Machine Learning for Meeting Analysis

Been Kim Cynthia Rudin
Massachusetts Institute of Technology
Cambridge, Masachusetts 02139
beenkim@csail.mit.edu, rudin@mit.edu

Abstract

Most people participate in meetings almost every day, multiple times a day. The study of meetings is important, but also challenging, as it requires an understanding of social signals and complex interpersonal dynamics. Our aim this work is to use a data-driven approach to the science of meetings. We provide tentative evidence that: i) there are common macro-patterns in the way social dialogue acts are interspersed throughout a meeting, and ii) it is often possible to predict whether a proposal during a meeting will be accepted or rejected based entirely on the language (the set of persuasive words) used by the speaker.

Introduction

"A meeting is indispensable when you don't want to get anything done." (Kayser 1990)

In the United States alone, an estimated 11 million meetings take place during a typical work day (Newlund 2012). Managers typically spend between a quarter and three-quarters of their time in meetings (Mackenzie and Nickerson 2009), and approximately 97% of workers have reported in a large-scale study (Hall 1994) that to do their best work, collaboration is essential.

In this work, we develop and use predictive modeling tools in order to provide a data-driven approach to the scientific study of meetings. Our specific goal in this work is to contribute insights to the new scientific field of meeting analysis, and more generally, to show that ML can be useful for exploratory study of meetings, leading towards to the eventual goal of increasing meeting productivity. We do not claim that our results definitively answer these questions, only that they yield hypotheses that can be tested more thoroughly through other surveys.

In our study, we use the most extensively annotated corpus of meetings data in existence, originating from the AMI (Augmented multi-party interaction) project (McCowan et al. 2005). This corpus is derived from a series of real meetings, controlled in the sense that each meeting has four participants who work as a team to compose a new design for a new remote control.

Learning Macro-Patterns in Meetings

We study the way in which social acts (positive or negative) interact with work-related acts, i.e., acceptance or rejection of proposals. More abstractly, we would like to know if there is a "quintessential representation" of interactions within a meeting, where a representation is a directed graph of social/work-related acts. To do this, we will present a discrete optimization approach that uses the notion of "template enumeration."

We selected the annotations that were meaningful in this context, namely: socially positive act, socially negative act, negative assessment act and positive assessment act. This allows us to focus on assessments (either social or work related), under the hypothesis that assessments have a generally more powerful effect on the trends in the conversation than other dialogue acts. That is, the assessments create a "macro-pattern" within the meeting that we want to learn. The selected data contain 12,309 dialogue acts, on average 130 acts per meeting.

We learn the directed graph by optimizing this regularized risk functional using discrete optimization:

$$\begin{aligned} &F(\text{template } t) \\ &\quad = \frac{1}{m} \sum_{\text{meetings } i} \min_{t_j \in \text{enum}(t)} [\text{dist}(t_j, \text{meeting } i)] \\ &\qquad + C_1 \text{length}(t) + C_2 \text{backw}(t). \end{aligned} \tag{1}$$

Intuitively, Equation 1 characterizes how well the set of meetings match the template using the first term (i.e. the edit distance the meeting and the closest template instantiation t_j, where each meeting is a string of social/work-related acts), and the other two terms are regularization terms that force the template to be sparser and simpler, with most of the edges pointing forwards in time. The length(t) is the number of nodes in template t. The value of backw(t) is the number of backwards directed edges in the template graph. (Note that it is possible to represent an arbitrary connected graph in terms of forward and backwards nodes if one is willing to duplicate nodes, as we are.) The set enum

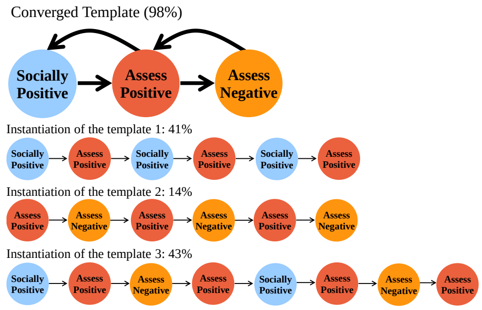

Figure 1: The template representing the interaction of social and work related acts.

is all *template instantiations* t_j of template t. We define a template instantiation as a path through the graph.

We optimize Equation 1 using simulated annealing, where the neighborhood is defined as a set of templates that are edit distance 1 away from the current template under allowable operations. The allowable operations include insertion of a new node between any two nodes or at the beginning or end of the template, deletion of a node, insertion or deletion of backwards directed edges.

We ran the algorithm starting from 95 different initial conditions, each initial condition corresponding to one of the true meetings in the database. The results were highly self-consistent, in the sense that in 98% of the total experimental runs, the algorithm converged to a template that is equivalent to, or a simple instantiation of, the one shown in Figure 1. This template has a very simple and strong message, which is that the next judgment following a negative assessment is almost never a socially positive act. The converse is also true, that a socially positive act is rarely followed by a negative assessment.

From a social perspective, this result can be viewed as somewhat counterintuitive, as one might imagine wanting to encourage a colleague socially in order to compensate for a negative assessment. In practice, however, sometimes positive social acts can sound disingenuous when accompanying a negative assessment. This can be demonstrated using the following pairs of dialog from within the AMI Corpus (that are not truly side by side): 1) "But I thought it was just completely pointless." "Superb sketch by the way." 2) "It'd be annoying." "Yeah, it was a pleasure working with you." In these cases a positive social act essentially becomes a negative social act.

Our algorithm is general and can be used in other domains to find a template characterizing sequences of events, where edges are mostly along one direction (e.g., forwards in time) allowing some backwards loops.

Learning Persuasive Words

A "good" meeting might be one where we have contributed to the team effort. We always want to suggest good ideas and want the team members to accept those ideas. Numerous articles claim that how we package our ideas, and our choice of words, is sometimes as important as the idea itself (Olsen 2009; Carlson 2012). We are interested in understanding this hidden factor in the team's decision making process. Are there patterns in suggestions that are accepted versus rejected? Can we use this to improve how we present ideas to the team?

We chose a bag-of-words representation for each "suggestion" dialogue act. We gathered a set of all words that occurred in all suggestions, excluding stop words, leading to a 1,839 dimensional binary vector representation for each suggestion (one feature per unique word). The labels were determined by the annotations, where accepted suggestions received a $+1$ label, and rejected suggestions received a -1 label.

We (i) used SVM coefficients to discover persuasive words from data alone, (ii) compared the SVM results to the published lists of persuasive words (Olsen 2009; Carlson 2012) and (iii) used Fisher's exact test to analyze each word independently.

Method (i) is to apply a linear SVM with cross-validation for the tradeoff parameter, and examine the values of its coefficients; this provides a ranking of features, and a rough measure of how important each word is to the overall classifier (Guyon et al. 2002). The set of 2,324 suggestions with 1,839 features was separated into 5 folds using each in turn as the test set, and the SVM accuracy was 83% $\pm$ 2.1%. Note that the SVM has no prior knowledge of what persuasive words are. Persuasive words identified by the SVM are considered to be words with large absolute coefficient value (normalized) over all 5 folds. Many of the persuasive words concerned the topic of marketing ("market," "presentation," "gimmick," "logo") and non-persuasive words included "buttons," "speech," "LCD," "recognition" and more often words specific to the topic of creating a remote control.

The next question (ii) is whether these learned persuasive words make sense. To answer this, we checked whether the persuasive words from published lists of persuasive words (Olsen 2009; Carlson 2012) had positive or negative SVM coefficients. SVM agrees with the published lists on "free," "good," "power," "avoid," "offer" and "save," but disagrees on "drive," "easy," "health," "guarantee," "explore," "money" and "purpose."

Finally, for (iii), Fisher's Exact test considers whether the suggestions containing the feature have significantly different proportions of acceptance and rejection. Using 0.1 as a threshold pvalue, the test selected 214 features. After this feature reduction, we applied SVM and achieved a test accuracy of 87.8% $\pm$ 1.2%, which is higher than before. Top ranked words include "fair," "speech," "possibilities," "selecting," "draw" and "amount." These are words that we believe are truly closer to being persuasive words, as they are persuasive both individually and together: they are each individually significant in predicting accepted suggestions, and they have the largest predictive coefficients.

References

Carlson, G. 2012. Uplifting word list: http://www.freelance-copy-writing.com/uplifting-word-list.html.

Guyon, I.; Weston, J.; Barnhill, S.; and Vapnik, V. 2002. Gene selection for cancer classification using support vector machines. *Mach. Learn.* 46(1-3):389–422.

Hall, J. 1994. Americans know how to be productive if managers will let them. *Organizational Dynamics*.

Kayser, T. 1990. *Mining group gold: How to cash in on the collaborative brain power of a group.* Serif Publishing.

Mackenzie, A., and Nickerson, P. 2009. *The time trap: The classic book on time management.* Amacom Books.

McCowan, I.; Carletta, J.; Kraaij, W.; Ashby, S.; Bourban, S.; Flynn, M.; Guillemot, M.; Hain, T.; Kadlec, J.; Karaiskos, V.; et al. 2005. The AMI meeting corpus. In *Proc. Methods and Techniques in Behavioral Research*, volume 88.

Newlund, D. 2012. Make your meetings worth everyone's time: http://www.usatoday.com/uscp/pni/business/2012-06-20-pni0620biz-career-getting-aheadpnibrd_st_u.htm.

Olsen, H. 2009. 108 persuasive words: http://www.high-output.com/uncategorized/108-persuasive-words/.