Relations Between Text Chunks

Joe Kurian Eappen
IIT Madras

Shajith M., Nitin G.
IBM Research India

Abstract

Relations are generally between entities. In this work, we explore the idea of relations between arbitrary chunks of text. The relation here is a summary of one text chunk with respect to the other. This relation is extracted by two methods, one a simple rule-based model and the other, a neural model with abstractive summarization capabilities.

1 Introduction

There is a significant body of work on extracting relations between entities in the same sentences. The question we aim to answer is: Given two text chunks A and B, can we label a directed edge from chunk A to chunk B with information about how chunk B is related to chunk A? How would we structure this label?

This could help us in question-answering systems and knowledge graph creation.

Related Work

The closest work we could find involved argumentative relation mining (Nguyen and Litman 2016). The premise of this paper was figuring out the relation between chunks of text in an argument. At its core, this was formulated as a multiclass classification problem for the chunks ('Major-Claim','Claim' and 'Premise') and a binary classification task on the relations (*vis.* Support vs Attack). Since the number of ways to label the relation in our case were very large and not fixed, a similar classification approach would not fit our use case.

Recent work (Nema et al. 2017) introduced a bi-textual sequence-to-sequence model. We adapt this model to fit our problem.

2 A Rule-based Approach

Our baseline involves using common entities extracted from OpenIE relation triples to rank (entity-verb) pairs in the chunks.

1. Extract noun groups from the relation triples extracted by coreNLP using Part-of-Speech Tags and the Dependency Tree obtained.
2. Find most frequent common entities (noun groups) between the two text chunks. Call this C.
3. Create a relation frequency counter on (entity,action) pairs in each text chunk for the set $R = E \times A$ where E =entities in a relation, A=actions in a relation.
4. To find a relation from chunk A to chunk B, return the pairs in chunk B's R which have entities in C.
5. If no results are obtained loosen the requirement by allowing words close to the entities in the word embedding space to match. We can extend this to a WordNet based measure as well.

Chunk A: Measures seeking to bring down the incidence of frauds perpetrated through bank drafts should be built into the draft form itself. Necessary changes in system and procedures to speed up issue and payment of drafts should be taken. Banks should ensure that demand drafts of 20,000/ and above are issued invariably with account payee crossing.
Chunk B: Duplicate draft, in lieu of lost draft, up to and including 5,000/ may be issued to the purchaser on the basis of adequate indemnity and without insistence on seeking non payment advice from drawee office irrespective of the legal position obtaining in this regard. Banks should issue duplicate Demand Draft to the customer within a fortnight from the receipt of such request.
A to B: Issue of Duplicate Demand Drafts
B to A: Issue of Demand Drafts

Table 1: An Example

We can compare our method with a simple word frequency based method which naively pairs the most common noun groups and verbs. These are on a small manually created dataset of 16 entries from RBI (Reserve Bank of India) Compliance documents.

Score	Word Count (WC)	Our Method
ROUGE-1	14.05	19.70
ROUGE-2	0.8	2.82

Table 2: Baseline results on the small dataset

Ground Truth:
A to B: Issue of Duplicate Demand Drafts
B to A: Issue of Demand Drafts
Predictions:
A to B: demand draft duplicate
A to B (WC): purchaser duplicate include
B to A: demand draft ensure
B to A (WC): draft issue ensure

Table 3: Results on the domain data example

3 Deep Neural Model

To determine an appropriate Neural model, we must create a system that

- takes as input two chunks of text (bi-textual)
- predicts a text sequence based on both of them (the relation).

The sequence can be thought of as a summary of one chunk with respect to the other. Recent work (Nema et al. 2017) introduces Query-based Abstractive Summarization that has a similar architecture (*vis.* a bi-textual sequence-to-sequence model).

Since this architecture assumes relatively small queries, a direct usage will not work with our large text chunks. A simple method to subvert this is to perform an extractive summarization step using algorithms like LexRank (Erkan and Radev 2004).

Architecture

The chunks are encoded separately via a recurrent network taking in each word sequentially. The decoder used an attention mechanism that attended to chunk A based on a weighted vector of the chunk B encoder hidden states. This weighting is done via a learnable non-linear projection of the decoder state and chunk B encoder hidden states. All RNNs used were GRUs (2014) and had a hidden layer dimension of 200.

The Diversity Module introduced in (Nema et al. 2017) to reduce repetition in the decoder output was also used with the same hidden layer size.

Some Points to Note

1. Relation structure is not fixed and is learned from training data.
2. Too address large chunk sizes we can consider extracting a portion or improving our attention model.

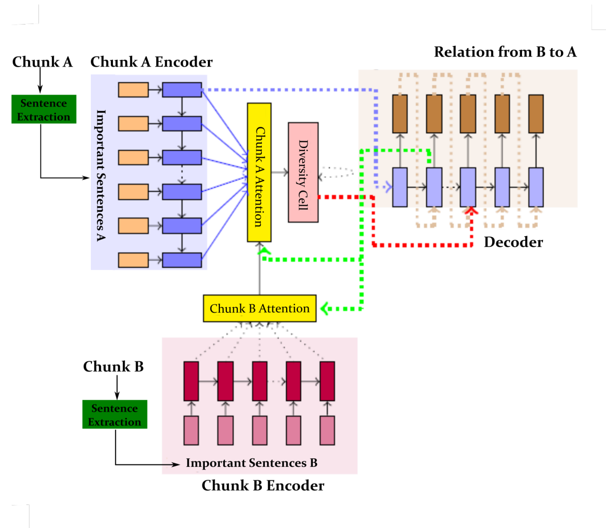

Figure 1: Deep Relation Generation Model

4 Dataset

To train our model, a sufficiently large dataset is required. We re-purposed the on-line encyclopedia, Wikipedia for this case by making two key observations:

- Every article on Wikipedia can be thought of as a collection of chunks under respective headings, all being about the same topic.
- We can try grouping chunks within an article as related and use the head chunk's title as the relation.

Thus we separated the article into chunks with the relations from chunk A to chunk B being the heading of chunk B. This ground truth relation is the concatenation of the headings in entire branch of the content hierarchy starting from the immediate heading of chunk B. To group the chunks together, a simple first-to-all order was followed wherein the first introductory portion of the article was chunk A and the other chunks were chunk B.

Dataset Metrics

To get an idea of the size of these chunks, we look at the following metrics.
ES nA: Extractive summarization to pull out n sentences from chunk A.

A2B relation average length $= 6.1$ words.

5 Results

The results are for different chunk sizes by varying the number of sentences extracted. We split the Dataset into 80% for training and 10% each for testing and validation.

From the scores in Table 6 we see that performance degrades with an increase in size. As expected, our Baseline would perform sub-optimally as our assumed structure does not hold in the general case.

Chunk A: the domestic cat is a small typically furry carnivorous mammal **Chunk B:** domestic cats especially young kittens are known for their love of play this behavior mimics hunting and is important in helping kittens learn to stalk capture and kill prey **A to B:** cat behavior play
Chunk A: australian rules football australian rules football officially known as australian football but also called aussie rules football or footy and in some regions marketed as afl after the australian football league is a contact sport played between two teams of eighteen players on an oval shaped field often a modified cricket ground possession of the ball is in dispute at all times except when a free kick or mark is paid **Chunk B:** both world war i and world war ii had a devastating effect on australian football and on australian sport in general in queensland the state league went into recess for the duration of the war vfl club university left the league and went into recess due to severe casualties **A to B:** australian rules football history effects of the two world wars

Table 4: Examples from the Wikipedia Dataset

Score	**ES 1A**	**ES 2A**	**ES 3A**	**Full Chunk**
Chunk A	29.7	55.4	78.7	259.1
Chunk B	74.6	74.6	74.6	241.1

Table 5: Chunk Lengths (in words)

Score	**ES 1A**	**ES 2A**	**ES 3A**	**Baseline**
ROUGE-1	52.71	53.78	39.94	10.72
ROUGE-2	33.06	33.68	19.82	00.64

Table 6: ROUGE scores of Deep Neural Model

6 Conclusion

We have demonstrated a method to help solve the problem of labeling the relation between two arbitrary text chunks.

Further experiments such as testing with two unrelated chunks will help check the capabilities of this approach.

7 Future Work

The extractive summarization algorithm step used prevents the model from being end-to-end trainable. One can consider adding a hierarchical attention (Yang et al. 2016) over the sentences and words to supplant this procedure. While initial results seem to be difficult to obtain due to memory constraints and large chunk size, our work is ongoing.

Dynamic Co-attention Networks (Xiong, Zhong, and Socher 2017) introduce a co-attention principle that could potentially improve performance here as well. We expect improvements by looking at using Chunk A context vector in the chunk B encoder and methods of co-attention over both chunks.

Ground truth: australian rules football history effects of the two world wars **ES 2A:** australian football history effects of the two world wars
Ground truth: invasion of italy and death death **ES 2A:** irish of italy and death death
Ground truth: computer timeline of analog computers precursors **ES 2A:** computer model of analog computing

Table 7: Example results on the Wikipedia Dataset

References

Chung, J.; Gülçehre, Ç.; Cho, K.; and Bengio, Y. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *CoRR* abs/1412.3555.

Erkan, G., and Radev, D. R. 2004. Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of Artificial Intelligence Research* 22:457–479.

Nema, P.; Khapra, M. M.; Laha, A.; and Ravindran, B. 2017. Diversity driven attention model for query-based abstractive summarization. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*.

Nguyen, H. V., and Litman, D. J. 2016. Context-aware argumentative relation mining. In *Proceedings of the Joint Conference of the 54th Annual Meeting of the ACL*. Association for Computational Linguistics.

Xiong, C.; Zhong, V.; and Socher, R. 2017. Dynamic coattention networks for question answering. In *Proceedings of the International Conference on Learning Representations (ICLR)*.

Yang, Z.; Yang, D.; Dyer, C.; He, X.; Smola, A.; and Hovy, E. 2016. Hierarchical attention networks for document classification. In *Proceedings of NAACL-HLT*, 1480–1489.