

CBR For Interpretable Response Selection In Conversational Modelling

Malavika Suresh^{1,*†}

¹Robert Gordon University, Aberdeen, United Kingdom

Abstract

Current state-of-the-art dialogue systems are increasingly complex. When used in applications such as motivational interviewing, the lack of interpretability is a concern. CBR offers to bridge this gap by using the most similar past cases to decide the outcome for a new problem, which then serves as a natural as well as accurate explanation of the outcome. This research proposes to extend the Abstract Argumentation CBR (AA-CBR) framework for predicting the next response type in an ongoing conversation by reusing the knowledge of previous conversations to achieve a desirable outcome for a new conversation context.

Keywords

Case Based Reasoning, Conversational Modelling, Motivational Interviewing, Abstract Argumentation

1. Introduction

There is recent research interest in automating motivational interviewing² (MI) conversations due to the effectiveness of MI and lack of trained MI interviewers [1]. While some work has shown that large-scale pre-trained language models can be useful to train MI interviewers by predicting responses [2], the model predictions are not explained. In such case, the decision to accept or reject a model's proposed response falls on the trainee. For example, without an explanation of possible outcomes, it is possible that the trainee may reject a model's suitable proposal for a less suited response that they prefer. This is a concern as it is already a difficulty for human MI practitioners to suppress the instinct to respond with premature advice [3]. Equally possible is that an unsuitable model prediction may be accepted by the trainee, which can be avoided when an explanation is available. Additionally, the type of response (i.e dialogue strategy) needs to be adapted based on the individual as the same strategy may result in different outcomes with different individuals. Thus, the interpretable CBR approach of using past cases to decide the outcome for a new case could be better suited for this problem.

CBR approaches to dialogue management have been studied in prior work [4] [5], where each utterance in a dialogue is considered as a case. In contrast, in this work we consider the whole dialogue as a case and propose the use of Abstract Argumentation for CBR (AA-CBR) for selecting next response type. As the framework represents past cases as a tree of arguments attacking and defending each other (i.e an argument graph, as in Fig 1), it can provide natural interactive explanations of the predicted outcomes [6].

ICCBR DC'22: Doctoral Consortium at ICCBR-2022, September, 2022, Nancy, France

✉ m.suresh@rgu.ac.uk (M. Suresh)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

²MI is a form of therapy that encourages people to realize on their own the need for change in attitude/belief

The expected research contributions are - (i) case representation of an MI conversation, (ii) adaptation and extension of AA-CBR for response type selection in MI conversations. These contributions can be extended to any task-oriented conversational model that requires interpretable response generation based on personalized context.

1.1. Background

This section illustrates an example to briefly summarize the AA-CBR framework originally proposed by [7]. AA-CBR is based on the argumentation framework [8] which is basically a set of arguments and a binary *attack* relationship between them which defines whether ArgumentA attacks ArgumentB. AA-CBR represents each case as a set of factors. When adding a new

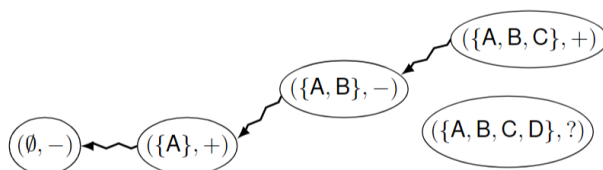


Figure 1: Example of an argument graph using the AA-CBR framework, taken from [7]: Alphabetical characters represent factors; Each node is a case in the case base and may consist of one or more factors and either a positive or negative outcome; The null node represents the default outcome in the absence of any factors; Arrow represents an attack relationship between two cases;

factor to the case changes the case outcome, it is considered an *attack*. Argumentation rules define the attack relationship between cases to then represent the case base as an argument graph. For instance in Fig 1, {A,B,C} which is a more specific case (i.e more factors) attacks {A,B} which is a less specific case with a different outcome. By inference, {A,B,C} defends {A} and attacks the default case. Note that {A,B,C} does not directly attack the default case because {A} already attacks the default case and {A} is more concise than {A,B,C}. The default case defines the assumed outcome for any new case unless the default case is *sufficiently attacked* by other cases in the case base. A *sufficient attack* against the default case occurs if all the unattacked nodes of the graph attack the default case (i.e., none of the unattacked nodes defend the default case).

For a new case, the outcome is decided by first determining which, if any, of the historical cases the new case attacks and subsequently inferring from the argument graph whether the default outcome is attacked or defended. If the default outcome is defended, then the outcome for the new case is the default outcome (in this example, negative). Argumentation rules define that a new case attacks a past case if the past case factors are not contained in the new case³. Here, the new case {A,B,C,D} does not attack any case because all historical cases are a subset of the factors of the new case. Since {A,B,C} is the closest *unattacked* case in the graph and by inference it attacks the default case, the outcome for the new case is positive.

Argumentation rules are also used to decide the outcome when multiple similar cases with differing outcomes exist in the case base. For instance, in the above example, if the case base

³This ensures that factors which are not present in the new case (and thus deemed irrelevant) do not contribute to the outcome

included a historical case such as $(\{A,D\},-)$, both $(\{A,D\},-)$ and $(\{A,B,C\},+)$ would be similar cases. By inference, the default outcome is now defended by at least one of the unattacked nodes (since $\{A,D\}$ attacks $\{A\}$) and is chosen as the outcome for the new case.

2. Research Plan

This section lists the research objectives, defines associated terminologies and describes the approaches considered. Fig 2 depicts an overview of the components in the research.

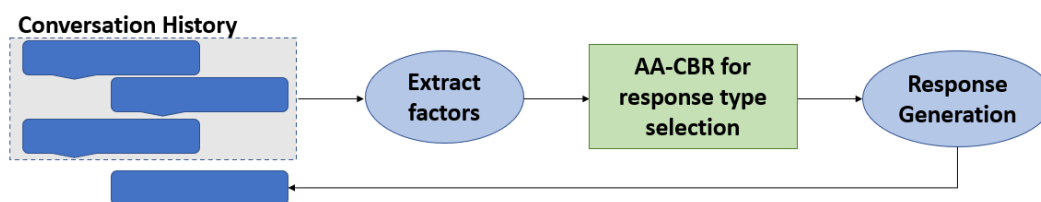


Figure 2: Overview of the research: CBR components in rectangle; Neural network components in oval

The research aims to build an interpretable conversational model for MI. An AA-CBR based approach is proposed to introduce interpretability when deciding the next response type. The following research questions will be investigated through the listed objectives:

1. How can an MI conversation be represented as a case of factors?
 - Identify a set of dialogue factors (i.e case attributes) to represent a given conversation history, which forms the problem component of the case.
 - Label each conversation as successful (good) or unsuccessful (bad), which forms the case outcome.
 - Identify a set of counsellor response types, which forms the solution for the case.
2. How can AA-CBR be applied for case retrieval?
 - Identify challenges in applying AA-CBR for MI conversations
 - Extend AA-CBR for MI conversations
 - Apply the extended AA-CBR framework and evaluate using a sample dataset

Definitions:

Case: A case comprises the entire available conversation history, represented as a set of dialogue factors and depicted as a node in the argument graph.

Dialogue factors: Dialogue factors can capture both relevant content such as the topic of the conversation as well as contextual features such as speaker sentiment and resistance or willingness to change (called MI talk-type). These will be annotated against each utterance.

Outcome: For MI, a good conversation outcome is either an explicit user expression of satisfaction at the end of a conversation or an implicit change in user perspective.

2.1. Approach / Methodology

Construct case base: First, the right set of dialogue factors that capture the separation between good and bad outcomes of an MI conversation need to be identified. Broadly, there are a few types of dialogue factors that may be considered - (i) frame of mind factors (eg. sentiment, LIWC markers [9], MI talk-type [10]) which are indicative of psychological state (ii) conversation topic (eg. addiction, weight-loss) (iii) linguistic factors (eg. utterance length, use of questions). AA-CBR assumes factors to be independent of each other while here, some of them such as sentiment and MI talk-type labels may be related and presence of one factor may entail the other. Such relationships will also need to be investigated. The final set of dialogue factors chosen will form the vocabulary knowledge container of the case base.

The quality of the AA-CBR framework will depend on the quality of the extracted factors and outcome labels. While public datasets such as [10] provide annotations for some factors, other factors may need to be either freshly annotated or predicted. Given the difficulty in obtaining expert annotations, this work will potentially adopt domain-transfer of well-studied models for classifying non-MI factors such as sentiment [11] since words indicating sentiment polarity are reasonably generalizable. Transparency of predictions will be enabled with explanation methods such as feature relevance scores [12]. The models will be trained for use in the continual learning setting [13] so that new data can be used to improve the model throughout its life. The overall approach towards case base construction is summarized as follows:

- Identify and define the dialogue factors to be used
- Annotate cases with factors (manually where expertise is available, or with existing models in literature, or by training a classifier on a few annotated cases)

Extend AA-CBR for response type selection: This work proposes to consider dialogue factors from available conversation history at each turn of a conversation as factors in the AA-CBR framework. Thus for a new case, the case representation would evolve with the unfolding of the conversation over time. It is worth noting that in reality not all possible factors may become available and new factors previously unseen in the case base may be added, which are both supported by the AA-CBR framework, making it suitable for this use case.

When retrieving a solution for a new case, some possible solutions (i.e. response types) are reflective, neutral or advice [10] and which of these is best suited will depend on the client's current frame of mind and other personality traits. For each of these possible solutions, similar cases can be retrieved and argumentation used to decide the outcome of choosing that solution. Accordingly, the response type(s) that leads to a positive predicted outcome can then be chosen for the next response from the counsellor. Thus, the AA-CBR framework argues why a particular response type should be chosen over other response types for a new client, based on the previous outcomes seen in the case base. Fig 3) depicts an example, where the default outcome for the response type of giving advice is taken to be positive. However, giving advice when the speaker sentiment is anger results in a negative outcome, and this node attacks the default outcome node. For the new conversation, the argument graph will result in proposing a positive outcome for choosing to advice, using the same logic as described in section 1.1.

It is likely that the AA-CBR framework may not be directly suitable for conversations. For instance, a crucial assumption in AA-CBR is that a particular combination of factors always

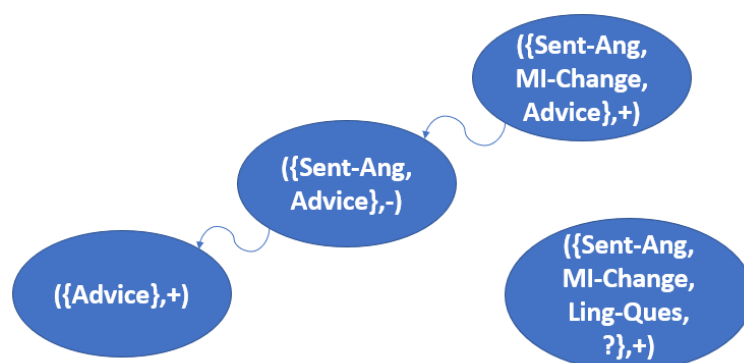


Figure 3: A simple example of the case structure and application of AA-CBR to MI: Each node represents a case; MI-Change is an expression of willingness to change; Linguistic-Question is the linguistic act of asking a question; Advice is the counsellor response type of providing advice

leads to the same outcome. However for conversations, this cannot be guaranteed in reality and outcomes may be probabilistic. Also, considering the temporal aspect, the same factors may appear at different times in the conversation and their ordering may result in different outcomes (eg. anger at the beginning vs end of the conversation are different). Therefore, the research will explore an extension of the AA-CBR framework to address such challenges.

Generate responses and evaluate: For a given conversation context, the next response-type as determined using the AA-CBR framework will be used as conditioning input to a suitable natural language generation model. The generated response will then be evaluated for:

- How well the response aligns to the given input response-type: by comparing the semantic similarity between outputs with and without the input conditioning.
- Whether the response-type conditioning can match the baseline performance using non-interpretable generative models as in [2], while providing interpretability.

3. Conclusion

This research proposes the use of AA-CBR for an interpretable modelling of motivational interviewing. The idea is to determine the response type at each counsellor turn in a new conversation by comparing it to similar conversations in the case base. Specifically, the case structure, i.e the representation of each conversation as a set of dialogue factors and the case structure evolution as the conversation progresses will be investigated. Further, approaches for extending the AA-CBR framework to allow for probabilistic attack relationships between cases and multi-outcome case representations will be explored and will form the major contribution of the research. The proposed work is currently in the initial stages and other research directions such as case adaptation in AA-CBR may also be explored in the future.

References

- [1] L. Tavabi, T. Tran, K. Stefanov, B. Borsari, J. Woolley, S. Scherer, M. Soleymani, Analysis of behavior classification in motivational interviewing, in: Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access, 2021.
- [2] S. Shen, C. Welch, R. Mihalcea, V. Pérez-Rosas, Counseling-style reflection generation using generative pretrained transformers with augmented context, in: Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue, 2020.
- [3] K. Resnicow, F. McMaster, Motivational interviewing: Moving from why to how with autonomy support, *The international journal of behavioral nutrition and physical activity* (2012).
- [4] N. Inui, T. Ebe, B. Indurkha, Y. Kotani, A case-based natural language dialogue system using dialogue act, in: IEEE International Conference on Systems, Man and Cybernetics, 2001.
- [5] K. Eliasson, An integrated discourse model for a case-based reasoning dialogue system, SAIS-SSL event on Artificial Intelligence and Learning Systems (2005).
- [6] K. Čyras, K. Satoh, F. Toni, Explanation for case-based reasoning via abstract argumentation, in: *Computational Models of Argument*, 2016.
- [7] K. Cyras, K. Satoh, F. Toni, Abstract argumentation for case-based reasoning, in: Fifteenth international conference on the principles of knowledge representation and reasoning, 2016.
- [8] P. M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, *Artificial Intelligence* (1995).
- [9] T. Althoff, K. Clark, J. Leskovec, Large-scale analysis of counseling conversations: An application of natural language processing to mental health, *Transactions of the Association for Computational Linguistics* (2016).
- [10] Z. Wu, S. Ballocu, V. Kumar, R. Helaoui, E. Reiter, D. R. Recupero, D. Riboni, Anno-mi: A dataset of expert-annotated counselling dialogues, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022.
- [11] N. Majumder, S. Poria, D. Hazarika, R. Mihalcea, A. Gelbukh, E. Cambria, Dialoguernn: An attentive rnn for emotion detection in conversations, *Proceedings of the AAAI Conference on Artificial Intelligence* (2019).
- [12] M. Danilevsky, K. Qian, R. Aharonov, Y. Katsis, B. Kawas, P. Sen, A survey of the state of explainable AI for natural language processing, in: Proceedings of the 10th International Joint Conference on Natural Language Processing, 2020.
- [13] M. Biesialska, K. Biesialska, M. R. Costa-jussà, Continual lifelong learning in natural language processing: A survey, in: Proceedings of the 28th International Conference on Computational Linguistics, 2020.