Computer School of Wuhan University, Wuhan China
Natural Language Processing

Reinforcement Learning for Relation Classification from Noisy Data

AAAI-2018

Weilong Hu
huweilong@whu.edu.cn

April 20, 2018

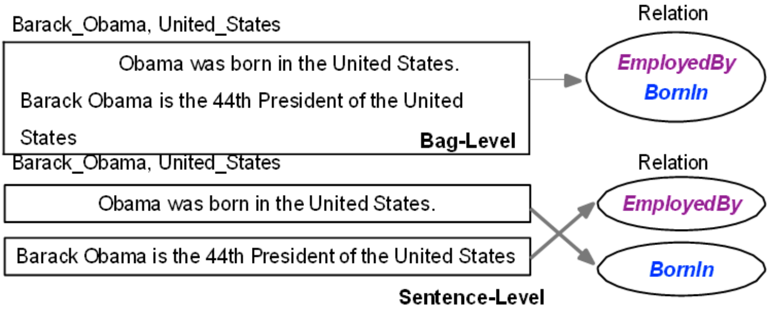
Content

- Most supervised methods require high-quality annotated data.

- Most supervised methods require high-quality annotated data.
- Distant supervision suffers from the `noisy labeling problem`.

- Most supervised methods require high-quality annotated data.
- Distant supervision suffers from the `noisy labeling problem`.
- Multi-instance learning suffers from two limitations:
  1. `Unable to handle the setence-level prediction.`
  2. `Sensitive to the bags with all noisy sentences which don't describe a relation at all.`

-



- 53% out of 100 sample bags have no valid sentences.

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 cleaned 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

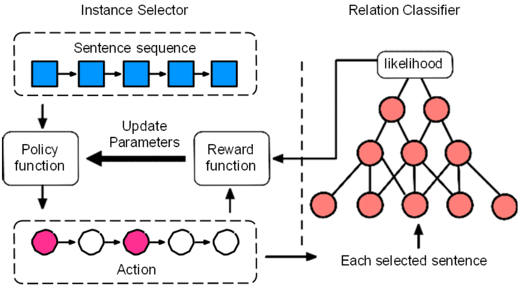
Supervised neural models:

- convolutional neural networks
- recursive neural network
- long short-term memory network
- two levels of attention

Distant supervised methods:

- multi-instance learning
- a sentence-level attention mechanism over multiple instances
- active learning
- negtive patterns

- instance selection problem:
  $X = \{(x_1, r_1), (x_2, r_2), \ldots, (x_n, r_n)\}$, where x_i is a sentence associated with (h_i, t_i)
- relation classification problem:
  estimate the probability of $p_\Phi(r_i | x_i, h_i, t_i)$



- Each sentence x_i has a action α_i to indicate whether or not x_i will be selected as a training instance
- The state s_i is represented by the current sentence x_i, the already chosen sentences among $\{x_1, \cdots, x_{i-1}\}$, and the entity pair h_i and t_i in sentence x_i.
- The instance selector samples an action given the current state according to a stochastic policy

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.

Real-valued vector $\mathbf{F}(s_i)$ encodes the following information:

- The non-linear layer of the CNN
- The average of the vector representations of all chosen sentences
- A pre-trained knowledge graph embedding table

The action $\alpha_i \in \{0, 1\}$

We sample α_i by its policy function $\pi_\Theta(s_i, \alpha_i)$

$$\begin{aligned}\pi_\Theta &= P_\Theta(\alpha_i | s_i) \\ &= \alpha_i \sigma(W * F(s_i) + b) \\ &+ (1 - \alpha_i)(1 - \sigma(W * F(s_i) + b))\end{aligned}$$

where $\Theta = \{W, b\}$

$B = \{x_1, \cdots, x_{|B|}\}$, the model has a terminal reward at terminal state $s_{|B|+1}$ when it finishes all the selection.

The reward is defined as follows:

$$r(s_i|B) = \begin{cases} 0 & \text{i < |B| + 1} \\ \frac{1}{|\hat{B}|} \sum_{x_j \in \hat{B}} \log p(r|x_j) & \text{i = |B| + 1} \end{cases}$$

where $\hat{B}$ is the set of selected sentences, r is the relation label of bag B. For $\hat{B} = \emptyset$, reward = the average likelihood of all sentences.

The above reward evaluates the overall utility of all the actions made by the policy, which supervises the instance selector to maximize the average likelihood of the chosen instances.

For a bag B, we aim to maximize the total reward. Objective function is defined as

$$
\begin{aligned}
J(\Theta) &= V_{\Theta}(s_0|B) \\
&= E_{s_0,\alpha_0,s_1,\ldots,s_i,\alpha_i,s_{i+1},\ldots}[\sum_{i=1}^{|B|+1} r(s_i|B)]
\end{aligned}
$$

where $\alpha_i \sim \pi_{\Theta}(s_i, \alpha_i)$, $s_{i+1} \sim P(s_{i+1}|s_i, \alpha_i) = 1$, since s_{i+1} is fully determined by s_i and α_i. V_{Θ} is the value function, and $V_{\Theta}(s_0|B)$ represents the expected future total reward that we can obtain by starting at certain state s_0 following policy $\pi_{\Theta}(s_i, \alpha_i)$.

$$
\Theta \leftarrow \Theta + \alpha \sum_{i=1}^{|B|} v_i \nabla_{\Theta} \log \pi_{\Theta}(s_i, \alpha_i)
$$

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-liner layer.

Loss function:

$$\mathcal{J}(\Theta) = -\frac{1}{|\hat{X}|}\sum_{i=1}^{|\hat{X}|}\log p(r_i|x_i;\Phi)$$

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

ALGORITHM 2: Reinforcement Learning Algorithm for the Instance Selector

Input: Episode number L. Training data $\mathrm{B} = \{B^1, B^2, \ldots, B^N\}$. A CNN and a policy network model parameterized by Φ and Θ, respectively

Initialize the target networks as: $\Phi' = \Phi, \Theta' = \Theta$

for *episode* $l = 1$ *to* L **do**

  Shuffle B to obtain the bag sequence $\mathrm{B} = \{B^1, B^2, \ldots, B^N\}$

  foreach $B^k \in \mathrm{B}$ **do**

    Sample instance selection actions for each data instance in B^k with Θ':

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

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

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

    Update the parameter Θ of instance selector:

    $\Theta \leftarrow \Theta + \alpha \sum_i v_i \nabla_\Theta \log \pi_\Theta(s_i, a_i)$, where

    $v_i = r(s_{|B|+1}|B)$

  end

  Update Φ in the CNN model

  Update the weights of the target networks:

  $\Theta' = \tau\Theta + (1-\tau)\Theta'$

  $\Phi' = \tau\Phi + (1-\tau)\Phi'$

end

Dataset:

	sentences	entity pairs	relational facts
training	522,611	281,270	18,252
testing	172,448	96,678	1,950

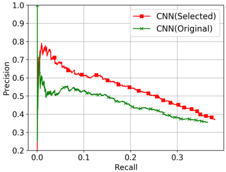
Word and entity embedding:

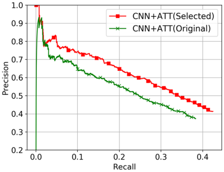
`word2vec and TransE model`

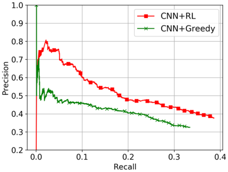
Experiment

Experiment Results

Method	Macro F_1	Accuracy
CNN	0.40	0.60
CNN+Max	0.06	0.34
CNN+ATT	0.29	0.56
CNN+RL(ours)	**0.42**	**0.64**







Bag I (**Entity Pair**: fabrice_santor, france; **Relation**:/people/person/nationality)	CNN+RL	CNN+ATT	CNN+Max
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.	1	0.60	0
in his quarterfinal , nalbandian overwhelmed unseeded **fabrice_santoro** of **france**	1	0.39	1
fabrice_santoro, 33 , of **france** finally reached the quarterfinals in a major on his 54th attempt by defeating the 11th-seeded spaniard david ferrer	1	0.01	0
Bag II (**Entity Pair**: jonathan_littel, france; **Relation**:/people/person/nationality)			
jonathan_littell, 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	0	0.89	1
after a languid intercontinental auction that stretched for more than a week, the american rights to **jonathan_littell**'s novel les bienveillantes, which became a publishing sensation in **france**, have been sold to harpercollins, the publisher confirmed yesterday.	0	0.11	0

Thank you for listening my presentation!