Relation Classification via Variational Autoencoder Multi-task Learning

Yixuan Su

Department of Engineering, University of Cambridge, Cambridge, U.K.

UNIVERSITY OF CAMBRIDGE

Objectives

Proposing general classifier, encoder-decoder ensemble framework to perform classification on various tasks. By using proposed framework, we demonstrate the following facts:

- Introducing encoder-decoder framework into classifier can further boost classifier performance and achieve on par state-of-the-art result on natural language processing (NLP) task.
- Using trained ensemble model we can generate new reasonable samples for different relation types.

Introduction

Relation classification is a crucial ingredient in numerous information extraction systems seeking to mine structured facts from text. We propose a novel ensemble model for this task. By combining classifier along with variational autoencoder we are able to produce on par state-of-the-art performance on SemEval 2010 Task 8 dataset [1] and KBP37 dataset [2].

Methodology

0.1 Long Short Term Memory Networks (LSTM)

For dealing with sentence in NLP task, it is important to model the long term dependency across all words contained in the sentence. Thus LSTM network is a proper choice of model for this goal. As a result, we use LSTM , described in Equation (1), as basic component for both classifier and variational autoencoder in the ensemble model.

$$\begin{aligned} i_t &= \sigma(W_{wi}w_t + W_{hi}h_{t-1}) \\ f_t &= \sigma(W_{wf}w_t + W_{hf}h_{t-1}) \\ o_t &= \sigma(W_{wo}w_t + W_{ho}h_{t-1}) \\ \hat{c}_t &= tanh(W_{wc}w_t + W_{hi}h_{t-1}) \\ c_t &= f_t \otimes c_{t-1} + i_t \otimes \hat{c}_t \\ h_t &= o_t \otimes tanh(c_t) \end{aligned} \tag{1}$$

0.2 Variational Autoencoder

The Variational Autoencoder (VAE) is a generative latent variable model for data in which $\mathbf{z_i} \sim \mathbf{N(0, I)}$ and $\mathbf{x_i} \sim \mathbf{p_\theta(x_i|z_i)}$. Because directly modelling the conditional probability is always intractable, thus a approximate conditional distribution $q_\phi(\mathbf{x_i}|\mathbf{z_i})$ is used, where this conditional is always parameterised by neural network models. To train VAE model, the objective function takes the form of Equation (2), where KL stands for KL-Divergence.

$$L(\theta; x) = -KL(q_\theta(z|x)||p(z)) + \mathbb{E}_{q_\phi(z|x)}[logp_\theta(x|z)] \leq logp(x) \tag{2}$$

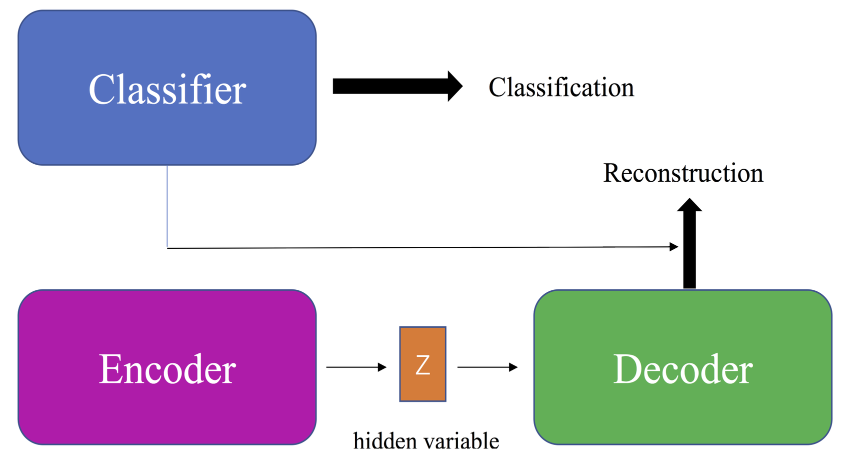

Figure 1: CVAE Model Framework

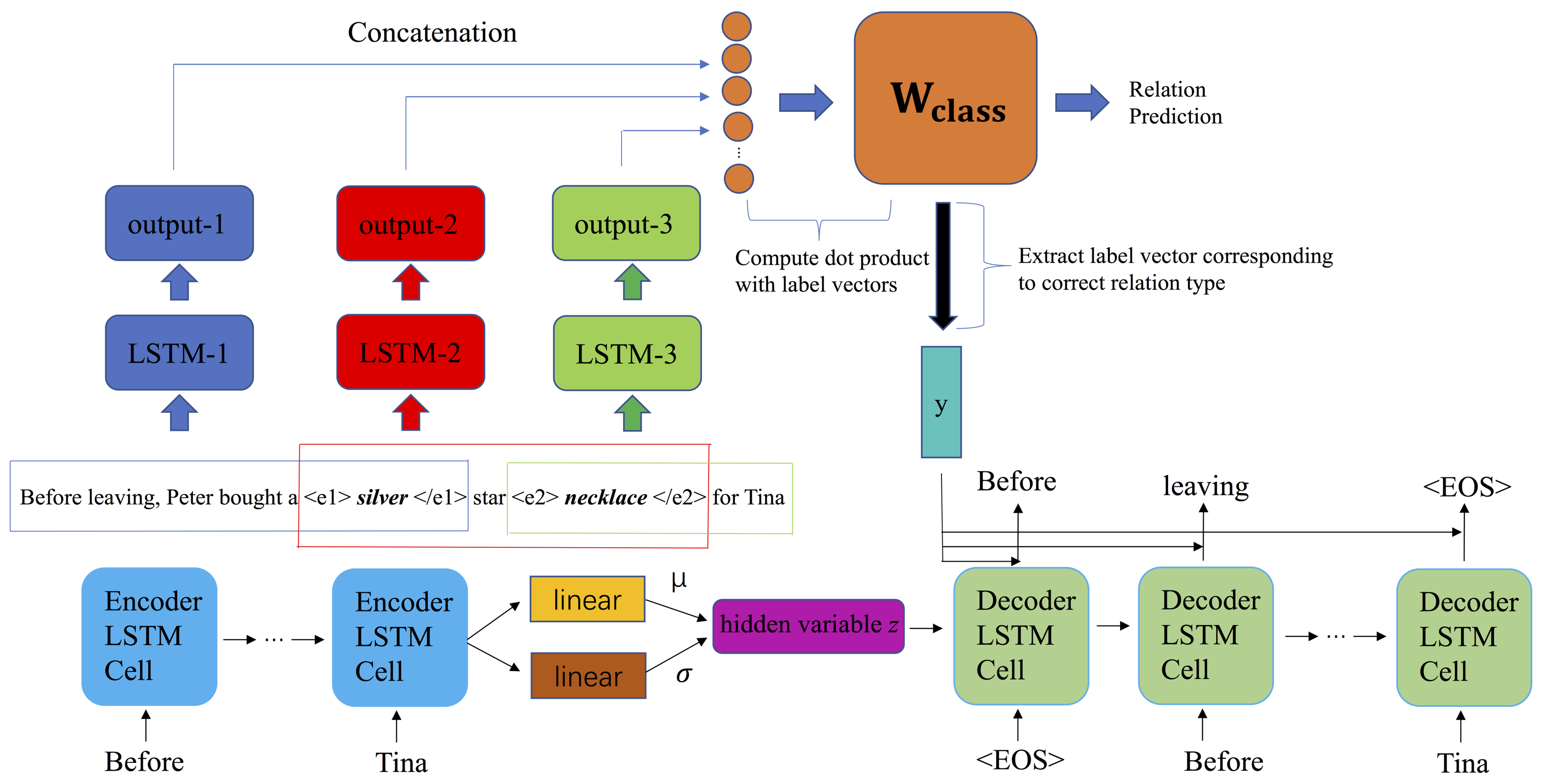

Figure 2: Detailed Model

0.3 CVAE Model

To build Class VAE (CVAE) model, we introduce class lable $\mathbf{y}$ into the whole framework and make assumptions that conditional distribution of hidden variable is independent of class label as described in Equation (3). And conditional distribution of data is dependent both on hidden variable and class lable which is modeled as $q_\phi(x|z, y)$. As a result, the objective function becomes Equation (4).

$$q_\theta(z|x, y) = q_\theta(z|x) \tag{3}$$

$$L(\theta; x) = -KL(q_\theta(z|x)||p(z)) + \mathbb{E}_{q_\phi(z|x)}[logp_\theta(x|z, y)] \leq logp(x) \tag{4}$$

The proposed ensemble model framework and detailed model are shown in Figure 1 and 2.

Experiments

Two relation datasets are experimented with which are Semeval 2010 Task 8 (19 relations) [1] and KBP37 (37 relations) [2]. System performance from only classifier and ensemble model are reported in Table 1, the performance is measured by marco F1 score across all relation types.

	Classifier	CVAE Model	state-of-the-art
Semeval 2010 Task 8	84.4	86.3	**88.0**
KBP37	59.1	**60.7**	58.8

Table 1: F1 score of different models on different dataset

In the next experiment, we use ensemble model to generate new sentence under different relation types which are shown in Table 2.

Relation Type	Generated Sentence
Cause-Effect(e1,e2)	the earthquake caused the destruction of death
Cause-Effect(e2,e1)	the disruption caused by the earthquake
Content-Container(e1,e2)	the box was contained in a box
Content-Container(e2,e1)	i found a bottle full of the water
Entity-Destination(e1,e2)	the government has been added into the country
Entity-Destination(e2,e1)	the book was in a large of chapters

Table 2: Generated sentence under different relation types

In Figure 3, we show the T-SNE visualization of classifier output under different type of inputs. And clear boundary can be seen between different relation types. Showing the classifier can learn informative knowledge.

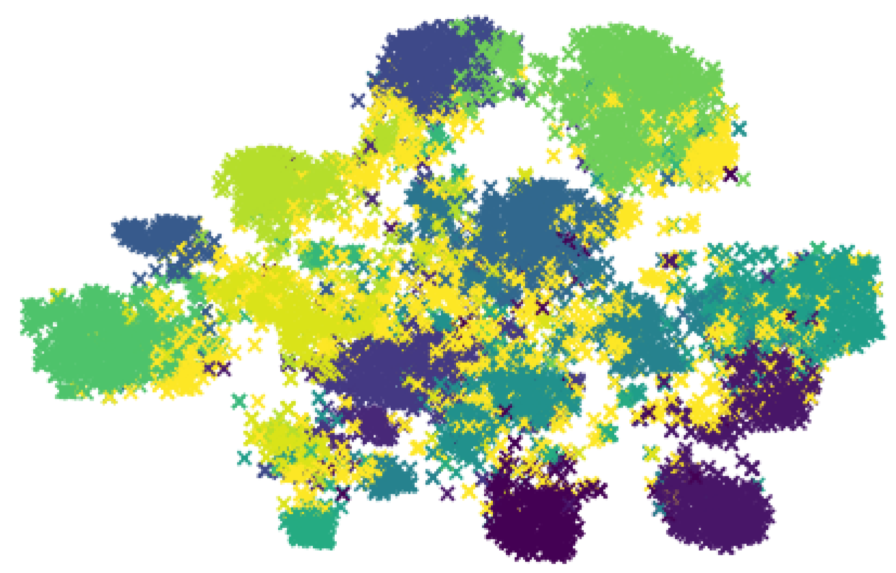

Figure 3: T-SNE visualization of classifier output of different relations

Results and Discussion

From the results, we see that when performing classification task, leveraging VAE along with classifier can further improve the result. The reason is that VAE serves as a kind of regularization to the classifier by forcing the model to reconstruct original data from learnt feature. The generated samples also prove our statement and the usefulness of our model.

References

[1] I. Hendrickx, N. K. Su, Z. Kozareva, P. Nakov, M. Pennacchiotti, L. Romano, and S. Szpakowicz, "Semeval-2010 task 8: multi-way classification of semantic relations between pairs of nominals," in *The Workshop on Semantic Evaluations: Recent Achievements and Future Directions*, pp. 94–99, 2009.

[2] D. Zhang and D. Wang, "Relation classification via recurrent neural network," *Computer Science*, 2015.