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

Connecting Small Group Affect and Social Network Centrality Measures

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

Keywords: network centrality, group affect, team cohesion, interaction, machine learning

Extended Abstract

Given a group interaction in a meeting, we aim to automatically predict group affect levels such as overall satisfaction and sense of information overload, using features of the interaction patterns in the discussion. The group interactions are represented using social network analysis, and we derive centrality scores from these representations. We describe how the various centrality measures relate to different affective ratings by the meeting participants. This work relates to recent research on detecting emotion in conversations [1] and predicting group task performance based on features of the group interaction [2].

For these experiments, we use the AMI meeting corpus[1], in which groups of four participants go through a series of four meetings. The members of each group role-play that they are employees of a company tasked with designing a product and bringing it to market. After each meeting, each participant answered several questions regarding their sense of how the meeting went. Here we focus on three of the criteria, which they rated on a 1-7 scale:

- Q7: *Overall Satisfaction:* 'All in all, I am very satisfied.'
- Q16: *Attention Satisfaction:* 'All team members received sufficient attention.'
- Q15: *Information Overload:* 'There was too much information.'

For each criterion, we sum the individual ratings to get an aggregate group score, with a maximum group score of 28 for each criterion. Those three group scores are the outcome variables in these experiments.

Most of the predictive features we use are derived from graph representations of the interaction patterns in the meetings. Each meeting is divided into non-overlapping windows of 20 dialogue act units each. Within each window, we represent the group interaction as a directed graph, where nodes represent meeting participants. There is a directed edge (A,B) from participant A to participant B if, within the current window, there is at least one immediate transition from A's speaking turn to B's speaking turn. Edge (A,B) has a cost that is the reciprocal of the number of transitions between A and B within that window. We then extract three centrality measures for each participant: betweenness centrality, closeness centrality, and degree centrality. We also extract the number of dialogue act units for each participant in that window. These four measures are averaged over the entire meeting for each participant. Finally, for each measure we take the maximum, minimum, and mean of the averages over the participants, resulting in 12 features in total.

The machine learning models we use are gradient boosting, random forests, and linear regression. There are 120 meetings in total, and we carried out 5-fold cross-validation. The predictive models are evaluated using mean-squared error (MSE). Table 1 shows the MSE

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

for all models and all three prediction tasks. Random forests and linear regression perform comparably to one another, with linear regression being the best overall. The MSE scores are best on Q7, regarding overall satisfaction. The R^2 values for the linear regression models for all three tasks are around 0.22, indicating that the models explain $\sim 22\%$ of the variation in the data. While these modest values show room for improvement, they are impressive given that we are using a small number of centrality and frequency features, and no other verbal or nonverbal sources of information.

Rating	Linear Regression	Gradient Boosting	Random Forests
Q7	6.98	9.33	7.63
Q15	15.83	20.14	16.06
Q16	11.07	14.00	11.25

Table 1: **Mean-Squared Error (MSE) Scores**

We subsequently evaluate each feature using an importance score from the random forests models, which indicates how much each feature tended to reduce the MSE. For Q7, the most useful features are the maximum centrality score, minimum dialogue act frequency, and minimum betweenness. Figure 1 shows the relationship between maximum centrality and overall group satisfaction – there is a statistically significant negative correlation.

For Q15, the most useful features are minimum frequency, maximum betweenness, and mean frequency. For Q16, on attention sufficiency, the most useful feature is minimum frequency – an intuitive result – as well as mean frequency and minimum betweenness.

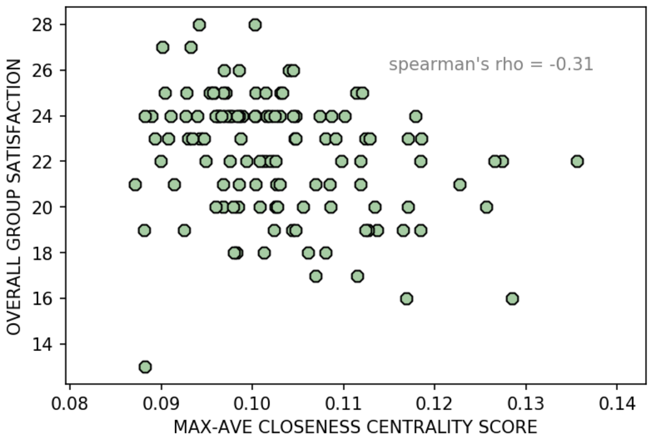

Figure 1: **Satisfaction Increases as Maximum Centrality Decreases**

Using centrality features derived from graph representations of small group interactions, we have automatically predicted three types of affective outcomes, and evaluated the usefulness of these features for each prediction task.

References

[1] HAZARIKA, D., PORIA, S., ZADEH, A., CAMBRIA, E., MORENCY, L.-P., AND ZIMMERMANN, R. Conversational memory network for emotion recognition in dyadic dialogue videos. In *Proc. of NAACL 2018, New Orleans, USA* (2018), pp. 2122–2132.

[2] MURRAY, G., AND OERTEL, C. Predicting group performance in task-based interaction. In *Proceedings of ICMI 2018, Boulder, USA* (2018), pp. 14–20.