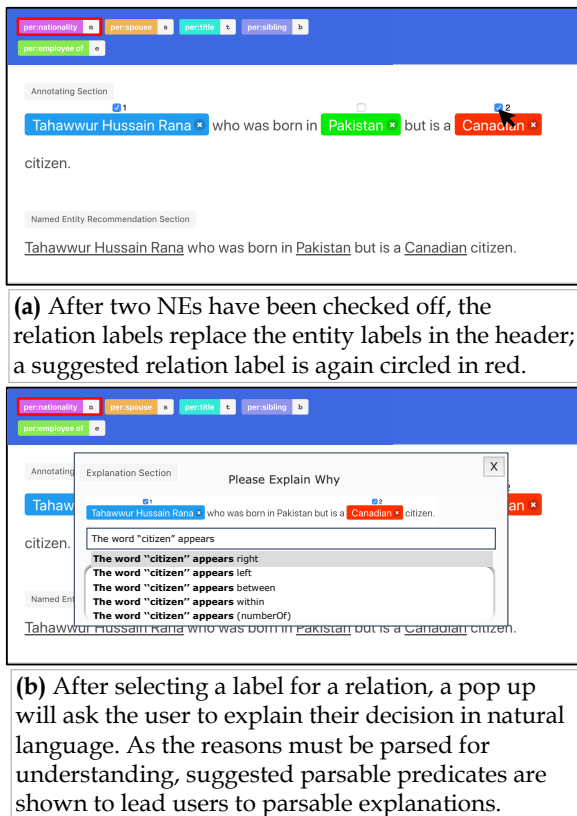


Figure 4: The workflow to annotate a relation label and NL explanation. (“per:nationality” as a relation label between “Tahawwur Hussain Rana” and “Canadian”).

4 LEAN-LIFE Framework

Our Weak Supervision Framework is composed of two main components, a weak labeling module that parses explanations to create labeling rules and a downstream model. The framework parses user-provided explanations to generate weakly labeled data and then trains the appropriate downstream model with this augmented training data. Our weak labeling module supports both explanation formats provided to the annotator in the UI—triggers and natural language. This section first introduces how the module utilizes triggers (§4.1) and then presents how the module deals with natural language (§4.2).

4.1 Input: Trigger

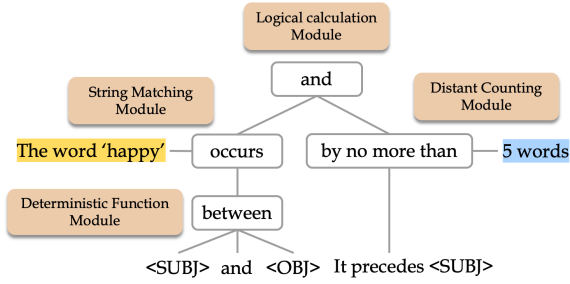
When a trigger is inputted into the system, we generate weak labels for our training data via soft-matching between trigger representations and unlabeled sentences (Lin et al., 2020). Each sentence may contain one or more triggers, but each trigger is associated with only one label. Our framework jointly learns a mapping between triggers and their label using a linear layer with a soft-max output

and a log-likelihood loss, as well as the semantic similarity between the triggers and their associated sentences using contrastive loss—we weigh both objectives equally. Through this joint learning, our trigger representations can capture label knowledge as well as semantic information. We use these representations to improve model training by generating weakly labeled data via soft matching on the unlabeled sentences. More specifically, for each unlabeled sentence, we first calculate the semantic similarity between the sentence and all collected triggers and then filter out all triggers where the similarity distance is larger than our fixed threshold. We then generate a trigger-aware sentence encoding for each threshold-passing trigger and feed these encodings into a downstream classifier for label inference. Finally, we conduct majority vote over outputted label sequences to finalize our weak labels for the unlabeled sentence. In this manner we are able to train over more data, where a good portion of it is weakly labeled.

4.2 Input: Natural Language

When natural language is inputted into the system, our module grows training data via soft-matching between logical forms parsed from natural language explanations and unlabeled sentences. The module follows the Neural Execution Tree framework of (Wang* et al., 2020) when dealing with natural language. First, the explanation is parsed into a logical form by a semantic parser. Previous works have suggested using similar logical forms to improve model training by strict matching on the pool of unlabeled sentences to generate additional labeled data. However, (Wang* et al., 2020) proposes an improved model training paradigm, which relaxes this strict matching constraint, subsequently improving weak labeling coverage and allowing for a larger pool of unlabeled data to be used for model training. Our module does assume each NL explanation can be broken down into a logical form composed of clauses consisting of predicates from four categories—hence the auto-suggest feature in the UI. At weak labeling, the module scores how likely a given unlabeled sentence fits each clause and then constructs an aggregate score representing the match between the logical form and the unlabeled sentence. If the final score is above configurable thresholds, we weakly label the sentence with the appropriate label.

As shown in Fig. 5, the scoring portion of our



Explanation : The word ‘happy’ occurs between <SUBJ> and <OBJ> and it precedes <SUBJ> by no more than 5 words

Figure 5: Weakly labeling module for exploiting natural language explanation. the keyword is ‘happy’

module has four parts: *String Matching Module*, *Distant Counting Module*, *Deterministic Function Module*, and the *Logical Calculation Module*. The first three modules are responsible for evaluating if different clauses in the logical form are applicable for the given unlabeled sentence, while the *Logical Calculation Module*’s job is to aggregate scores between the various clauses. The *String Matching Module* returns a sequence of scores $[s_1, s_2, \dots, s_n]$ indicating the similarity between each token w_i and the keyword q —“happy” in Fig. 5. Our *Distant Counting Module* aims to relax the distance constraint stated in the explanation, ex: “by no more than 5 words”. If the position of keyword q strictly satisfies the constraint, the score is set to 1, otherwise the score decreases as the constraint is less satisfied. Finally, the *Deterministic Function Module* deals with deterministic predicates like “LEFT”, “BETWEEN”, which can only be exactly matched in terms of the keyword q . Scores are aggregated by the *Logical Calculation Module* to output a final relevancy score.

5 Experiments

We conduct extensive experiments investigating label efficiency to prove the effectiveness of our annotation models. We found that using natural language explanations for RE and SA, and trigger explanations for NER provided the best results. For the downstream model portion of our weak supervision framework, we use common supervised method for each task: (1-RE) BLSTM+ATT (Bahdanau et al., 2015) adds an attention layer onto LSTM to encode an sequence. (2-SA) ATAE-LSTM (Wang et al., 2016) combines the aspect term information into both the embedding layer and attention layer to help the model concentrate on different parts of a sentence.

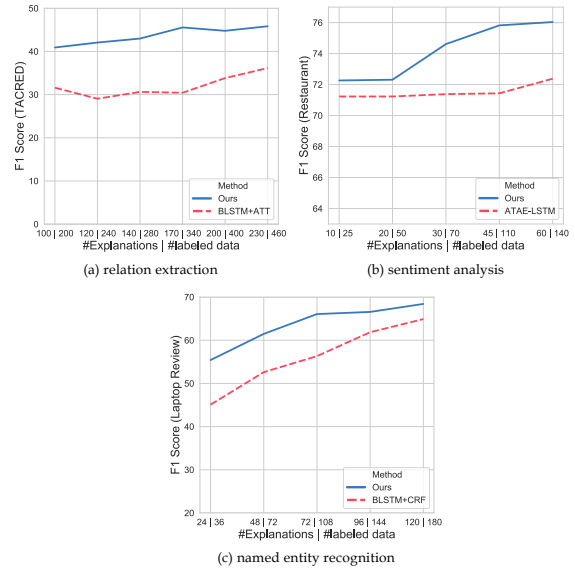


Figure 6: **Label Efficiency**. We choose commonly-used supervised baselines for comparison.

(3-NER) BLSTM+CRF (Ma and Hovy, 2016) encodes character sequences into a vector and concatenates the vector with pre-trained word embeddings to feed into word-level BLSTM. Then, it applies a CRF layer to predict sequence labels. Then we compare these methods as baselines.

Tasks and Datasets We test our implementation on three tasks: RE, SA, NER. We use TACRED (Zhang et al., 2017) for RE, Restaurant review from SemEval 2014 Task 4 for SA, and Laptop reviews (Pontiki et al., 2016) for NER.

Label Efficiency We claim that when starting with little to no labeled data, it is more effective to ask annotators to provide a label and an explanation for the label, than to just request a label. To support this claim, we conduct experiments to demonstrate the label efficiency of our explanation-leveraging-model. We found that the time for labeling one instance plus providing an explanation takes 2X times more time than just simply providing a label. Given this annotation time observation, we compare the performance between our improved training process and the traditional label-only training process by holding annotation time constant between the two trials. This means we expose the label-only supervised model to the appropriate multiple of labeled instances that the label-and-explanation supervised model is shown Fig. 6. Each marker on the x-axis of the plots indicate a certain interval of annotation time, which is represented by the number of

label+explanations our augmented model training paradigm is given vs. how many labels the traditional label-only model training is shown. We use the F-1 metric to compare the performances. As shown in Fig. 6, we see that our model not only is more time and label efficient than the label-only training process, but it also outperforms the label-only training process. Given these results, we believe it is worth to request a user to provide both a label and an explanation for the label. Not only does the improvement in performance justify the extra time required to provide the explanation, but we also can achieve higher performance with fewer datapoints / less annotation time.

6 Related Works

Leveraging natural language explanations for additional supervision has been explored by many works. (Srivastava et al., 2017) first demonstrated the idea of using natural language explanations for weak labeling by jointly training a task-specific semantic parser and label classifier to generate weak labels. This method is limited though, as the parser is too tightly coupled to the already labeled data, thus their weak learning framework is not able to build a much larger dataset than the one it already has. To address this issue, (Hancock et al., 2018) proposed a weak supervision framework that utilizes a more practical rule-based semantic parser. The parser constructs a logical form for an explanation that is then used as a labeling function—this resulted in a significant increase of the training set. Another effort can be found in (Camburu et al., 2018) work to extend the Stanford Natural Language Inference dataset with natural language explanations—this extension was done for the important textual entailment recognition task. They demonstrate the usefulness of explanations as an additional training signal for learning more comprehensive sentence representations. Even earlier (Andreas et al., 2016) explored breaking down natural language explanation into linguistic sub-structures for learning collections of neural modules which can be assembled into neural networks. Our framework is very related to the above weak supervision methods via explanation.

Another approach to weak supervision is attempting to transfer knowledge from a related source to the target domain corpus (Lin and Lu, 2018; Lan et al., 2020). Shang et al. (2018) and Yang et al. (2018) proposed using a domain-

specific dictionary for matching on the unannotated target corpus. Both efforts employ Partial CRFs (Liu et al., 2014) which assign all possible labels to unlabeled words and maximize the total probability. This approach addresses the incomplete annotation problem, but heavily relies on a domain-specific seed dictionary.

7 Conclusion

In this paper, we propose an open-source web-based annotation framework `LEAN-LIFE` that not only allows an annotator to provide the needed labels for a task, but can also capture explanation for each labeling decision. Such explanations enable a significant improvement in model training while only doubling per instance annotation time. This increase in per instance annotation time is greatly outweighed by the benefits in model training, especially in a low resource settings, as proven by our experiments. This is an important consideration for any annotation framework, as the quicker the framework is able to train annotation recommendation models to reach high performance, the sooner the user receives useful annotation recommendations, which in turn cut down on the annotation time required per instance.

Better training methods also allow us to fight the potential generation of noisy data due to inaccurate annotation recommendations. We hope that our work on `LEAN-LIFE` will allow for researches and practitioners alike to more easily obtain useful labeled datasets and models for the various NLP tasks they face.

Acknowledgements

This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via Contract No. 2019-19051600007, NSF SMA 18-29268, and Snap research gift. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. We would like to thank all the collaborators in USC INK research lab for their constructive feedback on the work.

References

- Jacob Andreas, Marcus Rohrbach, Trevor Darrell, and Dan Klein. 2016. [Neural module networks](#). In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 39–48.
- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. [Neural machine translation by jointly learning to align and translate](#). In *3rd International Conference on Learning Representations*.
- Kalina Bontcheva, Ian Roberts, Leon Derczynski, and Dominic Rout. 2014. [The gate crowdsourcing plugin: Crowdsourcing annotated corpora made easy](#). In *Proc. of EACL*, pages 97–100. Association for Computational Linguistics.
- Oana-Maria Camburu, Tim Rocktäschel, Thomas Lukasiewicz, and Phil Blunsom. 2018. [e-snli: Natural language inference with natural language explanations](#). In *Advances in Neural Information Processing Systems 31*, pages 9539–9549. Curran Associates, Inc.
- Richard Eckart de Castilho, Eva Mujdicza-Maydt, Seid Muhie Yimam, Silvana Hartmann, Iryna Gurevych, Anette Frank, and Chris Biemann. 2016. [A web-based tool for the integrated annotation of semantic and syntactic structures](#). In *LT4DH@COLING*, pages 76–84. The COLING 2016 Organizing Committee.
- Braden Hancock, Paroma Varma, Stephanie Wang, Martin Bringmann, Percy Liang, and Christopher Ré. 2018. [Training classifiers with natural language explanations](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1884–1895, Melbourne, Australia. Association for Computational Linguistics.
- Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. 2016. [Neural architectures for named entity recognition](#). In *Proc. of NAACL-HLT*, pages 260–270. Association for Computational Linguistics.
- Ouyu Lan, Xiao Huang, Bill Yuchen Lin, He Jiang, Liyuan Liu, and Xiang Ren. 2020. [Learning to contextually aggregate multi-source supervision for sequence labeling](#). In *Proceedings of Association for Computational Linguistics*.
- Bill Y. Lin, Frank Xu, Zhiyi Luo, and Kenny Zhu. 2017. [Multi-channel BiLSTM-CRF model for emerging named entity recognition in social media](#). In *Proceedings of the 3rd Workshop on Noisy User-generated Text*, pages 160–165, Copenhagen, Denmark. Association for Computational Linguistics.
- Bill Yuchen Lin, Dong-Ho Lee, Ming Shen, Ryan Moreno, Xiao Huang, Prashant Shiralkar, and Xiang Ren. 2020. [Triggerer: Learning with entity triggers as explanations for named entity recognition](#). In *Proceedings of Association for Computational Linguistics*.
- Bill Yuchen Lin, Dong-Ho Lee, Frank F. Xu, Ouyu Lan, and Xiang Ren. 2019. [AlpacaTag: An active learning-based crowd annotation framework for sequence tagging](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 58–63, Florence, Italy. Association for Computational Linguistics.
- Bill Yuchen Lin and Wei Lu. 2018. [Neural adaptation layers for cross-domain named entity recognition](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2012–2022, Brussels, Belgium. Association for Computational Linguistics.
- Yijia Liu, Yue Zhang, Wanxiang Che, Ting Liu, and Fan Wu. 2014. [Domain adaptation for CRF-based Chinese word segmentation using free annotations](#). In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 864–874, Doha, Qatar. Association for Computational Linguistics.
- Zhiyi Luo, Shanshan Huang, Frank F. Xu, Bill Yuchen Lin, Hanyuan Shi, and Kenny Zhu. 2018. [ExtRA: Extracting prominent review aspects from customer feedback](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3477–3486, Brussels, Belgium. Association for Computational Linguistics.
- Xuezhe Ma and Eduard Hovy. 2016. [End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1064–1074, Berlin, Germany. Association for Computational Linguistics.
- Thomas Morton and Jeremy LaCivita. 2003. [Word-Freak: An open tool for linguistic annotation](#). In *Companion Volume of the Proceedings of HLT-NAACL 2003 - Demonstrations*, pages 17–18.
- Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, Véronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeniy Kotelnikov, Nuria Bel, Salud María Jiménez-Zafra, and Gülşen Eryiğit. 2016. [SemEval-2016 task 5: Aspect based sentiment analysis](#). In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 19–30, San Diego, California. Association for Computational Linguistics.
- Jingbo Shang, Liyuan Liu, Xiaotao Gu, Xiang Ren, Teng Ren, and Jiawei Han. 2018. [Learning named entity tagger using domain-specific dictionary](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2054–2064, Brussels, Belgium. Association for Computational Linguistics.

- Yanyao Shen, Hyokun Yun, Zachary Lipton, Yakov Kronrod, and Animashree Anandkumar. 2017. [Deep active learning for named entity recognition](#). In *Proceedings of the 2nd Workshop on Representation Learning for NLP*, pages 252–256, Vancouver, Canada. Association for Computational Linguistics.
- Shashank Srivastava, Igor Labutov, and Tom Mitchell. 2017. [Joint concept learning and semantic parsing from natural language explanations](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1527–1536, Copenhagen, Denmark. Association for Computational Linguistics.
- Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun’ichi Tsujii. 2012. [brat: a web-based tool for NLP-assisted text annotation](#). In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 102–107, Avignon, France. Association for Computational Linguistics.
- Yequan Wang, Minlie Huang, Xiaoyan Zhu, and Li Zhao. 2016. [Attention-based LSTM for aspect-level sentiment classification](#). In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 606–615, Austin, Texas. Association for Computational Linguistics.
- Ziqi Wang*, Yujia Qin*, Wenxuan Zhou, Jun Yan, Qinyuan Ye, Leonardo Neves, Zhiyuan Liu, and Xiang Ren. 2020. [Learning from explanations with neural execution tree](#). In *International Conference on Learning Representations*.
- Jie Yang, Yue Zhang, Linwei Li, and Xingxuan Li. 2018a. [YEDDA: A lightweight collaborative text span annotation tool](#). In *Proceedings of ACL 2018, System Demonstrations*, pages 31–36, Melbourne, Australia. Association for Computational Linguistics.
- Yaosheng Yang, Wenliang Chen, Zhenghua Li, Zhengqiu He, and Min Zhang. 2018b. [Distantly supervised NER with partial annotation learning and reinforcement learning](#). In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2159–2169, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- Qinyuan Ye, Liyuan Liu, Maosen Zhang, and Xiang Ren. 2019. [Looking beyond label noise: Shifted label distribution matters in distantly supervised relation extraction](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3841–3850, Hong Kong, China. Association for Computational Linguistics.
- Daojian Zeng, Kang Liu, Yubo Chen, and Jun Zhao. 2015. [Distant supervision for relation extraction via piecewise convolutional neural networks](#). In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1753–1762, Lisbon, Portugal. Association for Computational Linguistics.
- Yuhao Zhang, Victor Zhong, Danqi Chen, Gabor Angeli, and Christopher D. Manning. 2017. [Position-aware attention and supervised data improve slot filling](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 35–45, Copenhagen, Denmark. Association for Computational Linguistics.
- Wenxuan Zhou, Hongtao Lin, Bill Yuchen Lin, Ziqi Wang, Junyi Du, Leonardo Neves, and Xiang Ren. 2020. [Nero: A neural rule grounding framework for label-efficient relation extraction](#). *The Web Conference*.