Adapting and Improving Nature Language Processing Methods for Conversations

Imperial College London

Contributors: Jonas Scholz, Shona Curtis-Walcott
Supervisor: Dr. Tim Evans
Research Group: Theoretical Physics

Introduction

1

In many areas of life decisions are made, problems are tackled, and ideas are developed through discussions (Fig. 1). It is therefore worthwhile to study how discussion performs as a means to reach conclusions: **How do ideas evolve over time? What linguistic features do certain speakers use to influence the flow of a discussion?** By statistically analysing conversation transcripts and extracting features of conversational structure, we can detect information that is otherwise obscured or unavailable.

This project investigates state-of-the-art data processing methods for conversational text analysis and display.

We base our research **on conversational podcast shows** due to the wide range of topics, opinions, personalities and conversation styles contained within the transcripts.

Figure 1: Illustration of the exchange of ideas and opinions that occurs during conversation. Image source: https://www.isystain.com/conversations.

While initially the aim was to apply existing Natural Language Processing (NLP) techniques specifically to podcasts, our evaluation of said methods within the context of podcast transcripts found that the methods were insufficient. Therefore, we instead **shifted our focus to improve existing techniques, and to adapt them specifically for the purpose of conversation analysis. We achieved results that significantly outperform the state of the art.**

Aims

2

- Evaluate existing NLP methods when applied to conversations, identify limitations and improve upon those techniques. Specifically, address the fields of **topic modelling** and **dialogue act classification**.
- Investigate how to best visualise the evolution of ideas and flow of conversation from podcast transcripts. Specifically, develop `Discussion Trees': graphical representations of transcripts which give a viewer insight into the structure of a given podcast, and the extent to which different topics were discussed.
- Apply our improved techniques to a large collection of podcast transcripts and publish a corpus of annotated transcripts for future work.

Dialogue Act Classification

4

What is the Goal?

Dialogue acts (DAs) describe the **social function** of any given sentence[5]. For example, the sentence “Hi how are you?” is a **greeting**. Other examples include statements, questions, answers, acknowledgements, and many others. **The goal is for a computer to map a set of utterances (sentences) to their appropriate dialogue acts.**

Existing Approaches

All state-of-the-art approaches use recurrent neural networks[6-12], which are algorithms capable of processing sequences of information, such as sequences of words and sequences of sentences at once. This is beneficial as the context of phrases (such as words, sentences) determines their meaning. For example, consider the following sections of conversation:

And you know what he did after that? *Yeah?*

Do you have a pet? *Yeah*

In the left example, “Yeah” is an acknowledgement, prompting the first person to continue. In the right, “Yeah” is an affirmative answer. **The context matters, and recurrent neural networks can capture it**.

We base our work on the specific model shown in Fig. 3. It achieves an accuracy of 79.2%[6]. The state of the art achieves 82.3% accuracy[12]. Both accuracies are found when evaluating the models on the publicly available Switchboard Dialogue Act Corpus[13].

Our Improvements

- We **fix a bug** in the original implementation which caused words adjacent to punctuation to be discarded.
- We **update the original model’s outdated recurrent neural network component**, LSTM cells[14], in favour of the more modern GRU[15] cells, which reduces the complexity of the model, improves accuracy and allows for faster training.
- The original model uses Stanford’s GloVe embeddings[16] to turn words into meaningful numbers. We instead use the **ConceptNet Numberbatch embeddings**[17], which outperform GloVe embeddings significantly[18].
- Overall, **we improve the accuracy of the model from 79.2% to 84.6 +- 0.3%**, which is currently the **state-of-the-art** and is higher than the inter-accuracy between the humans creating the dataset[13].

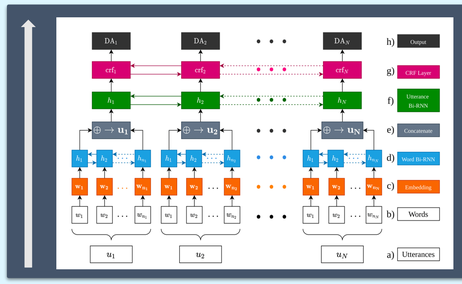

Figure 3: Model architecture

a) Utterances (i.e. sentences) are fed into the model as inputs
b) Utterances are split into constituent words
c) Words are embedded into a vector of numbers that represents the meaning of the word. We use the ConceptNet Numberbatch embeddings for this.
d) For each utterance, words are combined into a set of numbers that represents the whole utterance.
e) These numbers are concatenated into a single vector that can be understood as the embedding of the utterance.
f) The utterance embedding is fed into another RNN layer that takes the context of previous utterances into account.
g) A conditional random field layer makes the final classification and assigns each utterance a dialogue act.
h) The sequence of dialogue acts is the output.

Applications of Project

5

Visualising Conversation Structure

Trajectories through Topic Space

The visualisations in Fig. 4 use ConceptNet word embeddings to structure the layout of keywords in a topic space through which the discussion can be seen to navigate, providing immediate access to the key themes and topics discussed.

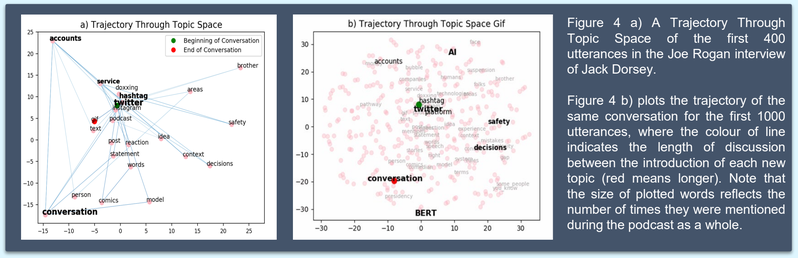

Figure 4 a) A Trajectory Through Topic Space of the first 400 utterances in the Joe Rogan interview of Jack Dorsey.

Figure 4 b) plots the trajectory of the same conversation for the first 1000 utterances, where the colour of line indicates the length of discussion between the introduction of each new topic (red means longer). Note that the size of plotted words reflects the number of times they were mentioned during the podcast as a whole.

Discussion Trees

A node is plotted for every utterance in a transcript, with its position decided based on the topics it contains: vertically-stacked nodes indicate consecutive utterances on the same topic, and horizontal steps between nodes indicate a change of conversation topic. When participants discuss a previously-mentioned topic, a new branch begins from where that topic was first mentioned. A Discussion Tree with many long branches indicates that the speakers covered many new topics in a 'linear' conversation (Fig. 5c); a Discussion Tree with many short branches indicates that the conversation loops back to the same topics (Fig. 5b).

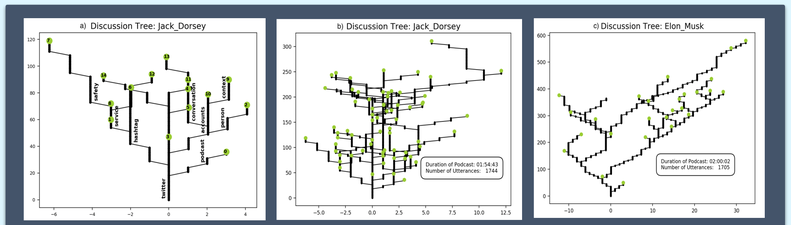

Figure 5: a) The Discussion Tree built from the first 400-utterances of the Joe Rogan interview of Jack Dorsey, with topics of discussion annotated. The full-transcript Discussion Tree of the same interview is given in b), and the full-transcript Discussion Tree of the Joe Rogan and Elon Musk is given in c), where the difference in tree structures reflects the nature of the conversations.

Further Research Ideas

One question we could answer is: **What makes a Conversation Interesting?** If we assert that a podcast being interesting is a necessary condition for popularity, we could use the number of listens and likes of a podcast episode as the popularity metric, and hence perform our linguistic and topical analysis on the most and least popular podcast episodes in our dataset to determine what led one podcast to be more successful than another.

Furthermore, if we were to employ large-scale **image analysis of Discussion Trees** we might extract a set of common patterns or templates in the trees. This could provide the basis of an investigation into **common conversation structures**, and by using our analysis of Dialogue Act usage we could then decipher why such structures arise, for example: what influences caused a meeting to consist of long, intricate discussions rather than quick, concise exchanges?

Topic Extraction

What is the Goal?

Given a conversation transcript, we would like to find out **which topics are raised and during which intervals of the conversation** they are active. A computer-generated example of the topical analysis of a conversation between Twitter CEO Jack Dorsey and podcast host Joe Rogan is shown in Fig. 2 below.

Existing Extraction Methods

For general text documents, the standard topic extraction method is latent Dirichlet allocation (LDA)[1]. LDA assumes that every document is generated by repeatedly randomly sampling a set of topics and then sampling words within these topics. By determining the most likely distributions fitting this assumption for a given document, the set of topics is extracted. Topic boundaries can be found as the gaps between cohesive distributions[2].

Limitations of Previous Methods

- In conversations a **lot of words are *not* part of a topic**: statements of politeness, jokes and acknowledgements diminish the effectiveness of LDA. This means LDA’s **generative assumption fails**.
- LDA is **unable to model temporal evolution of topics**. It assumes that topic distributions are global across the whole document. For other media such as newspaper articles, this may be a good approximation – for conversations it is not.

Our approach addresses these limitations by adapting the common **key phrase extraction** algorithm *TopicRank*[3] and using a modified version to only extract relevant words.

Graph of Embedded Extracted Keywords (GEEK)

3

- Our approach is a **graph-based model** that first extracts keywords such as nouns, proper nouns and named entities using pretrained neural networks.
- The range over which each of these keywords is active is found by **checking if it or semantically similar keywords are repeated**. The similarity of keywords is found as the cosine similarity of word embeddings.
- Certain keywords that fit this pattern but don’t describe topics, such as “dude” and “everything” and semantically similar keywords are removed through a manual filter.
- A graph is then created in which nodes, representing keywords, are connected if their keyword ranges overlap and if they are semantically similar. **Connected components** then **represent topics**.
- **The state-of-the-art conversation model** BayesSeg[2] achieved a windowDiff[4] score of 0.39+/-0.06. We beat this by a significant margin, **achieving 0.22+/-0.04** (lower is better).

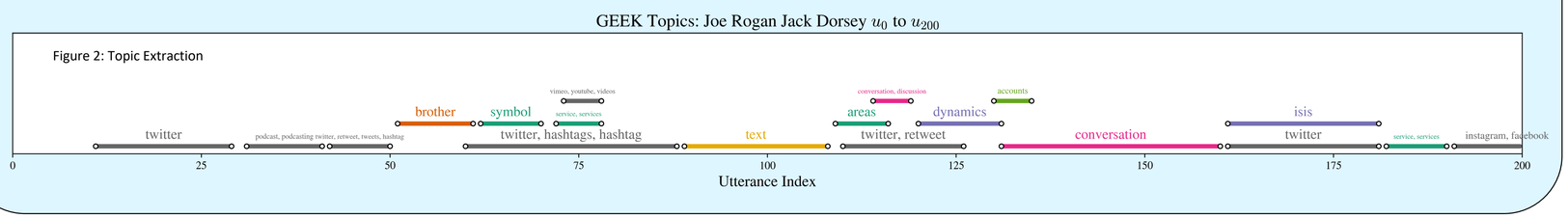

Figure 2: Topic Extraction

[1] David M Blei, Andrew Y Ng, and Michael I Jordan. “Latent dirichlet allocation”. In: Journal of machine Learning research 3.Jan (2003), pp. 993–1022
[2] Jacob Eisenstein and Regina Barzilay. “Bayesian unsupervised topic segmentation”. In: Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing. 2008, pp. 334–343.
[3] Adrien Bougouin, Florian Boudin, and B ́eatrice Daille. “TopicRank: Graph Based Topic Ranking for Keyphrase Extraction”. In: Proceedings of the Sixth International Joint Conference on Natural Language Processing
[4] Lev Pevzner and Marti A Hearst. “A critique and improvement of an evaluation metric for text segmentation”. In: Computational Linguistics 28.1 (2002), pp. 19– 36.
[5] McTear, Michael; Callejas, Zoraida; Griol, David (2016). The Conversational Interface: Talking to Smart Devices. Springer. pp. 162–166.
[6] Harshit Kumar et al. “Dialogue act sequence labeling using hierarchical encoder with crf”. In: arXiv preprint arXiv:1709.04250 (2017).
Conference on Empirical Methods in Natural Language Processing. 2018, pp. 887–893. 48 BIBLIOGRAPHY BIBLIOGRAPHY
[7] Vipul Raheja and Joel Tetreault. “Dialogue act classification with context-aware self-attention”. In: arXiv preprint arXiv:1904.02594 (2019).
[8] Ruizhe Li et al. “A dual-attention hierarchical recurrent neural network for dialogue act classification”. In: arXiv preprint arXiv:1810.09154 (2018).
[9] Yao Wan et al. “Improved dynamic memory network for dialogue act classification with adversarial training”. In: 2018 IEEE International Conference on Big Data (Big Data). IEEE. 2018, pp. 841–850.
[10] Zhejian Chen et al. “Dialogue act recognition via crf-attentive structured network”. In: The 41st international acm sigir conference on research & development in information retrieval. 2018, pp. 225–234.
[11] Chandrakant Bothe et al. “A context-based approach for dialogue act recognition using simple recurrent neural networks”. In: arXiv preprint arXiv:1805.06280 (2018).
[12] Sujith Ravi and Zornitsa Kozareva. “Self-governing neural networks for on-device short text classification”. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018, pp. 887–893.
[13] E. Holliman, J. Godfrey, and J. McDaniel. “SWITCHBOARD: telephone speech corpus for research and development”. In: Acoustics, Speech, and Signal Processing, IEEE International Conference on. Vol. 1. Los Alamitos, CA, USA: IEEE Computer Society, Mar. 1992, pp. 517–520.
[14] Sepp Hochreiter and J ̈urgen Schmidhuber. “Long short-term memory”. In: Neural computation 9.8 (1997), pp. 1735–1780.
[15] Junyoung Chung et al. “Empirical evaluation of gated recurrent neural networks on sequence modeling”. In: arXiv preprint arXiv:1412.3555 (2014).
[16] Jeffrey Pennington, Richard Socher, and Christopher D Manning. “Glove: Global vectors for word representation”. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014, pp. 1532–1543.
[17] Hugo Liu and Push Singh. “ConceptNet—a practical commonsense reasoning tool-kit”. In: BT technology journal 22.4 (2004), pp. 211–226.
[18] Robyn Speer and Joanna Lowry-Duda. “Conceptnet at semeval-2017 task 2: Extending word embeddings with multilingual relational knowledge”. In: arXiv preprint arXiv:1704.03560 (2017).