Novel features for capturing cooccurrence behavior in dyadic collaborative problem solving tasks

Vikram Ramanarayanan
Educational Testing Service R&D
90 New Montgomery St, #1500
San Francisco, CA
vramanarayanan@ets.org

Saad Khan
Educational Testing Service R&D
600 Rosedale Road
Princeton, NJ
skhan002@ets.org

1. INTRODUCTION

Research shows that complex interactive activities such as team work and collaboration are more effective when participants are not only engaged in the task but also exhibit behaviors that facilitate interaction [5]. Successful collaboration is often manifested in what is known as "entrainment" or convergence between the participants of such collaboration. In the educational context, entrainment between collaborators or between student and the tutoring system is important in understanding learning dynamics, learning gains and student performance in different learning environments [6]. Recently Luna Bazaldua et al. demonstrated a statistically significant synchronicity of cognitive and non-cognitive behavior between dyads engaged in online collaborative activity [1]. However, in their study participants were not able to see each other and only interacted over a text-based chat interface. This is an important point to note since the ability to converse face-to-face can significantly impact the nature of the dyadic interaction. Therefore, in this paper we focus on behavioral patterns of emotional expressions between dyads during face-to-face conversation through a video conferencing system. Our hypothesis is that dyads engaged in face-to-face collaborative activity demonstrate a significantly different pattern of behavior as opposed to nominal dyads who are artificially paired up with each other. Notation-wise, we use the term nominal dyad or artificial dyad interchangeably to mean two subjects whose data are analyzed as if they were interacting dyadically, but were actually not.

Explicitly modeling temporal information in such dyadic interaction data is important because each person's emotional state or behavior need not stay constant over the course of the interaction – they could get fatigued over time, or be more nervous at the very beginning (resulting in repetitive, cyclic fidgeting behavior), but gradually settle into a comfort zone later, as they get more familiar with the task and each other. For similar reasons their body language and emotional state can also fluctuate over the time series. However, current feature extraction approaches that aggregate information across time do not explicitly model temporal cooccurrence patterns; consider for instance that one person's emotional state – joy – generally follows his interlocutor's emotional state – say neutral – in a definitive pattern during certain parts of the interaction. Capturing such patterns might help us (i) explicitly understand the predictive power of different features (such as the occurrence of a given pair of emotions) in temporal context (such as how often did the emotional state of one person in the dyad occur given the previous occurrence of another emotional state of the other person in the dyad), thus allowing us to (ii) obtain features that are more interpretable on visual inspection. We would like to take an initial stab at bridging this gap in this paper. Specifically, we propose to adapt a feature based on histograms of cooccurrences [4] that was developed earlier for analyzing a single time-series (say, from one person), and extend it to the case of dyads (see Figure 1). The feature models how different "template" emotional states of one person in a dyad co-occur within different time lags of a "template" emotional states of the other person in the dyad over time. Such a feature explicitly takes into account the temporal evolution of emotional states in different interaction contexts.

2. DATA

2.1 The Tetralogue CPS Platform

We used an online collaborative research environment developed in-house – the Tetralogue [2, 1]. The participants, who may be in different locations, interact through an online chat box and system help requests (selecting to view educational videos on the subject matter). The main avatar, Dr. Garcia, introduces information on volcanoes, facilitates the simulation, and requires the participants to answer a set of individual and group questions and tasks. A second avatar, Art, takes the role of another student who shows his own answers to the questions posed by Dr. Garcia, in order to contrast his information with that produced by the dyad. Twenty-six subjects participated in this study and were paired in dyads using random selection.

3. ANALYSES AND OBSERVATIONS

In order to observe how well HoC features capture dyadic behavior, we randomly extracted 100 time-intervals (each 10 seconds long) from the post-processed and synchronized feature streams for all 26 subjects. We then computed HoC features for each of these intervals for each subject, respectively. Now recall that in this pool of subjects, each subject has one true dyad with whom they completed the Tetralogue task collaboratively. We hypothesize that the HoC features computed for true dyads will be significantly different as compared to the HoC features computed between artificial or nominal dyads (who did not actually engage in a dyadic interaction). We found that the distances computed between HoC features extracted from true dyads were significantly lower ($p \approx 0$) than those of distances between HoC features computed on artificial dyads. This finding suggests that (i) not only do true dyads engaged in a

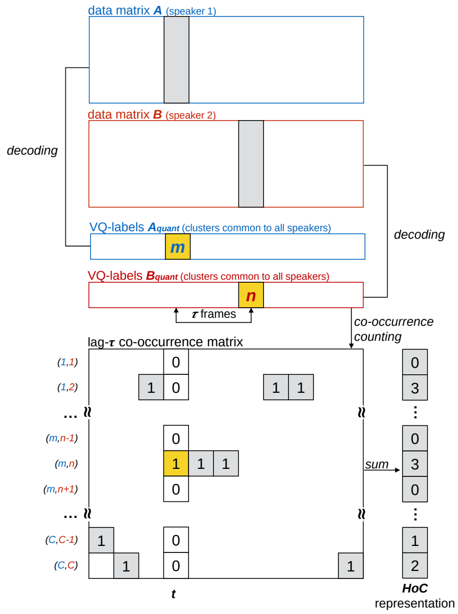

Figure 1: **Schematic depiction of the computation of histograms of cooccurrences (HoC) (adapted from [3]).**

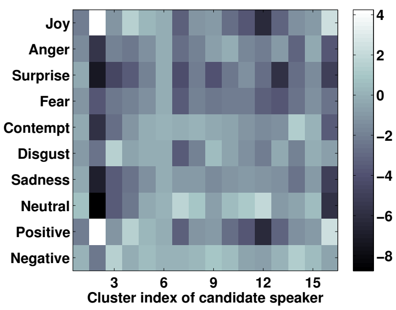

Figure 2: **Schematic illustrations of the emotion feature clusters computed for *all* speakers. Each column represents an emotional cluster centroid, which is a particular distribution of emotional state activations. There are 10 dimensions that describe an emotional state, represented by different rows. The colors represent the odds, in logarithmic (base 10) scale, of a target expression being present (typically range: $[-5,+5]$).**

collaborative interaction exhibit specific characteristic patterns of emotional state cooccurences that clearly sets them apart from artificial dyads, but (ii) such HoC features allow us to capture these differences in an effective manner.

Figures 2 and 3 gives us some more insight into why these features perform well. Figure 2 depicts the 16 cluster centroids computed on (and therefore common to) all speakers. Notice that each column of Figure 2 represents one cluster centroid, comprising different relative activation of different emotions – for instance, cluster 2 represents an emotional state with a higher activation of joy and positive emotion, while cluster 6 represents a more neutral emotional state, encompassing an equal (and approximately zero) activation of all emotions. Recall that these emotion clusters are common to *all* speakers. Figure 3 shows feature distributions of HoC features computed on one particular speaker and his/her actual dyadic partner, and those computed on that same speaker and an artificial dyadic partner. We observe that the feature distributions of the former are more peaky, with specific certain clusters of emotions

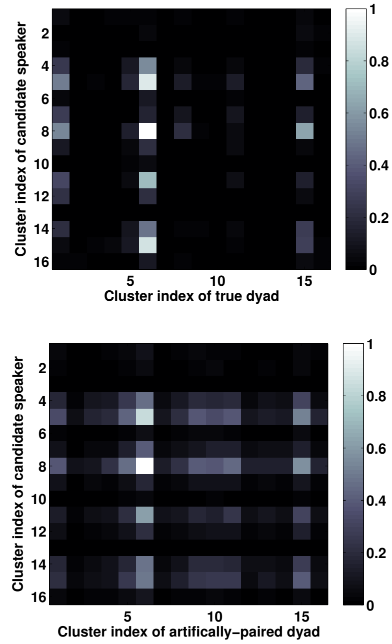

Figure 3: **Average HoC feature distributions (across lags) for the true and nominal dyad, respectively, of one particular speaker in the database. The color in the $(m,n)^{th}$ square represents the average normalized activation (between 0 and 1) of cluster m of the speaker represented along the y-axis co-occuring with cluster n of the speaker represented along the x-axis.**

co-occurring more often than others. However, in the case of the latter, this distribution is more flat and uniformly distributed. Note that while specific results shown in Figure 3 are particular to the chosen speaker, we observe the aforementioned trends are in general for all speakers. In other words, true dyads display specific patterns of behavioral cooccurrence and synchronicity that are not observed in artificial dyads, and such a HoC feature is helpful in understanding and bringing out these differences.

4. CONCLUSIONS AND OUTLOOK

This paper has made an initial attempt at proposing a novel feature, dubbed histograms of cooccurrences, that captures how often different prototypical behavioral states exhibited by one person cooccur with those exhibited by his/her partner over different temporal lags. We have shown that not only does this feature bring out the differences between dyads and non-dyads, but is also interpretable in that it tells us which behavioral states are most likely to occur in dyads as opposed to non-dyads.

5. REFERENCES

[1] D. L. Bazaldua, S. Khan, A. von Davier, J. Hao, L. Liu, and Z. Wang. On convergence of cognitive and non-cognitive behavior in collaborative activity. In *The 8th International Conference on Educational Data Mining (EDM 2015)*.

[2] L. Liu, J. Hao, A. A. von Davier, P. Kyllonen, and D. Zapata-Rivera. A tough nut to crack: Measuring collaborative problem solving. *Handbook of Research on Technology Tools for Real-World Skill Development*, page 344, 2015.

[3] V. Ramanarayanan, C. W. Leong, L. Chen, G. Feng, and D. Suendermann-Oeft. Evaluating speech, face, emotion and body movement time-series features for automated multimodal presentation scoring. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pages 23–30. ACM, 2015.

[4] V. Ramanarayanan, M. Van Segbroeck, and S. Narayanan. Directly data-derived articulatory gesture-like representations retain discriminatory information about phone categories. *Computer Speech and Language*, 36:330–346, 2016.

[5] A. A. Tawfik, L. Sanchez, and D. Saparova. The effects of case libraries in supporting collaborative problem-solving in an online learning environment. *Technology, Knowledge and Learning*, 19(3):337–358, 2014.

[6] J. Thomason, H. V. Nguyen, and D. Litman. Prosodic entraiment and Tutoring Dialogue Success. In H. Lane, K. Yacef, J. Mostow, and P. Pavlik, editors, *Artificial Intelligence in Education, AIED 2013*, pages 750–753. Springer, 2013.