Posterior-regularized REINFORCE for Instance Selection in Distant Supervision

Qi Zhang[1], **Siliang Tang**[1]*, **Xiang Ren**[3], **Fei Wu**[1], **Shiliang Pu**[2] **& Yueting Zhuang**[1]
[1]Zhejiang University, [2]Hikvision, [3]University of Southern California,
{zhangqihit,siliang,wufei,yzhuang}@zju.edu.cn,
pushiliang@hikvision.com,
xiangren@usc.edu

Abstract

This paper provides a new way to improve the efficiency of the REINFORCE training process. We apply it to the task of instance selection in distant supervision. Modeling the instance selection in one bag as a sequential decision process, a reinforcement learning agent is trained to determine whether an instance is valuable or not and construct a new bag with less noisy instances. However unbiased methods, such as REINFORCE, could usually take much time to train. This paper adopts posterior regularization (PR) to integrate some domain-specific rules in instance selection using REINFORCE. As the experiment results show, this method remarkably improves the performance of the relation classifier trained on cleaned distant supervision dataset as well as the efficiency of the REINFORCE training.

1 Introduction

Relation extraction is a fundamental work in natural language processing. Detecting and classifying the relation between entity pairs from the unstructured document, it can support many other tasks such as question answering.

While relation extraction requires lots of labeled data and make methods labor intensive, (Mintz et al., 2009) proposes distant supervision (DS), a widely used automatic annotating way. In distant supervision, knowledge base (KB) , such as Freebase, is aligned with nature documents. In this way, the sentences which contain an entity pair in KB all express the exact relation that the entity pair has in KB. We usually call the set of instances that contain the same entity pair a bag. In this way, the training instances can be divided into N bags $\mathbf{B} = \{B^1, B^2, ..., B^N\}$. Each bag B^k are corresponding to an unique entity pair $E^k = (e_1^k, e_2^k)$ and contains a sequence of instances $\{x_1^k, x_2^k, ..., x_{|B^k|}^k\}$. However, distant supervision may suffer a wrong label problem. In other words, the instances in one bag may not actually have the relation.

To resolve the wrong label problem, just like Fig.2 shows, (Feng et al., 2018) model the instance selection task in one bag B^k as a sequential decision process and train an agent $\pi(a|s, \theta_\pi)$ denoting the probability $P_\pi(A_t = a, |S_t = s, \theta_t = \theta_\pi)$ that action a is taken at time t given that the agent is in state s with parameter vector θ_π by REINFORCE algorithm (Sutton and Barto, 1998). The action a can only be 0 or 1 indicating whether an instance x_i^k is truly expressing the relation and whether it should be selected and added to the new bag $\overline{B^k}$. The state s is determined by the entity pair corresponding to the bag, the candidate instance to be selected and the instances that have already been selected. Accomplishing this task, the agent gets a new bag $\overline{B^k}$ at the terminal of the trajectory with less wrong labeled instances. With the newly constructed dataset $\overline{\mathbf{B}} = \{\overline{B^1}, \overline{B^2}, ..., \overline{B^N}\}$ with less wrong labeling instances, we can train bag level relation predicting models with better performance. Meanwhile, the relation predicting model gives reward to the instance selection agent. Therefore, the agent and the relation classifier can be trained jointly.

However, REINFORCE is a Monte Carlo algorithm and need stochastic gradient method to optimize. It is unbiased and has good convergence properties but also may be of high variance and slow to train (Sutton and Barto, 1998).

Therefore, we train a REINFORCE based agent by integrating some other domain-specific rules to accelerate the training process and guide the agent to explore more effectively and learn a better policy. Here we use a rule pattern as the Fig.1 shows (**?**). The instances that return true (match

*Corresponding author

Proceedings of NAACL-HLT 2019, pages 2831–2835
Minneapolis, Minnesota, June 2 - June 7, 2019.

the pattern and label in any one of the rules) are denoted as x_{MI} and we adopt posterior regularization method (Ganchev, 2010) to regularize the posterior distribution of $\pi(a|s, \theta_\pi)$ on x_{MI}. In this way, we can construct a rule-based agent π_r. π_r tends to regard the instances in x_{MI} valuable and select them without wasting time in trial-and-error exploring. The number of such rules is 134 altogether and can match nearly four percents of instances in the training data.

Our contributions include:

- We propose PR REINFORCE by integrating domain-specific rules to improve the performance of the original REINFORCE.
- We apply the PR REINFORCE to the instance selection task for DS dataset to alleviate the wrong label problem in DS.

2 Related Work

Among the previous studies in relation extraction, most of them are supervised methods that need a large amount of annotated data (Bach and Badaskar, 2007). Distant supervision is proposed to alleviate this problem by aligning plain text with Freebase. However, distant supervision inevitably suffers from the wrong label problem.

Some previous research has been done in handling noisy data in distant supervision. An expressed-at-least-once assumption is employed in (Mintz et al., 2009): if two entities participated in a relation, at least one instance in the bag might express that relation. Many follow-up studies adopt this assumption and choose a most credible instance to represent the bag. (Lin et al., 2016; Ji et al., 2017) employs the attention mechanism to put different attention weight on each sentence in one bag and assume each sentence is related to the relation but have a different correlation.

Another key issue for relation extraction is how to model the instance and extract features, (Zeng et al., 2014, 2015; Zhou et al., 2016) adopts deep

```
return true if match(' * born in * ', x) and r='born_in'
return true if match(' * killed in * ', x) and r='died_in'
return true if match(' * died in * ', x) and r='died_in'
...
...
```

Figure 1: Rule Pattern Examples

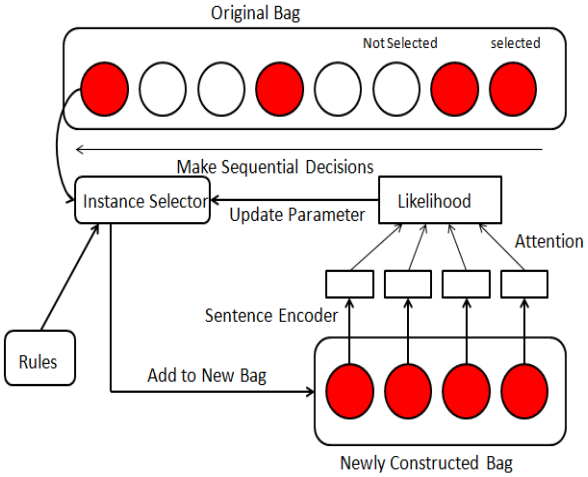

Figure 2: Overall Framework

neural network including CNN and RNN, these methods perform better than conventional feature-based methods.

Reinforcement learning has been widely used in data selection and natural language processing. (Feng et al., 2018) adopts REINFORCE in instance selection for distant supervision which is the basis of our work.

Posterior regularization (Ganchev, 2010) is a framework to handle the problem that a variety of tasks and domains require the creation of large problem-specific annotated data. This framework incorporates external problem-specific information and put a constraint on the posterior of the model. In this paper, we propose a rule-based REINFORCE based on this framework.

3 Methodology

In this section, we focus on the model details. Besides the interacting process of the relation classifier and the instance selector, we will introduce how to model the state, action, reward of the agent and how we add rules for the agent in training process.

3.1 Basic Relation Classifier

We need a pretrained basic relation classifier to define the reward and state. In this paper, we adopt the BGRU with attention bag level relation classifier f_b (Zhou et al., 2016). With $\mathbf{o}$ denoting the output of f_b corresponding to the scores associated to each relation, the conditional probability can be written as follows:

$$P_{f_b}(r|B^k, \theta_b) = \frac{exp(o_r)}{\sum_{k=1}^{n_r} exp(o_k)} \quad (1)$$

where r is relation type, n_r is the number of relation types, θ_b is the parameter vector of the basic relation classifier f_b and B^k denotes the input bag of the classifier.

In the basic classifier, the sentence representation is calculated by the sentence encoder network BGRU, the BGRU takes the instance x_i^k as input and output the sentence representation BGRU(x_i^k). And then the sentence level(ATT) attention will take $\{BGRU(x_1^k), BGRU(x_2^k), ..., BGRU(x_{|B^k|}^k)\}$ as input and output $\mathbf{o}$ which is the final output of f_b corresponding to the scores associated to each relation.

3.2 Original REINFORCE

Original REINFORCE agent training process is quite similar to (Feng et al., 2018). The instance selection process for one bag is completed in one trajectory. Agent $\pi(a|s, \theta_\pi)$ is trained as an instance selector.

The key of the model is how to represent the state in every step and the reward at the terminal of the trajectory. We use the pretrained f_b to address this key problem. The reward defined by the basic relation classifier is as follows:

$$R = \log P_{f_b}(r^k|\overline{B^k}, \theta_b) \quad (2)$$

In which r^k denotes the corresponding relation of B^k.

The state s mainly contained three parts: the representation of the candidate instance, the representation of the relation and the representation of the instances that have been selected.

The representation of the candidate instance are also defined by the basic relation classifier f_b. At time step t, we use BGRU(x_t^k) to represent the candidate instance x_t^k and the same for the selected instances. As for the embedding of relation, we use the entity embedding method introduced in TransE model (Bordes et al., 2013) which is trained on the Freebase triples that have been mentioned in the training and testing dataset, and the relation embedding re_k will be computed by the difference of the entity embedding element-wise.

The policy π with parameter $\theta_\pi = \{W, b\}$ is defined as follows:

$$P_\pi(A_t|S_t, \theta_\pi) = softmax(WS_t + b) \quad (3)$$

With the model above, the parameter vector can be updated according to REINFORCE algorithm (Sutton and Barto, 1998).

3.3 Posterior Regularized REINFORCE

REINFORCE uses the complete return, which includes all future rewards up until the end of the trajectory. In this sense, all updates are made after the trajectory is completed (Sutton and Barto, 1998). These stochastic properties could make the training slow. Fortunately, we have some domain-specific rules that could help to train the agent and adopt posterior regularization framework to integrate these rules. The goal of this framework is to restrict the posterior of π. It can guide the agent towards desired behavior instead of wasting too much time in meaninglessly exploring.

Since we assume that the domain-specific rules have high credibility, we designed a rule-based policy agent π_r to emphasize their influences on π. The posterior constrains for π is that the policy posterior for x_{MI} is expected to be 1 which indicates that agent should select the x_{MI}. This expectation can be written as follows:

$$E_{P_\pi}[\mathbf{l}(A_t = 1)] = 1 \quad (4)$$

where $\mathbf{l}$ here is the indicator function. In order to transfer the rules into a new policy π_r, the KL divergence between the posterior of π and π_r should be minimized, this can be formally defined as

$$minKL(P_\pi(A_t|S_t, \theta_\pi)||P_{\pi_r}(A_t|S_t, \theta_\pi)) \quad (5)$$

Optimizing the constrained convex problem defined by Eq.(4) and Eq.(5), we get a new policy π_r:

$$P_{\pi_r}(A_t|S_t, \theta_\pi) = \frac{P_\pi(A_t|S_t, \theta_\pi)exp(\mathbf{l}(A_t = 1) - 1)}{Z} \quad (6)$$

where Z is a normalization term.

$$Z = \sum_{A_t=0}^{1} P_{\pi_r}(A_t|X, \theta_\pi)exp(\mathbf{l}(A_t = 1) - 1)$$

Algorithm 1 formally define the overall framework of the rule-based data selection process.

4 Experiment

Our experiment is designed to demonstrate that our proposed methodologies can train an instance selector more efficiently.

Data: Original DS Dataset:
$\mathbf{B} = \{B^1, B^2, ..., B^N\}$, Max Episode:M, Basic Relation Classifier: f_b, Step Size: α
Result: An Instance Selector
initialization policy weight $\theta'_\pi = \theta_\pi$;
initialization classifier weight $\theta'_b = \theta_b$;
for *episode m=1 to M* **do**
  for B^k *in* ***B*** **do**
    $B^k = \{x_1^k, x_2^k, ..., x_{|B^k|}^k\}$, $\overline{B^k} = \{\}$;
    for *step i in* $|B^k|$ **do**
      construct s_i by $\overline{B^k}, x_i^k, re_k$;
      if $x_i^k \in x_{MI}$ **then**
        construct π_r;
        sample action A_i follow $\pi_r(a|s_i, \theta'_\pi)$;
      else
        sample action A_i follow $\pi(a|s_i, \theta'_\pi)$;
      end
      if A_i*=1* **then**
        Add x_i^k in $\overline{B^k}$;
      end
    end
    Get terminal reward: $R = \log P_{f_b}(r^k|\overline{B^k}, \theta'_b)$;
    Get step delayed reward: R_i=R;
    Update agent: $\theta_\pi \leftarrow \theta_\pi + \alpha \sum_{i=1}^{|B^k|} R_i \nabla_{\theta_\pi} \log \pi$
  end
  $\theta'_\pi = \tau\theta_\pi + (1-\tau)\theta'_\pi$;
  Update the classifier f_b;
end

Algorithm 1: PR REINFORCE

We tuned our model using three-fold cross validation on the training set. For the parameters of the instance selector, we set the dimension of entity embedding as 50, the learning rate as 0.01. The delay coefficient τ is 0.005. For the parameters of the relation classifier, we follow the settings that are described in (Zhou et al., 2016).

The comparison is done in rule-based reinforcement learning method, original reinforcement learning and method with no reinforcement learning which is the basic relation classifier trained on original DS dataset. We use the last as the baseline.

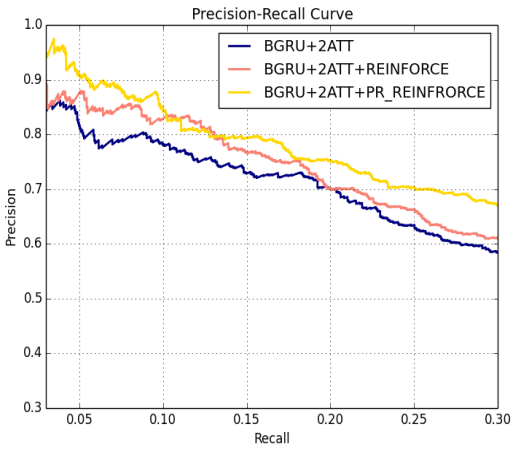

Figure 3: Precision/Recall Curves

4.1 Dataset

A widely used DS dataset, which is developed by (Riedel et al., 2010), is used as the original dataset to be selected. The dataset is generated by aligning Freebase with New York Times corpus.

4.2 Metric and Performance Comparison

We compare the data selection model performance by the final performance of the basic model trained on newly constructed dataset selected by different models. We use the precision/recall curves as the main metric. Fig.3 presents this comparison. PR REINFORCE constructs cleaned DS dataset with less noisy data compared with the original REINFORCE so that the BGRU+2ATT classifier can reach better performance.

5 Conclusions

In this paper, we develop a posterior regularized REINFORCE methodology to alleviate the wrong label problem in distant supervision. Our model makes full use of the hand-crafted domain-specific rules in the trial and error search during the training process of REINFORCE method for DS dataset selection. The experiment results show that PR REINFORCE outperforms the original REINFORCE. Moreover, PR REINFORCE greatly improves the efficiency of the REINFORCE training.

Acknowledgments

This work has been supported in part by NSFC (No.61751209, U1611461), 973 program (No. 2015CB352302), Hikvision-Zhejiang University

Joint Research Center, Chinese Knowledge Center of Engineering Science and Technology (CKCEST), Engineering Research Center of Digital Library, Ministry of Education. Xiang Ren's research has been supported in part by National Science Foundation SMA 18-29268.

References

Nguyen Bach and Sameer Badaskar. 2007. A review of relation extraction. *Literature review for Language and Statistics II*, 2.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Advances in neural information processing systems*, pages 2787–2795.

Jun Feng, Minlie Huang, Li Zhao, Yang Yang, and Xiaoyan Zhu. 2018. Reinforcement learning for relation classification from noisy data.

Kuzman Ganchev. 2010. *Posterior regularization for learning with side information and weak supervision*. Ph.D. thesis, University of Pennsylvania.

Guoliang Ji, Kang Liu, Shizhu He, Jun Zhao, et al. 2017. Distant supervision for relation extraction with sentence-level attention and entity descriptions. In *AAAI*, pages 3060–3066.

Yankai Lin, Shiqi Shen, Zhiyuan Liu, Huanbo Luan, and Maosong Sun. 2016. Neural relation extraction with selective attention over instances. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 2124–2133.

Mike Mintz, Steven Bills, Rion Snow, and Dan Jurafsky. 2009. Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*, pages 1003–1011. Association for Computational Linguistics.

Sebastian Riedel, Limin Yao, and Andrew McCallum. 2010. Modeling relations and their mentions without labeled text. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 148–163. Springer.

Richard S Sutton and Andrew G Barto. 1998. *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge.

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.

Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou, and Jun Zhao. 2014. Relation classification via convolutional deep neural network. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 2335–2344.

Peng Zhou, Wei Shi, Jun Tian, Zhenyu Qi, Bingchen Li, Hongwei Hao, and Bo Xu. 2016. Attention-based bidirectional long short-term memory networks for relation classification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pages 207–212.