Reinforcement Learning for Relation Classification from Noisy Data

Jun Feng, Minlie Huang, Li Zhao, Yang Yang, Xiaoyan Zhu

Abstract
Existing relation classification methods that rely on distant supervision assume that a bag of sentences mentioning an entity pair are all describing a relation for the entity pair. Such methods, performing classification at the bag level, cannot identify the mapping between a relation and a sentence, and largely suffers from the noisy labeling problem. In this paper, we propose a novel model for relation classification at the sentence level from noisy data. The model has two modules: an instance selector and a relation classifier. The instance selector chooses high-quality sentences with reinforcement learning and feeds the selected sentences into the relation classifier, and the relation classifier makes sentence-level prediction and provides rewards to the instance selector. The two modules are trained jointly to optimize the instance selection and relation classification processes. Experiment results show that our model can deal with the noise of data effectively and obtains better performance for relation classification at the sentence level.

Introduction
Relation classification, aiming to categorize semantic relations between two entities given a plain text, is an important problem in natural language processing, particularly for knowledge graph completion and question answering. Most existing works for relation classification adopt supervised learning approaches, either based on traditional handcrafted features (Mooney and Bunescu 2005; Zhou et al. 2005) or based on the features automatically generated by deep neural networks (Zeng et al. 2014; dos Santos, Xiang, and Zhou 2015), but all require high-quality annotated data.

In order to obtain large-scale training data, distant supervision (Mintz et al. 2009) was proposed by assuming that if two entities have a relation in a given knowledge base, all sentences that contain the two entities will mention that relation. Although distant supervision is effective to label data automatically, it suffers from the noisy labeling problem. Taking the triple (Barack, Obama, BornIn, United, States) as an example, the noisy sentence “Barack Obamba is the 44th president of the United State” will be regarded as a positive instance by distant supervision and a BornIn relation is assigned to this sentence, although the sentence does not describe the relation BornIn at all.

To address the issue of noisy labeling, previous studies adopt multi-instance learning to consider the noises of instances (Riedel, Yao, and McCallum 2010; Hoffmann et al. 2011; Surdeanu et al. 2012; Zeng et al. 2015; Lin et al. 2016; Ji et al. 2017). In these studies, the training and test process is proceeded at the bag level, where a bag contains noisy sentences mentioning the same entity pair but possibly not describing the same relation. As a result, previous studies suffer from two limitations: 1) Unable to handle the sentence-level prediction; 2) Sensitive to the bags with all noisy sentences which do not describe a relation at all.

To better explain the first limitation, we show an example in Figure 1. Bag-level prediction can find the two relations “EmployedBy” and “BornIn” between the entity pair “Barack, Obama” and “United, States”. However, sentence-level prediction is able to further map each relation to the corresponding sentences. As for the second limitation, for each bag, previous bag-level methods retain at least one sentence, even if all the sentences in a given bag are noisy (not describing the relation). Such bags, produced by distant supervision, are quite common. For instance, our investigation on a widely used dataset1 shows that 53% out of 100 sample bags have no sentences that describe the relation. Such noisy bags will definitely decrease the performance of relation classification.

Figure 1: Bag-level: Relations are mapped to a bag of sentences, each of which contains the same entity pair; Sentence-level: Each sentence is mapped to a specific relation.

---

1Copyright © 2018, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.
1http://iesl.cs.umass.edu/riedel/ecml/
In this paper, to handle the above two limitations, we propose a novel relation classification model consisting of two modules: instance selector and relation classifier. By having an explicit instance selector\(^2\), we are able to first select high-quality sentences from a sentence bag, and then predict a relation at the sentence level by the relation classifier. To handle the second limitation, our instance selector will filter the entire bag if all sentences are labeled incorrectly. The major challenge here is how to train the two modules jointly, particularly when the instance selector has no explicit knowledge about which sentences are labeled incorrectly.

We address this challenge by casting the instance selection task as a reinforcement learning problem (Sutton and Barto 1998). Intuitively, although we do not have an explicit supervision for the instance selector, we can measure the utility of the selected sentences as a whole. Thus, the instance selection process has the following two properties: first, trial-and-error-search, meaning that the instance selector attempts to choose some sentences and obtain feedback (or reward) on the quality of the selected sentences from the relation classifier; second, the feedback from the relation classifier can be obtained only when we finish the instance selection process, which is typically delayed. These two properties naturally inspire us to utilize reinforcement learning techniques.

Our contributions in this work include:

- We propose a new model for relation classification, which consists of an instance selector and a relation classifier. This formalization enables our model to extract relations at the sentence level on the cleansed data.
- We formulate instance selection as a reinforcement learning problem, which enables the model to perform instance selection without explicit sentence-level annotations but just with a weak supervision signal from the relation classifier.

**Related Work**

Relation classification is a common task in natural language processing. Many approaches have been developed, particularly with supervised methods (Mooney and Bunescu 2005; Zhou et al. 2005; Zelenko, Aone, and Richardella 2003). However, such supervised methods heavily rely on high-quality labeled data.

Recently, neural models have been widely applied to relation classification (Zeng et al. 2014; dos Santos, Xiang, and Zhou 2015; Mooney and Bunescu 2005; Yang et al. 2016) including convolutional neural networks, recursive neural network (Ebrahimi and Dou 2015; Liu et al. 2015), and long short-term memory network (Miwa and Bansal 2016; Xu et al. 2015; Miwa and Bansal 2016). In (Wang et al. 2016), two levels of attention is proposed in order to better discern patterns in heterogeneous contexts for relation classification.

In general, a large amount of labeled data are required to train neural models, which is quite expensive. To address this issue, distant supervision was proposed (Mintz et al. 2009) by assuming that all sentences that mention two entities of a fact triple describe the relation in the triple. In spite of the success of distance supervision, such methods suffer from the noisy labeling issue. To alleviate this issue, many studies formulated relation classification as a multi-instance learning problem (Riedel, Yao, and McCallum 2010; Hoffmann et al. 2011; Surdeanu et al. 2012; Zeng et al. 2015). In (Lin et al. 2016; Ji et al. 2017; Tianyu Liu and Sui 2017), a sentence-level attention mechanism over multiple instances was proposed and incorrect sentences can be down-weighted. However, such multi-instance learning models all predict relations at the bag level but not at the sentence level, and they can not deal with the bags in which all sentences are not describing a relation at all. There are other approaches to reduce the noise of distant supervision using active learning (Sterckx et al. 2014) and negative patterns (Takamatsu, Sato, and Nakagawa 2012).

Previous methods are all at the bag level but not at the sentence level and as such, they cannot find the exact mapping between a relation and a sentence. Furthermore, these methods are unable to handle the bags in which all the sentences are not describing the relation. To address these issues, we propose a new framework which first selects correct sentences in the framework of reinforcement learning (Sutton and Barto 1998; Narasimhan, Yala, and Barzilay 2016) and then predicts relations from each sentence in the cleansed data.

**Methodology**

We propose a new relation classification framework, which is able to select correct sentences from noisy data for better relation classification. The proposed framework can predict relations at the sentence level from the cleansed data, rather than at the bag level. Sentence-level prediction is more friendly to the tasks that need to comprehend sentences such as question answering and semantic parsing.

Our framework consists of two key modules: the instance selector which selects correct sentences from noisy data, and the relation classifier which predicts relation and updates its parameters with cleaned data. The two modules interacts with each other during the training process.

**Problem Definition**

Formally, we decompose the task of relation classification into two sub-problems in this paper: instance selection and relation classification.

We formulate the instance selection problem as follows: given a set of <sentence, relation label> pairs as \(X = \{(x_1, r_1), (x_2, r_2), \ldots, (x_n, r_n)\}\), where \(x_i\) is a sentence associated with two entities \((h_i, t_i)\) and \(r_i\) is a noisy relation label produced by distant supervision. The goal is to determine which sentence truly describes the relation and should be selected as a training instance.

The relation classification problem is formulated as follows: given a sentence \(x_i\) and the mentioned entity pair \((h_i, t_i)\), the goal is to predict the semantic relation \(r_i\) in \(x_i\). Essentially, the model estimates the probability: \(p_{\Phi}(r_i|x_i, h_i, t_i)\).
for relation classification. The state whether or not \( x \) current sentence \( i \) is the set of selected sentences, which is a subset of \( B \). We will introduce (i.e., \( \text{state}, \text{action}, \text{and reward} \)) as follows. To be clear, we will omit the superscript \( k \) which denotes the bag index. Thus, the formulation hereafter is based on only one bag.

**State.** The state \( s_i \) represents the current sentence, the already selected sentences, and the entity pair when making decision on the \( i \)-th sentence of the bag \( B \). We represent the state as a continuous real-valued vector \( F(s_i) \), which encodes the following information: 1) The vector representation of the current sentence, which is obtained from the non-linear layer of the CNN for relation classification; 2) The representation of the chosen sentence set, which are the average of the vector representations of all chosen sentences; 3) The vector representations of the two entities in a sentence, obtained from a pre-trained knowledge graph embedding table.

**Action.** We define an action \( a_i \in \{0, 1\} \) to indicate whether the instance selector will select the \( i \)-th sentence of the bag \( B \) or not. We sample the value of \( a_i \) by its policy function \( \pi_{\Theta}(s_i, a_i) \), where \( \Theta \) is the parameters to be learned. In this work, we adopt a logistic function as the policy function:

\[
\pi_{\Theta}(s_i, a_i) = \frac{e^{a_i \sigma(W \ast F(s_i) + b)}}{1 + e^{(1 - a_i)(1 - \sigma(W \ast F(s_i) + b))}}
\]

(1)

where \( F(s_i) \) is the state feature vector, and \( \sigma(\cdot) \) is the sigmoid function with the parameter \( \Theta = \{W, b\} \).

**Reward.** The reward function is an indicator of the utility of the chosen sentences. For certain bag \( B = \{x_1, \ldots, x_{|B|}\} \), we sample an action for each sentence, to determine whether the current sentence should be selected or not. We assume that the model has a terminal reward when it finishes all the selection. Therefore we only receive a delayed reward at the terminal state \( s_{|B|+1} \). The reward is zero at other states. Therefore, the reward is defined as follows:

\[
r(s_i | B) = \begin{cases} 
0, & i < |B| + 1 \\
\frac{1}{|B|} \sum_{x_i \in B} \log p(r | x_i), & i = |B| + 1
\end{cases}
\]

(2)

where \( B \) is the set of selected sentences, which is a subset of \( B \), and \( r \) is the relation label of bag \( B \). As shown in Figure 2.
2, \( p(r|x_j) \) is calculated by the relation classifier which is given by a CNN model. For the special case \( B = \emptyset \), we set the reward as the average likelihood of all sentences in the training data, which enables our instance selector to exclude noisy bag effectively.

Note that the relation classifier is at the sentence-level since it computes \( p(r|x) \) for each sentence. The reward is computed on a new bag of sentences selected by the instance selector. Essentially, the above reward evaluates the overall utility of all the actions made by the policy. It supervises the instance selector to maximize the average likelihood of the chosen instances, which makes the objective function of the instance selector consistent with the relation classifier.

In the selection process, not only the final action contributes to this reward, but also all the previous actions do. Therefore, this reward is delayed, and can be handled very well by reinforcement learning techniques (Sutton and Barto 1998).

**Optimization.** For a bag \( B \), we aim to maximize the expected total reward. More formally, our objective function is defined as

\[
J(\Theta) = V_\theta(s_1|B) = E_{s_1,a_1,s_2,...,s_i,a_i,s_i+1,...,s_{|B|+1}} \left[ \sum_{i=0}^{|B|+1} r(s_i|B) \right]
\]

where \( a_i \sim \pi_\theta(s_i,a_i), s_i+1 \sim P(s_{i+1}|s_i,a_i). \) The transition functions \( P(s_{i+1}|s_i,a_i) \) are equal to 1, since the state \( s_{i+1} \) is fully determined by the state \( s_i \) and \( a_i \). \( V_\theta(s_1|B) \) is the value function, and \( V_\theta(s_1|B) \) represents the expected future total reward that we can obtain by starting at certain state \( s_1 \) following policy \( \pi_\theta(s_i,a_i) \).

According to the policy gradient theorem (Sutton et al. 1999) and the REINFORCE algorithm (Williams 1992), we update the gradient in the following way. For each bag \( B \), we sample an action for each state sequentially according to the current policy. We then get a sampled trajectory \( \{ s_1, a_1, s_2, a_2, ..., s_{|B|}, a_{|B|}, s_{|B|+1} \} \) and a corresponding terminal reward \( r(s_{|B|+1}|B) \). Since we only have a non-zero terminal reward, the value function is the same for all states from \( s_1 \) to \( s_{|B|} \), namely \( v_i = V(s_i|B) = r(s_{i+1}|B), \) for \( i = 1, 2, ..., |B| \). We update the current policy using the following gradient:

\[
\Theta \leftarrow \Theta + \alpha \sum_{i=1}^{|B|} v_i \nabla \log \pi_\theta(s_i,a_i)
\]

(4)

**Relation classifier**

In the relation classifier, we adopt a CNN architecture to predict relations. The CNN network has an input layer, a convolution layer, a max pooling layer and a non-linear layer from which the representation is used for relation classification.

**Input layer.** For each sentence \( x \), we represent it as a list of vectors \( \mathbf{x} = (w_1, w_2, \ldots, w_m) \). Each representation vector consists of two parts: one is the word embedding; the other is the position embedding. Word embeddings are obtained from word2vec\(^3\), and the dimension is \( d^w \). Similar to

\[\text{ALGORITHM 1: Overall Training Procedure}\]

1. Initialize the parameters of the CNN model of relation classifier and the policy network of instance selector with random weights respectively
2. Pre-train the CNN model to predict relation \( r_i \) given the sentence \( x_i \) by maximizing \( \log p(r_i|x_i) \)
3. Pre-train the policy network by running Algorithm 2 with the CNN model fixed.
4. Run Algorithm 2 to jointly train the CNN model and the policy network until convergence

\[\text{ALGORITHM 2: Reinforcement Learning Algorithm for the Instance Selector}\]

**Input:** Episode number \( L \). Training data \( B = \{ B_1, B_2, \ldots, B_N \} \). A CNN and a policy network model parameterized by \( \Phi \) and \( \Theta \), respectively

**Initialize** the target networks as: \( \Phi' = \Phi, \Theta' = \Theta \)

for episode \( l = 1 \) to \( L \)

Shuffle \( B \) to obtain the bag sequence

\[ B = \{ B^1, B^2, \ldots, B^N \} \]

foreach \( B_k \in B \)

Sample instance selection actions for each data instance in \( B^k \) with \( \Theta' \):

(To be clear, we omit the superscript \( k \) below)

\[ A = \{ a_1, \ldots, a_{|B|} \}, \quad a_i \sim \pi_{\Theta'}(s_i,a_i) \]

Compute delayed reward \( r(s_{|B|+1}|B) \)

Update the parameter \( \Theta \) of instance selector:

\[ \Theta \leftarrow \Theta + \alpha \sum_{i=1}^{|B|} v_i \nabla \log \pi_\theta(s_i,a_i) \]

where \( v_i = r(s_{i+1}|B) \)

end

Update \( \Phi \) in the CNN model

Update the weights of the target networks:

\[ \Theta' \leftarrow \tau \Theta + (1 - \tau) \Theta' \]

\[ \Phi' \leftarrow \tau \Phi + (1 - \tau) \Phi' \]

end

(Zeng et al. 2014), we use \( d^p \)-dimensional position embeddings, which are vector representations of the relative distances from the current word respectively to the head or tail entities in this sentence. We concatenate the word and position embeddings of each word to form a new vector \( \mathbf{w}_i \in \mathbb{R}^d \), and \( d = d^w + 2 \times d^p \), and then input these vectors to the CNN model.

**CNN.** In order to obtain high-level and abstractive representation of the raw input of a sentence, we apply a CNN structure for relation classification. This can be briefly described as below:

\[ \mathbf{L} = \text{CNN}(\mathbf{x}) \]

(5)

where \( \mathbf{x} \) is the input vectors as described in the input layer and \( \mathbf{L} \in \mathbb{R}^{d^p} \) is the output of the max pooling layer. In this structure, there is a convolution layer, and a max pooling layer. The convolution operation is performed on 3 consecutive words, and the number of feature maps \( d^p \) is set to 230, the same as the setting of (Lin et al. 2016). Hence, the convolution parameters are \( \mathbf{W}_f \in \mathbb{R}^{d^p \times (3d)} \) and \( \mathbf{b}_f \in \mathbb{R}^{d^p} \).
Then, the probability for relation prediction \( p(r|x; \Phi) \) is given as follows:

\[
p(r|x; \Phi) = \text{softmax}(W_r \ast \text{tanh}(L) + b_r)
\]

(6)

where \( W_r \in \mathbb{R}^{n_r \times d'} \) and \( b_r \in \mathbb{R}^{n_r} \) are parameters in the fully-connected layer, \( n_r \) is the total number of relation types, and \( \Phi = \{W_f, b_f, W_r, b_r\} \).

The key difference between our relation classifier and other studies lies in that our classifier performs relation classification at the sentence level. The input to the relation classifier in other studies is a bag of sentences. Instead, the input is a sentence, instead of for each bag. For example, the task in Figure 1 needs to map the first sentence to relation "BornIn" and the second sentence to "EmployedBy".

Since the data obtained from distant supervision are noisy, we randomly chose 300 sentences and manually labeled the relation type for each sentence to evaluate the classification performance. We adopted accuracy and macro-averaged \( F_1 \) as the evaluation metric.

**Baselines.** We adopted three state-of-the-art baselines:

- **CNN** (Zeng et al. 2014) is a sentence-level classification model. It does not consider the noisy labeling problem.
- **CNN+Max** (Zeng et al. 2015) is a bag-level classification model. It assumes that there is one sentence describing the relation in a bag. It chooses the most correct sentence in each bag.
- **CNN+ATT** (Lin et al. 2016) is also a bag-level model, similar to CNN+Max. It adopts a sentence-level attention
<table>
<thead>
<tr>
<th>Method</th>
<th>Macro $F_1$</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN</td>
<td>0.40</td>
<td>0.60</td>
</tr>
<tr>
<td>CNN+Max</td>
<td>0.06</td>
<td>0.34</td>
</tr>
<tr>
<td>CNN+ATT</td>
<td>0.29</td>
<td>0.56</td>
</tr>
<tr>
<td>CNN+RL (ours)</td>
<td>0.42</td>
<td>0.64</td>
</tr>
</tbody>
</table>

Table 1: Performance on sentence-level relation classification.

over the sentences in a bag and thus can down weight noisy sentences in a bag.

CNN is a sentence-level model that is trained directly on noisy data. For bag-level models (CNN+Max and CNN+ATT), the training process is the same as the referenced papers. During test, each sentence is treated as a bag and a relation is predicted for each bag. In this scenario, the bag-level relation prediction is exactly the same as the sentence-level prediction. All the baselines were implemented with the source codes released by (Li et al. 2016).

**Results.** Results in Table 1 reveal the following observations.

- CNN+RL obtains superior performance than CNN, indicating that filtering noisy data by instance selection benefits the task.
- CNN+RL outperforms CNN+Max and CNN+ATT remarkably. It shows the effectiveness of instance selection with reinforcement learning.
- The sentence-level models (CNN and CNN+RL) perform much better than the bag-level models (CNN+Max and CNN+ATT), indicating that bag-level models do not perform well for sentence-level prediction.

**Instance Selection**

We then evaluated the effectiveness of our instance selector from several aspects. First, we evaluated whether the selected data by our instance selector are better for relation classification. Second, we justified the accuracy of selection decision in the selector by manually checking the decisions on sentences. Third, we compared the proposed RL selection strategy in our selector with greedy selection. Last, we assessed whether the selector has the ability of filtering those bags that contain all noisy sentences.

**Relation classification on selected data.** To measure the quality of the selected data by our instance selector, we performed relation classification experiments on the selected data. We first used our instance selector to select the high-quality sentences from the original data. Then, we trained two state-of-the-art models, CNN and CNN+ATT with two settings. One setting is to train them on the original data, named as CNN(Original) and CNN+ATT(Original). The other setting is to train them on the selected data, which are named as CNN(Selected) and CNN+ATT(Selected). We compared the performance of CNN(Original) (CNN+ATT(Original)) with CNN(Selected) (CNN+ATT(Selected)) on the relation classification task. The results are compared under the held-out evaluation configuration (Mintz et al. 2009) which provides an approximate measure of relation classification without expensive human annotations. The held-out evaluation compares the predicted relational fact from the test data with the facts in Freebase, but it does not consider the mapping between a relational fact and a sentence.

As shown in Figure 3 and Figure 4, the models trained on the selected data achieve much better performance than the counterparts trained on the original dataset. The results also indicate our instance selector has the ability of filtering out noisy sentences and distilling high-quality sentences, resulting better classification performance.

**Accuracy of instance selection decision.** To assess how accurate the decision is by the instance selector, we manually checked each sentence selected and rejected by the instance selector in a sampled dataset. For each sentence, the instance selector makes a correct decision if the sentence’s label is correct and our instance selector selects it as a training instance, or, if its label is wrong and our instance selector rejects it. Otherwise, we judged that the instance selector makes a wrong decision.

Specifically, we sampled 300 sentences from the training data. Our instance selector chooses 64 sentences as the training instances, among which 45 sentences are correctly selected. The selector also rejects 236 instances, and 177 of them are noisy instances (not describing the relation). To summarize, the accuracy of our instance selector is $(45 + 177)/300 = 74\%$, which demonstrates the effectiveness of our instance selector.

**Different instance selection strategies.** To show the ne-
Table 2: Instance selection examples by different models. For CNN+RL and CNN+Max, 1 or 0 means the sentence is selected or not. For CNN+ATT, the value is the attention weight.

<table>
<thead>
<tr>
<th>Bag I</th>
<th>Entity Pair: fabrice_santoro, france; Relation:people/person/nationality)</th>
<th>CNN+RL</th>
<th>CNN+ATT</th>
<th>CNN+Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>though not without some struggle, federer, the world’s top-ranked player, advanced to the fourth round with a thrilling, victory over the crafty fabrice_santoro of france, who is ranked 76th.</td>
<td>1</td>
<td>0.60</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>in his quarterfinal, nalbandian overwhelmed unseeded fabrice_santoro of france</td>
<td>1</td>
<td>0.39</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>fabrice_santoro, 33, of france finally reached the quarterinals in a major on his 54th attempt by defeating the 11th-seeded spaniard david ferrer</td>
<td>1</td>
<td>0.01</td>
<td>0</td>
</tr>
<tr>
<td>Bag II</td>
<td>Entity Pair: jonathan_littel, france; Relation:people/person/nationality)</td>
<td>jonathan_littel, a new york-born writer whose french-language novel about a murderous and degenerate officer has been the sensation of the french publishing season, on monday became the first american to win france’s most prestigious literary award, the prix goncourt</td>
<td>0</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>after a languid intercontinental auction that stretched for more than a week, the american rights to jonathan_littel’s novel les bienveillantes, which became a publishing sensation in france, have been sold to harpercollins, the publisher confirmed yesterday.</td>
<td>0</td>
<td>0.11</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 5: Comparison of instance selection with reinforcement learning against greedy selection.

In this paper, we propose a novel model for sentence-level relation classification from noisy data using a reinforcement learning framework. The model consists of an instance selector and a relation classifier. The instance selector chooses high-quality data for the relation classifier. The relation classifier predicts relation at the sentence level and provides rewards to the selector as a weak signal to supervise the instance selection process. Extensive experiments demonstrate that our model can filter out the noisy sentences and perform sentence-level relation classification better than state-of-the-art baselines from noisy data.

Further, our solution for instance selection can be generalized to other tasks that employ noisy data or distant supervision. For instance, a possible attempt might be to perform sentiment classification on noisy data (Go, Bhayani, and Huang 2009). We leave this as our future work.

Acknowledgement

This work was partly supported by the National Science Foundation of China under grant No.61272227/61332007.

References


Miwa, M., and Bansal, M. 2016. End-to-end relation extraction using lstms on sequences and tree structures. In ACL.


