PDTB-Ji

1.Introduction

PDTB itself provides a dataset and annotates it, but the difficulty remains in automatically identifying implicit relationships.

Example:

Bob gave Tina the burger.

She was hungry.

We add “because”.

applying a discriminatively-trained model of compositional distributed semantics to discourse relation classification

The discourse relation can be predicted as a bilinear combination of these vector representations.

combined with a small number of surface features, better than PDTB

To address the problem of the confusing relationship between entities and roles, we compute vector representations with each discourse argument and also for each entity description. The aim is to capture the roles played by entities in the text.

In short, a combination of surface features, distributed representations of discursive arguments and distributed representations of entities

2. Entity augmented distributed semantics

2.1 Upward pass: argument semantics

a feed-forward “upward” pass: each non-terminal in the binarized syntactic parse tree has a K-dimensional vector representation that is computed from the representations of its children, bottoming out in pre-trained representations of individual words.

Using RNN

For each parent node i:

Leaves: pre-trained word vector representations

feedforward, no cycles and all nodes can be computed in linear time.

2.2 Downward pass: entity semantics

we augment the representation of each argument with additional vectors, representing the semantics of the role played by each coreferent entity in each argument.

Additional distributed vectors

its parent ρ(i), and its siblings(i).upward vector of the sibling us(i)

algorithm is designed to maintain the feedforward nature of the neural network, so that we can efficiently compute all nodes without iterating.

the upward and downward passes are each feedforward

finish in time that is linear in the length of the input

3. Predicting discourse relations

Deciding function

avoid overfitting, we apply a lowdimensional approximation to each Ay,

Surface features

4. Large-margin learning framework

Two things to learn:

the classification parameters

the composition parameters

define a large margin objective

use backpropagation to learn all parameters of the network

using stochastic gradient descent

final learning method:

a single argument pair (m, n) with the gold discourse relation y\*

objective function:

4.1 Learning the classification parameters

In the objective function,

Otherwise…… <

The gradient for the classification parameters therefore depends on the margin value between gold label and all other labels.

4.2 Learning the composition parameters

two composition matrices U and V, corresponding to the upward and downward composition procedures

5.Implementation

Learning: used AdaGrad to tune the learning rate in each iteration

To avoid the exploding gradient problem, we used the norm clipping trick proposed by Pascanu et al., fixing the norm threshold at = 5.0.

Hyperparameters: 3 tunable hyperparameters: the latent dimension K for the distributed representation, the regularization parameter , and the initial learning rate .

for the latent dimensionality, for the regularization (on each training instance), and for the learning rate.

Initialization: All the classification parameters are initialized to 0.

follow Bengio (2012) and initialize U and V with uniform random values drawn from the range

Word representations

a word2vec model on the PDTB corpus, standardizing the induced representations to zeromean, unit-variance

Syntactic structure

We run the Stanford parser to obtain constituent parse trees of each sentence in the PDTB, and binarize all resulting parse trees

Coreference

The impact of entity semantics on discourse relation detection is inherently limited by two factors: (1) the frequency with which the arguments of a discourse relation share coreferent entity mentions, and (2) the ability of automated coreference resolution systems to detect these coreferent mentions.

improvements in resolution

Additional features

using additional surface features proposed by Lin et al.. These include four categories: word pair features, constituent parse features, dependency parse features, and contextual features.