Cross-Lingual Sentiment Classification Based on Denoising Autoencoder

Huiwei Zhou, Long Chen, and Degen Huang

Dalian University of Technology
Natural Language Processing laboratory
Outline

- Introduction
- Denoising Autoencoder (DAE)
- The Combination CLSC Approach Based on Denoising Autoencoder
- Experiments and Analysis
- Conclusion and Future Work
Outline

- Introduction
- Denoising Autoencoder (DAE)
- The Combination CLSC Approach Based on Denoising Autoencoder
- Experiments and Analysis
- Conclusion and Future Work
Sentiment Classification

- Sentiment classification technique is the task of predicting sentiment polarity for a given text.

- Generally, sentiment classification approaches can be divided into two categories:
  - Lexicon based approach
  - Machine learning based approach
Labeled data are very imbalanced

The lack of sentiment resources limits the research progress in some languages.

Cross-Lingual Sentiment Classification (CLSC).

leverage resources on one language (source language) to resource-poor language (target language) for improving the classification performance on target language
The problems to be settled in CLSC task:
- Machine translation services adopted in CLSC task, bring translation errors in training process
- The language gap between source language and target language also influences the performance of system
Decrease the effects of noisy examples

- **To reduce the effect of translation errors**
  - Denoising autoencoder is adopted

- **To eliminate the language gap**
  - Two classifiers are trained in English view and Chinese view respectively, and the final results are combined from two classification outputs
Outline

• Introduction
• Denoising Autoencoder (DAE)
• The Combination CLSC Approach Based on Denoising Autoencoder
• Experiments and Analysis
• Conclusion and Future Work
Autoencoder

\[ L_H(x, z) = H(B_x \parallel B_z) = - \sum_{k=1}^{d} [x_k \log z_k + (1-x_k) \log(1-z_k)] \]
Outline

- Introduction
- Denoising Autoencoder (DAE)
- The Combination CLSC Approach Based on Denoising Autoencoder
- Experiments and Analysis
- Conclusion and Future Work
The Combination CLSC Approach Based on Denoising Autoencoder

- **Training phase**
  - English training set $D_{el}$
  - English classifier based on denoising autoencoder $M_{e}$
  - Chinese classifier based on denoising autoencoder $M_{c}$

- **Classification phase**
  - Chinese-to-English test set $D_{elu}$
  - Machine translation (Chinese-to-English)
  - Machine translation (English-to-Chinese)

- **Combination**
  - English results
  - Chinese results
  - final results

- **English View**
  - English-to-Chinese training set $D_{el}$

- **Chinese View**
  - Chinese test set $D_{elu}$
The Advantages of This Approach

- Denoising autoencoder is adopted to reduce the impacts of training errors.

- Training classifiers in multi-views helps to bridge the gap between English and Chinese.
Feature Setting

- Sentiment Word Features Selection
  - High-Frequency Words Method
  - CHI Statistical Method

\[ \chi^2(t_i, C_j) = \frac{N \times (A \times D - B \times C)^2}{(A + C) \times (B + D) \times (A + B) \times (C + D)} \]

- A: \((t_i, C_i)\)
- B: \((t_i, \overline{C}_i)\)
- C: \((\overline{t}_i, C_i)\)
- D: \((\overline{t}_i, \overline{C}_i)\)
Feature Setting

- Negation Features

\[ \text{vector} = (\text{neg}_1, \text{sent}_1, ..., \text{neg}_i, \text{sent}_i, ..., \text{neg}_{2000}, \text{sent}_{2000}) \]

- Feature Weight Calculation
  - Boolean Method
  - Word Frequency Method
  - TF-IDF method

\[ w_{ij} = tf_{ij} \times \log \frac{N}{n_i} \]
Outline

- Introduction
- Denoising Autoencoder (DAE)
- The Combination CLSC Approach Based on Denoising Autoencoder
- Experiments and Analysis
- Conclusion and Future Work
Experimental Settings

- **Dataset**
  - NLP&CC 2013 cross-lingual sentiment classification dataset, including book, DVD and music categories

- **Evaluation method**
  - category accuracy
  - average accuracy

- **Model parameters**
  - architecture: 4000-500-2
  - epoch: 30
  - learning rate: 0.1

\[
Accuracy_c = \frac{\text{#system_correct}}{4000}
\]

\[
Accuracy = \frac{1}{3} \sum_{i=1}^{3} Accuracy_c
\]
## Table 1. Effect of Sentiment Word Features Selection

<table>
<thead>
<tr>
<th>System</th>
<th>Methods</th>
<th>Book</th>
<th>DVD</th>
<th>Music</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>High-frequency</td>
<td>74.53%</td>
<td>75.43%</td>
<td>73.8%</td>
<td>74.58%</td>
</tr>
<tr>
<td></td>
<td>CHI statistic</td>
<td>73.03%</td>
<td>76.93%</td>
<td>75.15%</td>
<td>75.04% (+0.46%)</td>
</tr>
<tr>
<td>Chinese</td>
<td>High-frequency</td>
<td>78.40%</td>
<td>74.45%</td>
<td>73.15%</td>
<td>75.33%</td>
</tr>
<tr>
<td></td>
<td>CHI statistic</td>
<td>78.15%</td>
<td>75.05%</td>
<td>74.30%</td>
<td>75.83% (+0.50%)</td>
</tr>
</tbody>
</table>
Effect of Negation Features

Fig. 3. Performance comparison with or without negation features
Effect of Feature Weight Calculation Methods

Fig. 4. Performance comparison with different weight calculation methods
# Performance of Combination CLSC Systems

Table 2. Performance of combination CLSC systems

<table>
<thead>
<tr>
<th>System</th>
<th>Book</th>
<th>DVD</th>
<th>Music</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>English system</td>
<td>73.03%</td>
<td>76.93%</td>
<td>75.15%</td>
<td>75.04%</td>
</tr>
<tr>
<td>Chinese system</td>
<td>78.15%</td>
<td>75.05%</td>
<td>74.30%</td>
<td>75.83%</td>
</tr>
<tr>
<td>Combination system</td>
<td>79.68%</td>
<td>78.33%</td>
<td>78.08%</td>
<td>78.70%</td>
</tr>
</tbody>
</table>
Effect of Destruction Fraction in Denoising Autoencoders

Fig. 5. Accuracy vs. Destruction fraction
Comparison with Related Work

Table 3. CLSC performance comparison on the NLP&CC 2013 Share Task test data

<table>
<thead>
<tr>
<th>Team</th>
<th>Book</th>
<th>DVD</th>
<th>Music</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chen et al. 2014</td>
<td>77.00%</td>
<td>78.33%</td>
<td>75.95%</td>
<td>77.09%</td>
</tr>
<tr>
<td>HLT-Hitsz</td>
<td>78.50%</td>
<td>77.73%</td>
<td>75.13%</td>
<td>77.12%</td>
</tr>
<tr>
<td>Gui et al. 2013</td>
<td>78.70%</td>
<td>79.65%</td>
<td>78.30%</td>
<td>78.89%</td>
</tr>
<tr>
<td>Our Approach</td>
<td>80.63%</td>
<td>80.95%</td>
<td>78.48%</td>
<td>80.02%</td>
</tr>
</tbody>
</table>
Outline

- Introduction
- Denoising Autoencoder (DAE)
- The Combination CLSC Approach Based on Denoising Autoencoder
- Experiments and Analysis
- Conclusion and Future Work
Conclusion

- Denoising autoencoder and combination approach could improve the sentiment classification performance.
  - decrease the impacts of translation errors
  - eliminate the language gap
- The feature setting of CHI feature selection method together with TF-IDF weight calculation method works well on CLSC task.
Future Work

- Deep combination of classifiers rather than linear combination only
  - Such as co-training or transfer learning strategy
- Select high-quality translated reviews for training Chinese classifier to further reduce the impacts of translation errors.
Thank you!