Decision Conversations Decoded

Léa A. Deleris, Debasis Ganguly, Killian Levacher, Martin Stephenson, Francesca Bonin

IBM Research - Ireland
Dublin, Ireland
`lea.deleris, debasis.ganguly1, killian.levacher,`
`martin_stephenson, fbonin@ie.ibm.com`

Abstract

We describe the vision and current version of a Natural Language Processing system aimed at group decision making facilitation. Borrowing from the scientific field of Decision Analysis, its essential role is to identify alternatives and criteria associated with a given decision, to keep track of who proposed them and of the expressed sentiment towards them. Based on this information, the system can help identify agreement and dissent or recommend an alternative. Overall, it seeks to help a group reach a decision in a natural yet auditable fashion.

1 Our Vision

Decision Analysis is the scientific discipline that formally studies decisions: procedures, methods, and tools for identifying, representing, and assessing important aspects of a decision, ultimately to recommend actions to the decision maker (Matheson and Howard, 1977). One of the focus of decision analysis is on practical aspects of formulating the decision problem (rather than focusing solely on its mathematical resolution). This includes (i) defining the utility function of the decision maker including criteria, risk attitudes and trade-offs, (ii) identifying the relevant uncertainties and (iii) investigating the benefits of gathering additional information. In order to achieve this, the decision analysis process needs certain inputs, specifically **alternatives** (options available to the decision maker), **criteria** (values, risk preferences, and time preferences of the decision maker), **frame** including the constraints associated with the decision (Howard and Abbas, 2015).

Many decisions are taken collaboratively, whether truly collaboratively (where everyone has a voice) or rather when a decision maker consults with a group of trusted advisers. For instance, for complex cases in medicine, it is common to have multiple experts meet to discuss the patient's situation and come up with a recommended course of action. When recruiting, different perspectives are typically taken into account to help inform a final manager of her decision to make an offer to a candidate. In large projects, multiple stakeholders can take part of important architectural decisions. However, collaborative decision discussions are typically unstructured, inefficient and can be frustrating for all participants, as illustrated by the Abilene Paradox [1].

With the proliferation of recording devices in our professional and personal lives (e.g., teleconferencing, intelligent personal assistant or group chat exchanges such as Slack), it would be helpful to develop NLP-based engines to automatically extract decisions related concepts such as alternatives and criteria from decision conversations and make use of that information to facilitate the decision discussions. As a starting point, such a technology could provide the input to generate a visualisation of the decision discussion so that a group can consult it to identify underdeveloped ideas or options, and to recall points of consensus and dissent. It would serve as a summary, enabling people who have missed a decision discussion to catch up or more simply reminding a decision maker of the arguments that were raised so she can make her decision at a later time.

The system output can also be used to document the decision making process in a structured way. This information in turn is key to better understanding power plays and negotiation in group decision making. More practically, it can be essential to prove compliance with processes, e.g., a financial advisor proving she has presented reasonable investment alternatives to her customers.

Note that our objective is to follow how a de-

---

[1] `https://en.wikipedia.org/wiki/Abilene_paradox` retrieved on February 15th 2018

cision is made, rather than focusing solely on its outcome i.e., the final choice (though this is a by-product).

2 Related works

Decisions are often presented as one of the most important outcomes of business meetings (Whittaker et al., 2006). Banerjee et al. (2005) show that updates about the decisions of a meeting are beneficial for persons who had missed the meeting to prepare for the next one. Interest on meeting developments is shown also by the large amount of corpus collections on the topic, e.g., ICSI (Janin et al., 2004) , AMI (Carletta et al., 2005), CHIL (Mostefa et al., 2007) or VACE (Chen et al., 2006). While some annotations in these corpora consider decisions from meetings, the annotated labels (text spans) are either too specific (dialogue acts) or too general (meeting summaries) to study the decision making process.

Some studies have investigated automatic detection of decisions. Hsueh and Moore (2007) attempted to identify patterns of the decision gists, relying on the relevant annotated Dialogue Acts (DAs) in meeting transcripts. Fernández et al. (2008) extended the annotations with new decision-related DAs, and formulated the problem as a classification problem for each class of DAs. They designed an annotation scheme that takes into account the different roles that DAs play in the decision-making process, for instance to initiate a discussion by raising a topic, to propose a resolution, or to express agreement. However, in all this work, the objective was to detect the span of the conversation where the decision is taken. We intend to go further and identify the elements that belong to the *content* of decision-making processes, whether or not a final decision is taken. Cadilhac et al. (2012), while focusing more specifically on the representation of preferences, have proposed an approach to extract what we refer to as alternatives and which in their framework is described as outcomes. They do not pursue the extraction of criteria.

3 System Architecture

3.1 Overall

The various components, that together enable to decode decision conversations, are presented in Figure 1. In this diagram, we present both components that are currently implemented along with others that are in development (italics).

Input Processing Module. Input to the system is in the form of text. This text can originally come from a recording or live dialog, which is converted to text using Speech-To-Text technology. Speaker attribution is also performed as part of this step. Alternatively, input can come from text entered via the UI or from a set of pre-existing transcripts. The text is then pre-processed so as to provide a clean transcript with speakers identified to the Extraction and Summarization module.

Resources. The main part of the resources consists in a set of Machine Learning (ML) algorithms, which are described in Section 3.2. Annotated data used for training models can be enriched via user feedback of already identified criteria and alternatives, i.e the user can verify or refute an identified criteria or alternative. This annotated data can then be used to re-train the models. External resources, such as DBpedia and WordNet are also leveraged in the pipeline.

Extraction and Summarization Module. This module constitutes the core of the NLP pipeline and is composed of multiple sub-components. A) **Decision Segmentation** - As more than one decision may be discussed in a conversation, this component segments the conversation into the corresponding multiple discussion threads. B) **Decision Topic Analysis** - This component determines the topic of the decision. C) **Decision Element Extraction** - Using ML models, this component identifies the location of the decision alternatives and criteria in the text. D) **Semantic Grouping** - This component clusters semantically similar alternatives and criteria. E) **Mapping Alternative to criteria** - This component associates identified alternatives to criteria. F) **Wikification** - This component further enriches the transcript by linking words and phrases to external resources (e.g., Wikipedia). G) **Identification of Expressed Sentiment** - This component determines the expressed sentiment of speakers towards extracted alternatives and criteria.

Finally, the output of this text processing is recorded in a JSON data structure called Structured Decision Summary Output.

Summary Analysis Module. This module analyses the Structured Decision Summary Output based on the following two components. First the recommendation module can make use of the information to identify which alternative seems the
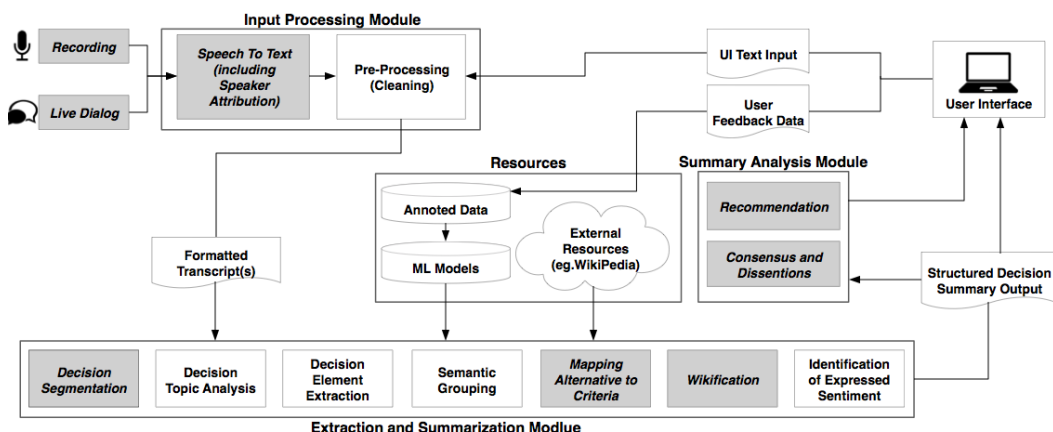
Figure 1: Architecture of our System

most supported by the group. Similarly, it can search for dominated alternatives and suggest they be discarded. Second the consensus and dissent module analyzes where and when in the discussion people agree and disagree on the proposed alternatives.

User Interface. Its main functions are to allow the user to input text directly for analysis and to subsequently present him/her with the alternative and criteria extraction output, in addition to the option to cluster and/or summarise this output as described in Section 4. Finally, it enables the user to accept or refute identified alternatives and/or criteria identified by the system.

3.2 Machine Learning Module

Corpus. We leverage the AMI (Augmented Multi-party Interaction) (Carletta et al., 2005) corpus, which we annotated with alternative and criteria. Description of our annotation process along with access to the corpus is summarised in (Deleris et al., 2018). We use supervised classification settings, where 80% sentences from the AMI corpus are used for training the models and the rest for testing. **Sequence Prediction**. Our automatic identification of alternative and criteria is based on standard sequence prediction approaches. We experimented with many common models namely naive Bayes, MaxEnt, SVM, CRF and LSTM based RNN. As expected the bag-of-words based models (with fixed length context features), i.e. naive Bayes, MaxEnt and SVM were outperformed by the sequence models, namely CRF and RNN. The linear CRF is currently our model with the highest performance as outlined in Table 1 (due to space constraints the results of other models are not shown).

Label	Precision	Recall	F-score
Alternative	0.6311	0.4667	0.5366
Criteria	0.7368	0.3394	0.4647

Table 1: Performance of the CRF model (based on token level evaluation on the AMI corpus).

4 Demonstration Flow

4.1 Describing Interface

Our demo interface starts from a text box where a user can enter the transcript to be analyzed as presented on Fig. 2. Clicking on the button *Analyze Text* located underneath the text box will run the topic analysis and extraction algorithms whose results will then be shown to the user underneath the text box.
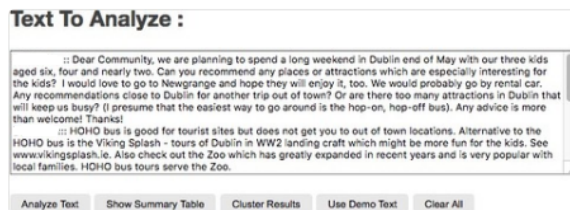

Figure 2: Initial Screen into the System

Specifically the topic analysis provides background information about the main themes of the decision discussion and more importantly describes the frame for the decision that is being discussed, specifically the decision topic e.g. *Can you recommend any places or attractions which are especially interesting for the kids?*, and the context of this decision, e.g., *Dear Community, we are planning to spend a long weekend in Dublin end of May with our three kids*. The results of

93

the extraction algorithm results are then displayed. The input text is presented with sections highlighted to indicate detected alternatives and detected criteria, as shown on Fig. 3.
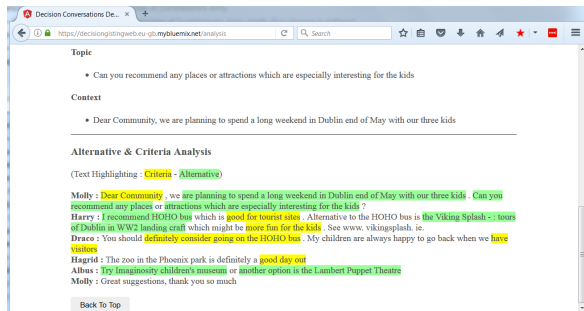

Figure 3: Alternatives and Criteria Extraction

While this representation of the output of the NLP algorithms is instructive to understand how the system operates, we feel a more useful summary to effectively guide decision discussions should be based on grouping alternatives and criteria by person as in Fig 4 and also grouping alternatives and criteria by semantic topic. Those two subsequent analyses are obtained by clicking on *Show Summary Table* and *Cluster Results* shown in Fig. 2. Note that the summary table format also allows to indicate the expressed sentiment of the person towards the alternative or criteria (as represented by the smiley faces). Finally when a user hovers over a detected fragment, we show in an overlay window the part of the transcripts from where it was extracted, so as to provide context for its interpretation if needed.

Figure 4: Summarization in a Table with Sentiment and Context Overlay

4.2 Examples of Analyses

In this section, we provide some illustrative results of the use of our technologies on diverse kinds of discussions. Note that we have slightly edited the text, mainly changing the names of the speakers

and cutting some long utterances.

Figure 5 top shows an excerpt from the AMI Corpus (Carletta et al., 2005) which corresponds to face-to-face discussions about remote control design (specifically ES 2012). Figure 5 middle relates to a discussion about a visit to Machu Pichu on a travel website where a user has requested advice from other users. Finally, the text in Figure 5 bottom is extracted from the discussions of the European Parliament, using the Europarl Corpus (Koehn, 2005).

Figure 5: Examples (top: AMI Corpus - ES 2012; middle: Travel Forum Discussion; bottom: European Parliament - September 18th 1996)

5 Conclusion

We make countless decisions every day, some of which are bound to be collaborative, making the decision process all the more challenging. Our system proposes to automatically follow the decision process. It tracks the options being considered, why they are proposed (i.e., which criteria are brought up), by whom and with whose support. It then organizes all collective thoughts into a summary in order to facilitate further discussions, guide the final decision, explain how a decision was made or make recommendations.

As a virtual facilitator, the system objective is to augment collaborative decision making, empowering all stakeholders involved to contribute their perspective and making the decision making process effective and transparent.

References

Satanjeev Banerjee, Carolyn Penstein Rosé, and Alexander I. Rudnicky. 2005. The necessity of a meeting recording and playback system, and the benefit of topic-level annotations to meeting browsing. In *INTERACT*.

Anaïs Cadilhac, Nicholas Asher, Farah Benamara, Vladimir Popescu, and Mohamadou Seck. 2012. Preference extraction from negotiation dialogues. In *Proceedings of the 20th European Conference on Artificial Intelligence*, pages 211–216. IOS Press.

Jean Carletta, Simone Ashby, Sebastien Bourban, Mike Flynn, Mael Guillemot, Thomas Hain, Jaroslav Kadlec, Vasilis Karaiskos, Wessel Kraaij, Melissa Kronenthal, et al. 2005. The ami meeting corpus: A pre-announcement. In *International Workshop on Machine Learning for Multimodal Interaction*, pages 28–39. Springer.

Lei Chen, R. Travis Rose, Ying Qiao, Irene Kimbara, Fey Parrill, Haleema Welji, Tony Xu Han, Jilin Tu, Zhongqiang Huang, Mary Harper, Francis Quek, Yingen Xiong, David McNeill, Ronald Tuttle, and Thomas Huang. 2006. Vace multimodal meeting corpus. In *Proceedings of the Second International Conference on Machine Learning for Multimodal Interaction*, MLMI'05, pages 40–51, Berlin, Heidelberg. Springer-Verlag.

Lea A Deleris, Tuan Tran, Francesca Bonin, Debasis Ganguly, and Killian Levacher. 2018. Preparing a dataset for extracting decision elements from a meeting transcript corpus. Technical report, IBM Research.

Raquel Fernández, Matthew Frampton, Patrick Ehlen, Matthew Purver, and Stanley Peters. 2008. Modelling and detecting decisions in multi-party dialogue. In *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*, pages 156–163. ACL.

Ronald A Howard and Ali E Abbas. 2015. *Foundations of decision analysis*. Pearson.

Pei-Yun Hsueh and Johanna D Moore. 2007. Automatic decision detection in meeting speech. In *International Workshop on Machine Learning for Multimodal Interaction*, pages 168–179. Springer.

Adam Janin, Jeremy Ang, Sonali Bhagat, Rajdip Dhillon, Jane Edwards, Javier Macas-guarasa, Nelson Morgan, Barbara Peskin, Elizabeth Shriberg, Andreas Stolcke, Chuck Wooters, and Britta Wrede. 2004. The icsi meeting project: Resources and research. In *in Proc. of ICASSP 2004 Meeting Recognition Workshop*. Prentice Hall.

Philipp Koehn. 2005. Europarl: A parallel corpus for statistical machine translation. In *MT summit*, volume 5, pages 79–86.

James E Matheson and Ronald A Howard. 1977. An introduction to decision analysis. In Miller Howards, Matheson, editor, *Readings in Decision Analysis*, chapter 1, pages 9–43. Stanford Research Institute, California.

D. Mostefa, N. Moreau, K. Choukri, G. Potamianos, S. Chu, A. Tyagi, J. Casas, J. Turmo, L. Cristoforetti, F. Tobia, A. Pnevmatikakis, V. Mylonakis, F. Talantzis, S. Burger, R. Stiefelhagen, K. Bernardin, and C. Rochet. 2007. The chil audiovisual corpus for lecture and meeting analysis inside smart rooms. *Language resources and evaluation*, 41(3):389–407.

Steve Whittaker, Rachel Laban, and Simon Tucker. 2006. *Analysing Meeting Records: An Ethnographic Study and Technological Implications*. Springer Berlin Heidelberg, Berlin, Heidelberg.