Project Proposal
Chinese Natural Language Inference with LSTM

Yingbin Zhao
Department of Computer Science
Cornell University
yz672@cornell.edu

Mingtong Han
School of Applied & Engineering Physics
Cornell University
mh2242@cornell.edu

Abstract

Identifying entailment and similarity relationship between two sentences is an important building block to a variety of NLP applications, such as news event identification, semantic representation and summarization. Recent research using variations of Long Short-Term Memory Networks (LSTMN) on Stanford Natural Language Inference (SNLI) corpus show promising improvements over feature-based models for textual entailment. However, the SNLI corpus consists of only sentence pairs describing Flickr images in experiment context, and the generalizability of LSTMN on real world setting remains unexplored. In this project, we plan to implement LSTMN on BosonNLP News Language Inference (BosonNLI) corpus with sentences from business news articles in Chinese, and compare its performance with feature-based classifiers.

1 Motivation and Dataset

Large amounts of news articles are published online each day describing events happening around the world. The ability to identify entailment and similarity among sentences in these articles is critical to a lot of real world NLP scenarios such as summarization across many articles to an event and news search re-ranking to present the most informative and diversified results at the first page.

Provided a pair of premise and hypothesis, a typical natural language inference (NLI) task classify the pair in three classes: entailment, neutral and contradiction. For example, Stanford Natural Language Inference (SNLI) corpus [Bowman et al.’15] is a dataset balanced among the three labels.

SNLI is to date the largest and non-synthetic dataset for recognizing textual entailment. By constraining each sentence pair to describe the scenario of a given Flickr image, it obtains specific premises and hypotheses that has less ambiguity. Many models such as [Wang & Jiang’15] and [Cheng et al.’16] achieved promising results with test accuracy above 85% on this dataset. Therefore, it is interesting to explore how well the state-of-the-art NLI models perform in understanding language in real world settings such as news articles, and what improvements can be made. This vision motivates the creation of BosonNLP News Language Inference (BosonNLI) dataset and the work in this proposal.

BosonNLI defines a variation of NLI tasks with four classes: A_entails_B, B_entails_A, similar and neutral. This definition is different from the SNLI setting in three ways. First, the original entailment label becomes two labels, A_entails_B and B_entails_A, representing entailment in different direction. This settings treats the identification of premise and
hypothesis as part of the learning task, simply as they cannot be obtained directly from news articles. Second, a new label **similar** is added for sentence pairs that are almost equivalent to each other, where entailment holds in both direction. Third, **contradiction** is merged into **neutral** because practically the proportion of contradicting statements among news articles are relatively small, and this can be treated as a separate learning task in the future.

Examples from BosonNLI are shown in Table 1.

2 Approach

LSTMN is a form of recurrent neural network (RNN) capable of storing information and learning long-term dependencies. Compared to traditional RNN, LSTMN is constructed by a more complicated type of cell called Long Short-Term Memory (LSTM) which consists of four interacting components to control the flow of information (including cell state and output) regarding what to keep from previous cell state, what to update in the new cell state and what to output.

Among all research on the SNLI corpus, the top records are dominated by LSTMN based methods, and the best result to date is given by an extension of LSTMN with neural attention and word-by-word attention [Rocktäschel et al. ’15].

In this project, one or two neural network based models such as [Rocktäschel et al. ’15] and [Cheng et al. ’16] will be adapted and evaluated. Improvements on these models will be made based on observations from experiment results.

3 Evaluation

For evaluation, the plan is to develop feature based models as baseline to compare with the LSTMN method. There are 7 candidate features that we consider now: (1) BLEU score, (2) length difference as a real value, (3) overlap both as an absolute count and as a percentage, (4) unigram and bigram, (5) cross-unigrams, (6) cross-bigrams [Bowman et al.’15] and (7) cross-name-entities. We will also try different machine learning classifiers such as logistic regression and random forest to experiment on different subsets of the features.
At this time, Samsung is shifting its focus from CPT to LCD. It didn’t work with SEG Samsung, but moved the 4th generation LCD production line to Shenzhen to set up a new company.

【赛迪网讯】3月28日消息，据国外媒体报道，谷歌或正在致力开发的第三方评论平台应用即将推出。[By CCIDNET] According to foreign media reports on March 28, Google might unveil its dedicated third-party review platform soon.

Salesforce与和创科技(原图搜天下)达成一篮子合作。Salesforce and Hechuang Inc. (formerly Tusou World) reaches a whole package of cooperation.

今年7月，和创科技宣布和阿里旗下的钉钉达成合作。In July, Hechuang Inc. announced cooperation with Dingding, an affiliate company of Alibaba.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>而此时,三星自身也正在收缩彩管业务,转型平板,它把液晶四代线移到深圳,但没有和赛格三星合作,而是单独设立新公司。</td>
<td>三星后来在大陆建液晶模组生产线时选择单独设立公司,没有和赛格三星合作。</td>
<td>A_entails_B</td>
</tr>
<tr>
<td>At this time, Samsung is shifting its focus from CPT to LCD. It didn’t work with SEG Samsung, but moved the 4th generation LCD production line to Shenzhen to set up a new company.</td>
<td>Samsung later set up a separate company in the mainland for building LCD module production line, and didn’t cooperate with SEG Samsung.</td>
<td></td>
</tr>
<tr>
<td>【赛迪网讯】3月28日消息，据国外媒体报道，谷歌或正在致力开发的第三方评论平台应用即将推出。[By CCIDNET] According to foreign media reports on March 28, Google might unveil its dedicated third-party review platform soon.</td>
<td>谷歌或正在致力开发的第三方评论平台应用即将推出。Google might unveil its dedicated third-party review platform soon.</td>
<td>Similar</td>
</tr>
<tr>
<td>Salesforce与和创科技(原图搜天下)达成一篮子合作。Salesforce and Hechuang Inc. (formerly Tusou World) reaches a whole package of cooperation.</td>
<td>今年7月，和创科技宣布和阿里旗下的钉钉达成合作。In July, Hechuang Inc. announced cooperation with Dingding, an affiliate company of Alibaba.</td>
<td>Neutral</td>
</tr>
</tbody>
</table>

Table.1 Example sentence pairs in BosonNLI

References


