Aspect-Based Relational Sentiment Analysis Using a Stacked Neural Network Architecture

1 Introduction

Sentiment analysis needs to be regarded thus as a relation extraction problem consisting of three parts:

1.the extraction of aspect and opinion terms with respect to the discussed product, theme or event,

2. the labeling of these opinion terms with a sentiment (e.g. “positive”, “neutral”, “negative”), and

3. the extraction of relations between aspect and opinion

The extraction of aspect and opinion terms can essentially be regarded as a tagging task and can potentially be tackled by sequence modeling techniques such as Hidden Markov Models, Conditional Random Fields (CRFs) etc.

Firstly, we propose a component that combines convolutional neural networks with recurrent neural networks to extract aspect and opinion terms.

Secondly, a recurrent neural network extracts the expressed sentiment of each opinion term by using Part-of-Speech (POS) tags, word and distance embedding features.

Thirdly, we extract aspect-opinion relations by using a similar RNN model to classify extracted aspect and opinion terms in a pairwise fashion.

1.1 Related Work

Aspect and opinion term extraction for sentiment analysis has also been addressed using probabilistic graphical models.

2 Datasets

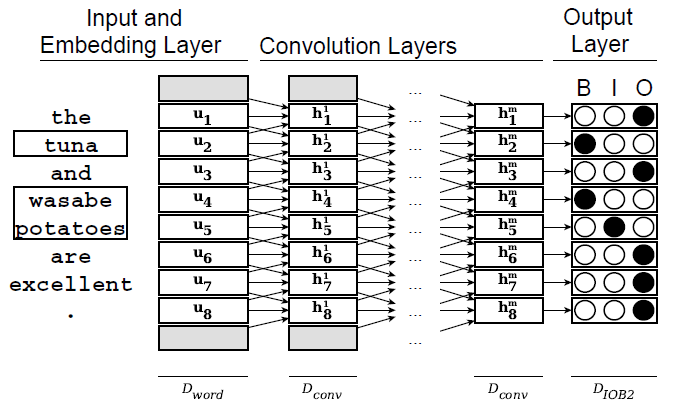
2.1 SemEval2015

2.2 USAGE

3 Aspect and Opinion Term Extraction

3.1 Features

3.2 Convolutional Neural Network Model



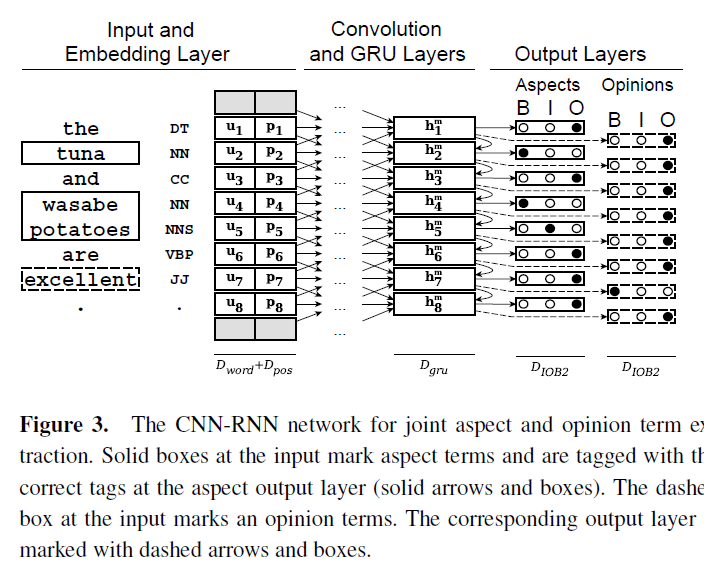
3.3 Recurrent Neural Network Model

Our RNN architecture comprises of an embedding layer for our Dword = 100 dimensional (pretrained) word embeddings, a GRU layer with Dgru = 100 hidden units, and a dense layer with a softmax activation applied to each single output vector of the GRU’s output sequence. The hidden layer sizes for this component are therefore 100-100.

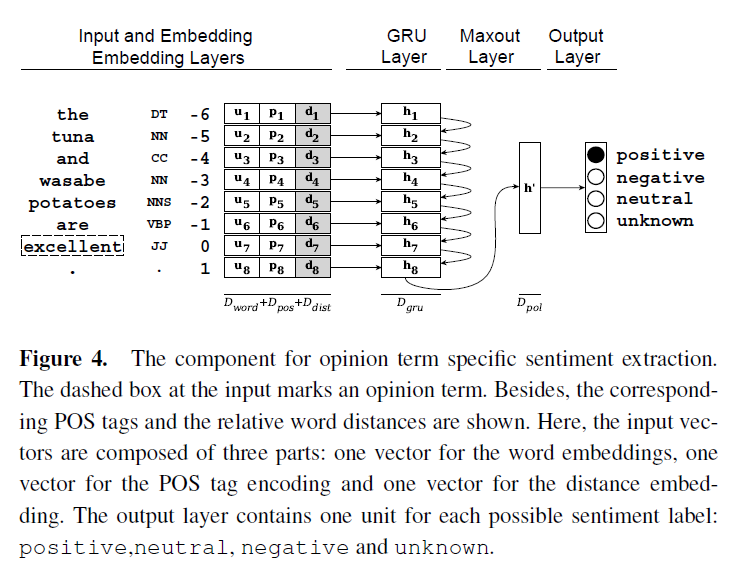
3.4 Stacked Model

We design the new combined model as follows. First, we apply convolutional layers to the input sequence, similar to the model in 3.2, yet only up to the final hidden layer. On top of this sequence of high-level features, we stack a GRU layer that learns temporal dependencies of its input sequence. Again, a dense layer with a softmax activation is used to map the recurrent hidden states to tag probabilities. The stacked CNN-RNN model uses the hidden layer sizes 100-50-50-50-100 (or 146-50-50-50-100 when using additional POS tag features).

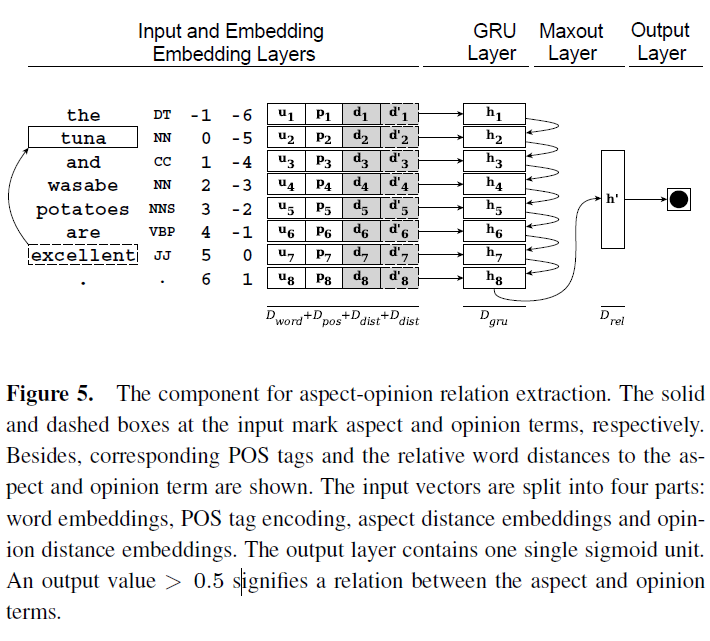
3.5 Joint Model



4 Opinion Term Specific Sentiment Extraction



5 Aspect-Opinion Relation Extraction



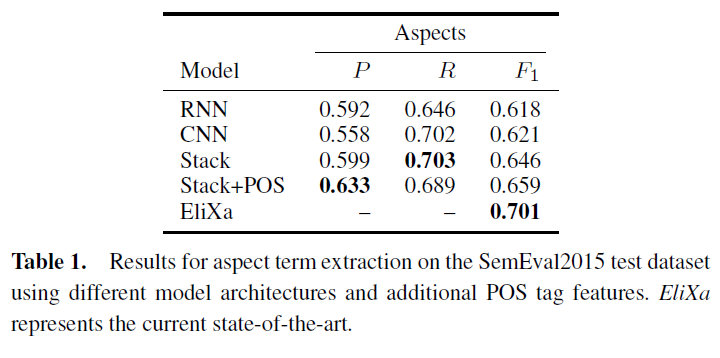
6 Experiments and Results

6.1 Training

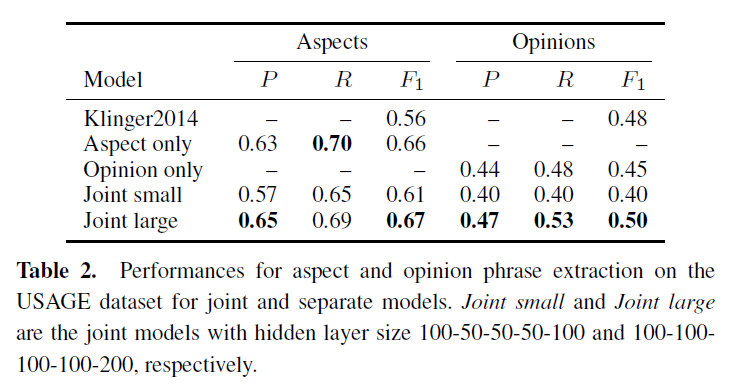
6.2 Initialization of Word Embeddings

6.3 Evaluation: Aspect and Opinion Term Extraction

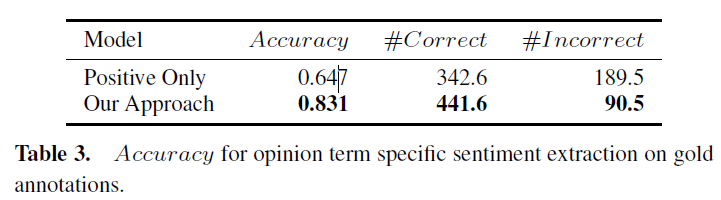
6.3.1 CNN vs. RNN vs. Stacked Models



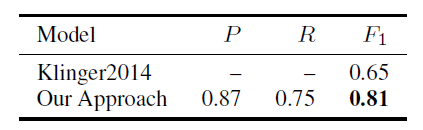
6.3.2 Joint vs. Separate Models



6.4 Evaluation: Opinion Term Specific Sentiment Extraction



6.5 Evaluation: Aspect-Opinion Relation Extraction



7 Conclusion and FutureWork