5^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2019, University of Amsterdam, The Netherlands

Predicting Small Group Interaction Dynamics with Social Network Analysis

Gabriel Murray
University of the Fraser Valley
gabriel.murray@ufv.ca

Keywords: group interaction, turn-taking, social networks, meetings, machine learning

Extended Abstract

Given a meeting participant's activity level and interaction patterns at one point in a meeting, our aim is to automatically predict their activity level at a later point of the meeting. Such predictions could be useful for an automated meeting assistant that provides feedback on interaction patterns, with the goal of encouraging participation. The predictive models use features derived from a social network representation of each small group. The best automatic prediction models consistently outperform two baseline models at multiple time lags. We analyze which interaction features are most predictive of later meeting activity levels. Recent related work includes time-lagged prediction of speech rate [2] and prediction of group affect [1].

In these experiments we use the AMI meeting corpus[1], where each meeting consists of four people role-playing as a design team. Each group goes through a series of four meetings. For each meeting, we randomly choose one of the four participants as the target participant whose activity levels we will try to predict. Each meeting is segmented into non-overlapping windows of 20 dialogue act units each. Within each window, we represent the group interaction dynamics as a directed graph, with nodes representing participants. There is an edge (A,B) between participants A and B if there is at least one immediate transition within that window from A's speaking turn to B's speaking turn. The edge (A,B) has a cost which is the reciprocal of the number of times that transition was made within the current window.

After building the graph for the current window, we extract betweenness centrality, closeness centrality, and degree centrality features for each participant. We also extract the number of dialogue units by each speaker in the current window, and the location index in the meeting. We extract the same features from the preceding window as well, to provide additional context. We then predict the dialogue act frequency of the target participant after a time lag of n windows, where we vary n with the values 1,2,3,4,5. That later window is the *target window*.

There are 86 meetings in the training set and 52 in the test set. We ensure that each group has its meetings contained entirely in the training set or entirely in the test set, to make the prediction task more challenging. The number of actual observations in the training and test set depends on the value of the time lag n. For example, with $n = 1$, there are 3428/2088 training/test examples, and with $n = 5$ there are 3084/1880 training/test examples. We experiment with linear regression, random forest, and gradient boosting models. These models are contrasted with a baseline that predicts the mean dialogue act frequency from the training data, and a baseline that predicts that the target dialogue act frequency will be the same as the current dialogue act frequency. This latter baseline assumes that a participant's activity level will stay roughly the same between the current window and target window.

Figure 1 shows the results for all prediction models and all time lags. The three machine learning models are all significantly better than the baseline that predicts the same participation

[1] `http://groups.inf.ed.ac.uk/ami/corpus/`

levels in the target and current windows, according to a paired t-test on the squared residuals (all $p<0.05$), and these significant results are found across all time lags. That baseline gets much worse as the lag gets larger, meaning that the assumption that the participant will have the same participation level in the future as they currently do becomes an increasingly poor assumption as the time lag grows.

The linear regression model and gradient boosting model are the best overall prediction models, and perform very comparably to each other. They are significantly better than both baselines across all time lags (all $p<0.05$). However, the improvement over both baselines at lag 1 is very substantial (e.g. 32% relative reduction in MSE over the mean prediction baseline), while the improvement over the mean prediction baseline at lag 5 is very small. Generally, all of the prediction methods perform worse as the time lag between the current window and the target window gets larger, with the exception of the mean prediction baseline.

Based on average MSE reduction, the most useful features at lag 1 are the frequency features in the context windows, followed by the position in the group discussion and the closeness centrality features of all of the participants. The feature importance scores are much the same at lags 2 and 3, but at lag 4 the position in the discussion becomes the most useful predictor, and at lag 5 the closeness centrality of the target participant and the position in the discussion are the top features.

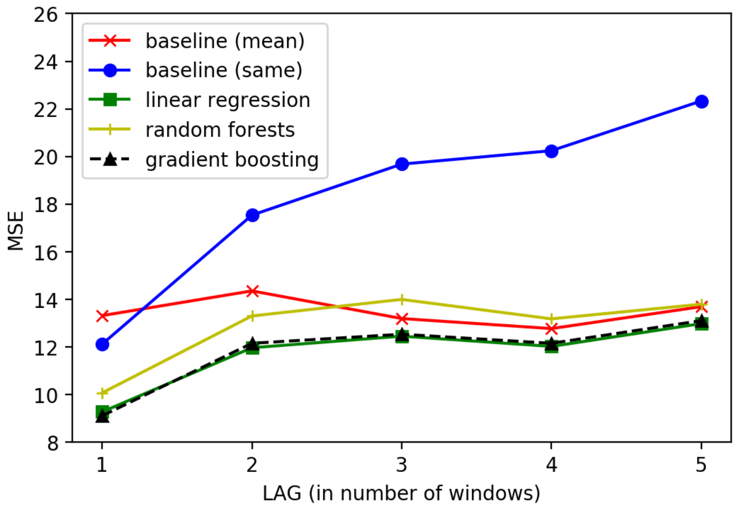

Figure 1: **Mean Squared Errors (MSE) for All Models & Time Lags**

Using features derived from social network representations of group interactions, we have shown that automatic predictive models can significantly outperform two baseline models in predicting the future activity level of meeting participants, even at long time lags.

References

[1] LAI, C., AND MURRAY, G. Predicting group satisfaction in meeting discussions. In *Proc. of the Workshop on Modeling Cognitive Processes from Multimodal Data, ICMI, Boulder, USA* (2018).

[2] STEWART, A., KEIRN, Z., AND D'MELLO, S. Multimodal modeling of coordination and coregulation patterns in speech rate during triadic collaborative problem solving. In *Proc. of ICMI, Boulder, USA* (2018), pp. 21–30.