Relation Classification via Variational Autoencoder Multi-task Learning

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

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.

\[ h_t = \sigma(W_{wx}x_t + W_{hh}h_{t-1}) \]
\[ f_t = \sigma(W_{wf}W_{w}x_t + W_{bf}h_{t-1}) \]
\[ o_t = \sigma(W_{wo}W_{w}x_t + W_{ho}h_{t-1}) \]
\[ c_t = \tanh(W_{wc}W_{w}x_t + W_{hc}h_{t-1}) \]
\[ h_t = o_t \odot \tanh(c_t) \]  (1)

0.2 Variational Autoencoder
The Variational Autoencoder (VAE) is a generative latent variable model for data in which \( z_i \sim N(0,1) \) and \( x_i \sim p_i(x_i|z_i) \). Because directly modelling the conditional probability is always intractable, thus a approximate conditional distribution \( q_i(x_i|z_i) \) 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.

\[ L(\theta, x) = -KL(q_i(x_i|z_i)||p_i(z_i)) + E_{q_i(x_i|z_i)}[\log p_i(x_i|z_i)] \leq \log p_i(x) \]  (2)

0.3 CV AE Model
To build Class VAE (CVAE) model, we introduce class label \( y \) 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 label which is modeled as \( q_i(x_i|z_i, y) \). As a result, the objective function becomes Equation (4).

\[ q_i(z_i|y) = q_i(z_i) \]
\[ L(\theta, x, y) = -KL(q_i(z_i|y)||p_i(z_i)) + E_{q_i(x_i|z_i, y)}[\log p_i(x_i|z_i, y)] \leq \log p_i(x) \]  (3)

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.

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

Table 1: F1 score of different models on different dataset

<table>
<thead>
<tr>
<th>Relation Type</th>
<th>Generated Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cause-Effect(e1,e2)</td>
<td>the earthquake caused the destruction of death</td>
</tr>
<tr>
<td>Cause-Effect(e2,e1)</td>
<td>the disruption caused by the earthquake</td>
</tr>
<tr>
<td>Content-Container(e1,e2)</td>
<td>the box was contained in a box</td>
</tr>
<tr>
<td>Content-Container(e2,e1)</td>
<td>I found a bottle full of the water</td>
</tr>
<tr>
<td>Entity-Destination(e1,e2)</td>
<td>the government has been added into the country</td>
</tr>
<tr>
<td>Entity-Destination(e2,e1)</td>
<td>the book was in a large of chapters</td>
</tr>
</tbody>
</table>

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 1: CV AE Model Framework

Figure 2: Detailed Model

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