

Introduction

The Sentences Involving Compositional Knowledge (SICK) dataset consists of 9,840 pairs of sentences. Each sentence pair is labelled as either contradiction, neutral or entailment. The following experiment uses a deep, Siamese, bidirectional, Long Short-Term Memory (LSTM) network to predict sentence entailment using *Word2Vec* embeddings. The data set is split into 4,934 training pairs and 4,906 test pairs.

Word2Vec

Word2Vec is a technology that uses a two-layer neural network to produce a dictionary of vectors from a set of sentences [1]. The vector representation of a word is formed using its context and surrounding vocabulary, which enables the embedding process to carry a words semantic meaning.

It has been demonstrated that words with comparable meanings are embedded as similar vectors, which means they can be manipulated to find synonyms, or to search for similar words in the dictionary of vectors. An example of this being $vec(\text{king}) - vec(\text{man}) = vec(\text{queen})$.

Embedding the entirety of the SICK data set using *Word2Vec* yields a vocabulary of 2,217 individual words.

Network Architecture

The network reads-in two input sentences as sequences of 300-dimensional *Word2Vec* vectors. 300 dimensions were chosen, as a previous study had success using this embedding dimensionality [2].

Each sentence is fed into a bidirectional LSTM branch. The output from each branch is merged, and input into two dense layers. Softmax is then applied to one-hot encode the output. A diagram of the network architecture is shown in Fig 1.

Results

An accuracy of 76.03%, and a loss value of 0.5905 has been achieved when predicting entailment on the unmodified test pairs. Out of 4,906 test predictions, 1,176 were incorrectly classified.

Accuracy is limited by the number of training examples, which is demonstrated by an accuracy of 77.09%, and a loss value of 0.5586 when splitting the training/test data sets at 75:25. Out of 2,432 test predictions, 562 were incorrectly classified.

Further Work

Accuracy may be improved by generating more training examples. This has been done in another study by replacing random words with synonyms [2].

The SICK dataset also holds a relatedness score for each sentence pair. Further work could be done to test the network configuration to predict this score by changing the final layer to output a probability.

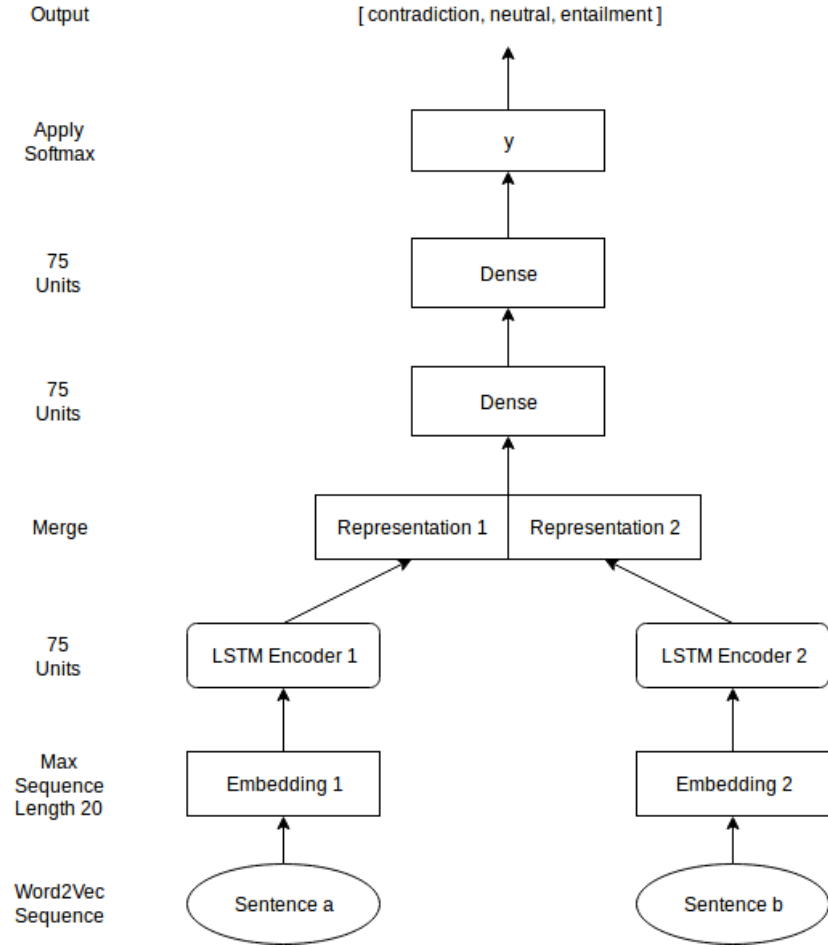


Figure 1: Architecture of the network. Two sentences are passed through identical LSTM branches before being processed through two dense layers and categorised using Softmax.

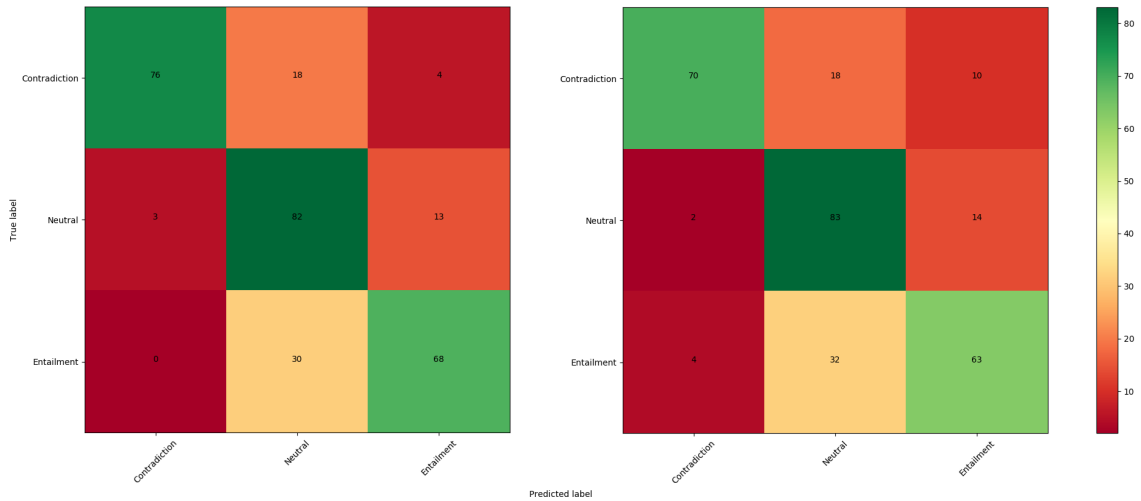


Figure 2: Normalised confusion matrices showing the number of true labels plotted against their predicted classes. The right-hand side matrix has been trained using a 50:50 train/test split. The left-hand matrix has been trained using a 75:25 train/test split.

References

- [1] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.
- [2] Jonas Mueller and Aditya Thyagarajan. Siamese recurrent architectures for learning sentence similarity. In *AAAI*, volume 16, pages 2786–2792, 2016.