Discovering Group Nonverbal Conversational Patterns with Topics

Dinesh Babu Jayagopi[1,2] and Daniel Gatica-Perez[1,2]
[1] Idiap Research Institute, Martigny, Switzerland
[2] Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland
(djaya,gatica)@idiap.ch

ABSTRACT

This paper addresses the problem of discovering conversational group dynamics from nonverbal cues extracted from thin-slices of interaction. We first propose and analyze a novel thin-slice interaction descriptor - a bag of group nonverbal patterns - which robustly captures the turn-taking behavior of the members of a group while integrating its leader's position. We then rely on probabilistic topic modeling of the interaction descriptors which, in a fully unsupervised way, is able to discover group interaction patterns that resemble prototypical leadership styles proposed in social psychology. Our method, validated on the Augmented Multi-Party Interaction (AMI) meeting corpus, facilitates the retrieval of group conversational segments where semantically meaningful group behaviours emerge, without the need of any previous labeling.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing

General Terms

Human Factors

Keywords

Meetings, Characterizing groups, Nonverbal Cues

1. INTRODUCTION

Characterizing group conversations using nonverbal behaviour is a key problem in human interaction modeling, with applications related to browsing and retrieval of specific conversations where certain types of behaviours are exhibited. Modeling group interaction is challenging both in social science [9] and in computing, where methods for understanding group conversations from audio and visual nonverbal cues have started to become popular, motivated by the fact that nonverbal behaviours carry a wealth of information about the group members' relationships (in parallel to the spoken words) that can be decoded from relatively limited observations (thin-slices of interaction).

Various supervised learning approaches have been explored to characterize group conversations using a suite of nonverbal features. Several authors have employed layered sequential approaches to structure group conversations into group meeting activities (e.g., [12, 4]), where the first layer modeled the individuals' behaviour, and the second layer the interaction type (monologue, presentations, or discussions). Other authors have for instance classified conversations into conversational regimes (convergence or monologue, dyad-link and divergence) [10].

Overall, little work has been done on unsupervised learning of group conversational patterns [1], where emerging patterns are discovered rather than predefined by hand. Our paper addresses this issue, presenting a discovery method based on the use of probabilistic topic models (specifically Latent Dirichlet Allocation (LDA)), which are generative models originally used in text modeling to cluster documents according to semantic or topic similarity.

The contributions of this paper are threefold. First, we define a novel descriptor of thin-slices of group interaction - a bag of group-nonverbal-patterns (NVP) - that represents the nonverbal characteristics of a group as a whole. The bag-of-NVP allows robust fusion of nonverbal cues and tolerates variations in the number of group participants. Second, we show that LDA extracts meaningful topics that, after manual inspection, could be related to classical concepts in social psychology regarding leadership styles in groups [8]. Finally, we provide an analysis of nonverbal group dynamics emerging at different time scales to understand the effect of the time-interval of observation (the "thin-slice width") w.r.t. both NVP representation and topic discovery.

Sections 2 discusses our method in detail. Section 3 introduces our experimental setup, and presents and discusses the results. Section 4 provides the conclusions of our analysis.

2. OUR APPROACH

Figure 1 shows the overview of our work. First, we extract low-level nonverbal cues from thin-slices of small-group meeting data. Second, we quantize these cues into a bag-of-NVPs. Finally, we input these "documents" (bags-of-NVPs words) into a probabilistic topic model to discover patterns of group nonverbal behaviour.

Various nonverbal cues are known to be correlated with interpersonal relations [5]. A bag-of-NVPs representation facilitates fusion of individual cues, and makes the cues more robust as compared to raw low-level cues. Also, the use of group NVPs facilitates the comparison of groups of varying

ICMI-MLMI'09, November 2–4, 2009, Cambridge, MA, USA.

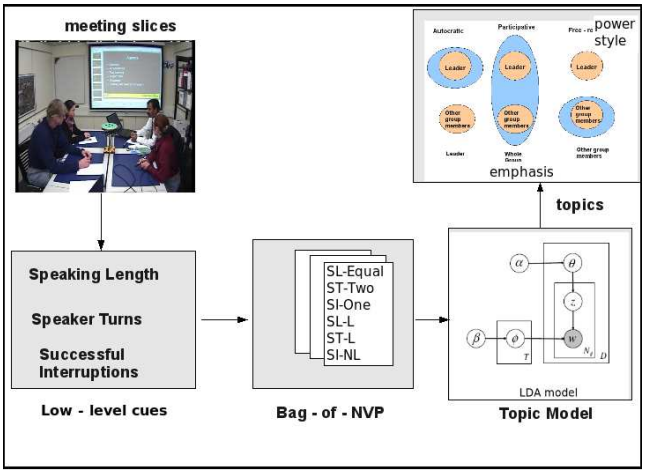

Figure 1: Overview of the group NVP discovery process using topic models.

sizes, and the use of principled methods for unsupervised learning like topic models.

Our bag-of-NVPs describes the conversational patterns of the group as a whole. Floor occupation related features signal various behaviours related to social verticality [5]. We extract speaking activity-based cues and construct the bag-of-NVP as follows.

Low-level cue extraction: From any given group conversation recording (with close-talk microphones), we extract two vocalic cues for the N_p participants.

Speaking energy: Using a window of 40 ms with a 10 ms time shift, the real-valued speaker energy is computed for each participant at each time step.

Speaking status: Traditionally, speaking status is computed from the speaking energy, by thresholding the energy values. In our work we employ a slightly sophisticated approach which deals with the presence of cross-talk efficiently [13]. This binary variable indicates the speaking / non-speaking (1/0) status of each participant at each time step. A turn is a continuous period of time for which the person's speaking status is 1.

Alternatively, the speaking status could also be obtained by speaker diarization on far-field microphone data. Then, *Speaking Length* (TSL) - total time that a participant speaks, *Speaking Turns* (TST) - the accumulated number of turns over the entire meeting for a participant, and *Successful Interruptions* (TSI) - the cumulative number of times that a participant starts talking while another participant speaks, and the latter finishes his turn before the former does are computed.

Let **TSL** denote the vector composed of N_p elements, whose elements are TSL for each participant after normalization (elements sum up to 1). We employ an analogous notation for **TST** and **TSI**.

Bag-of-NVPs generation: Our bag model includes two types of words discussed in the following.

Generic group patterns: We quantize each of the vectors **TSL**, **TST**, **TSI** into one of the five classes - *Silence*, *One*, *Two*, *Rest*, *Equal* - to describe a group. The class depends on whether silence ('0'), one-person ('1'), two-persons ('2'), three or more ('3'), or all participants ('4') share most of the probability mass for a particular nonverbal cue. The goal is to map a joint cue over a thin slice (e.g. speaking length) into a prototypical case (e.g. an interaction pattern in which all people talk about the same time, one person speaks most of the time, etc.) where identity is not important, and therefore making the description generic. The actual rule is described as follows: Let $SortedVector$ represent the input vector after sorting in descending order. The output is '1' if the first element of $SortedVector$ satisfies the following condition: $SortedVector(1) > 2 * \frac{1}{N_p}$. The output is '2' if $SortedVector(1) + SortedVector(2) > 3 * \frac{1}{N_p}$. and the output is '4' if $SortedVector(N_p) > \Delta$, where Δ represents a small interval. Finally, the output '3' is used as a catch-all class. Figure 2 shows an example histogram ($SortedVector$) for each of the classes other than silence for a group with $N_p = 4$ participants.

The 15 words corresponding to the generic groups patterns are *SL-Silence*, *SL-One*, *SL-Two*, *SL-Rest*, *SL-Equal*; *ST-Silence*, *ST-One*, *ST-Two*, *ST-Rest*,*ST-Equal*; and *SI-Silence*, *SI-One*, *SI-Two*, *SI-Rest*, *SI-Equal*.

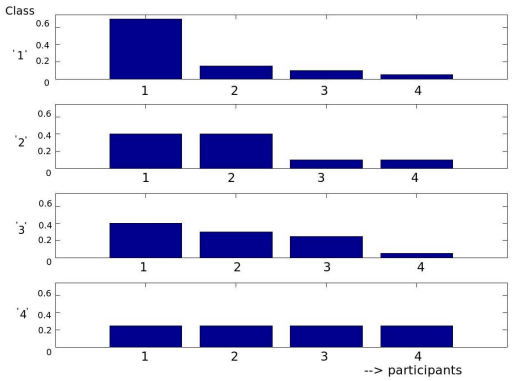

Figure 2: Example joint histograms for each of the NVP words other than *Silence*.

Leadership patterns: Very often there are meetings with a designated leader (e.g. a manager). Previous works have shown that speaking length correlates to dominance [6], speaker turns correlates with role-based status [7] and successful interruption signal real status and power [11]. All these behavioural aspects closely relate to social verticality [5] in groups. We add two more words for each of the 3 sets of features to indicate whether the designated leader ('L') or someone else ('NL') is the one who has the maximum. For example the presence of *SL-L* means that in this time slice, the leader has the maximum speaking length. Together with the words that characterize the generic group patterns, these words describe the position of the leader. The 6 words corresponding to the leadership patterns are *SL-L*, *SL-NL*; *ST-L*, *ST-NL*; *SI-L*, *SI-NL*.

The overall size of the vocabulary is 21 and each document contains six words. An important advantage of such a representation is that it is independent of the number of participants and hence allows the comparison of groups of different sizes. The size of the vocabulary can be increased by considering more nonverbal cues that are of behavioural interest, in a similar fashion.

Latent Dirichlet Allocation (LDA): Topic models are probabilistic generative models that were originally used in text modeling. In Latent Dirichlet Allocation [2], a text document is modeled as a distribution over topics, and a topic as a multinomial distribution over words. The topics discover patterns based on word co-occurrence.

Let there be D documents in a corpus and let a document contain N_d words. The probability of a given word w_i assuming T topics is $P(w_i) = \sum_{t=1}^{T} P(w_i|z_i = t)P(z_i = t)$, where z_i is a latent variable indicating the topic from which the i^{th} word was drawn. Each document is generated by choosing a distribution over topics $P(z = t) = \theta_t^{(d)}$. Each topic is characterised by a word distribution $P(w|z = t) = \phi_w^{(t)}$. In LDA, $P(\theta)$ is a Dirichlet(α) and $P(\phi)$ is a Dirichlet(β), where α and β are hyperparameters. We use Gibbs sampling to infer the parameters $\theta_t^{(d)}$ and $\phi_w^{(t)}$ and then interpret the T topics using the top words (with highest probability), and the documents as mixture of these topics.

3. EXPERIMENTS AND RESULTS

Dataset: Our dataset is 37 meetings (approximately 17 hours) from the Augmented Multi-Party Interaction (AMI) corpus [3] consisting of 10 groups of participants. Each group consists of 4 participants, who were given the task of designing a remote control over a series of meeting sessions. Each participant was assigned distinct roles: 'Project Manager', 'User Interface specialist', 'Marketing Expert', and 'Industrial Designer'. To encourage natural behaviour, the meetings were not scripted and the teams met over several sessions to achieve the common goal. We assume the project manager to be the leader of the group.

Distribution of bag-of-NVPs on AMI meeting slices: Figure 3 visualizes the distributions of the generic group patterns of **TSL**, **TST** and **TSI** among the 5 classes ('0' to '4') at different time scales. It is interesting to observe that the group interactions look more like a monologue at finer time scales (e.g. 1 min) and like a discussion at coarser time scales (e.g. 5 min), as observed through speaking length and speaker turns. Also, successful interruptions are not very common at fine time scales, as seen by the significant mass at class 0. Figure 4 shows the distribution of leadership patterns at two different time scales. If all the four participants had equal status (egalitarian groups) the probability mass at 'L'(resp. 'NL') should be close to 0.25 (resp. 0.75). The distribution shows that the average statistics of AMI data is close to egalitarian at several time scales, though individual leaders could have different styles, which we discover using the LDA model.

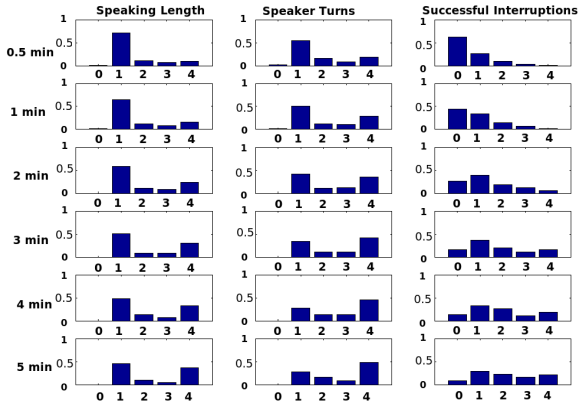

Figure 3: Empirical distribution of generic group patterns at different time scales.

In the experiments, we use 5-minute and 2-minute meeting slices. The number of documents for 2-minute slices is 501

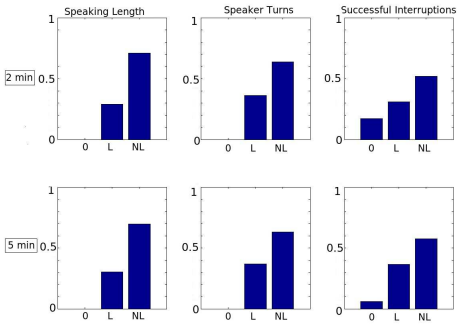

Figure 4: Empirical distribution of leadership patterns at two different time scales. '0' corresponds to the case when there is silence, 'L' (resp. 'NL') when leader (resp. someone else) has maximum feature value.

Topic 1 - LDA		Topic 2 - LDA		Topic 3 - LDA	
$P(z) = 0.32$		$P(z) = 0.33$		$P(z) = 0.35$	
'Autocratic'		'Participative'		'Free-rein'	
Word	$P(w\|z)$	**Word**	$P(w\|z)$	**Word**	$P(w\|z)$
SL-L	0.2	ST-Equal	0.25	SL-One	0.22
ST-L	0.2	SL-Equal	0.18	SL-NL	0.19
SI-L	0.16	ST-NL	0.15	SI-One	0.16
SI-Two	0.11	SL-NL	0.15	SI-NL	0.16
SI-Rest	0.08	SI-NL	0.14	ST-NL	0.13
ST-Rest	0.07	SI-Equal	0.12	ST-One	0.12

Table 1: LDA based discovery at 5-minute scale.

(without any overlapping slices) and 5-minute slices is 873 (using overlapping slices). We set $\Delta = 0.05$.

LDA-based pattern discovery at 5-minute scale: The LDA hyperparameters were set to standard values. We applied our LDA-based discovery procedure varying the number of topics T; we report the results using $T = 3$ topics. Table 1 shows the resulting top 6 words for each of the topics. Looking at the top words of Topic 1 (*SL-L*, *ST-L*, *SI-L* terms), it resembles a meeting where the leader is dominant or autocratic (talks more, more often, and interrupts more) and hence the title 'autocratic'. Topic 2 seems to characterize an egalitarian or participative meeting (top words being *ST-Equal*, *SL-Equal*), whereas Topic 3 represents a meeting where there is a single dominant person who, interestingly, is not the leader (top words being *SL-One*, *SI-One*, *SL-NL*, *ST-NL*, *SI-NL*). Based on manual inspection these patterns for the project managers of AMI meeting slices discovered for $T = 3$ topics seem to resemble the three classic leadership styles of Lewin et al. [8] as illustrated in Figure 5. The three styles - 'autocratic', 'participative', and 'free-rein', differ according to the emphasis (in terms of power) it places on the leader, the whole group, or the rest of the group.

Using the above representation, Figure 6 shows the average topic distribution (over all documents) for the 10 groups of participants. As one can observe, different groups have different signature distribution of topics. For example, groups 1,2 seem to have an autocratic leader, whereas groups 4, 9, 10 are more participative than others.

LDA-based pattern discovery at 2-minute scale: The same experiments were repeated with $T = 3$ topics on 2-minute meeting slices (see Table 2). For the case of the 'free-rein' topic, the top six words are exactly the same though in a different order. For the other two topics, we observe that the words in 'autocratic' and 'participative' topics are also similar to those of the 5-minute case (SL and ST related words are the same). A new word *SI-Silence* becomes

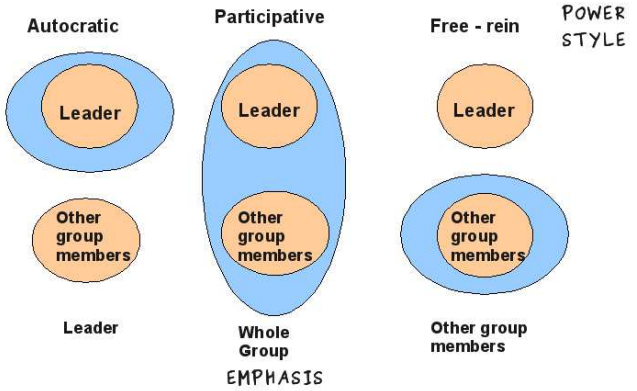

Figure 5: Leadership styles by Lewin et al. The blue envelope shows the emphasis (in terms of power) that is placed on the various group members.

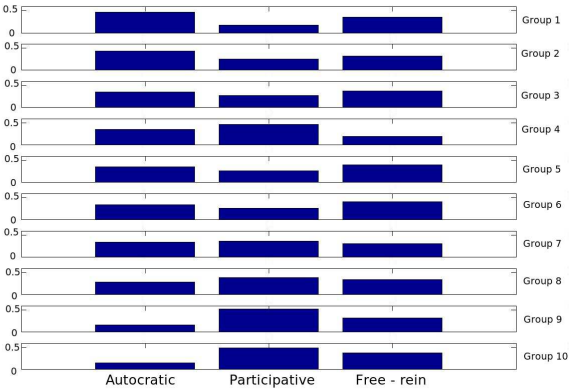

Figure 6: Average topic distribution over groups at 5-minute scale.

significant at 2-minute and appears in 'autocratic' topic.

Figure 7 shows the average topic distribution for the 10 groups of participants at 2-minute scale. As compared to the 5-minute case, the distribution is more balanced across the three topics. This suggests that the interaction styles (as defined here in terms of discovered topics) seem to emerge more strongly over longer intervals of time. Nevertheless, in a few cases some trends are stable. For instance, groups like group 4 which are more participative than other groups at both 5-minute and 2-minute scales, make a more egalitarian group, as compared to group 1 which looks very autocratic at both scales.

4. CONCLUSIONS

In this paper we investigated the problem of discovering nonverbal group patterns using topic models. We proposed a novel bag-of-NVPs approach to characterize groups. Our model described the groups in terms of the joint distribution

Topic 1 - LDA		Topic 2 - LDA		Topic 3 - LDA	
$P(z) = 0.31$		$P(z) = 0.35$		$P(z) = 0.34$	
'Autocratic'		'Participative'		'Free-rein'	
Word	$P(w\|z)$	**Word**	$P(w\|z)$	**Word**	$P(w\|z)$
ST-L	0.22	ST-Equal	0.19	SL-NL	0.30
SL-One	0.20	SI-L	0.16	SI-NL	0.19
SL-L	0.19	SL-Equal	0.12	ST-NL	0.18
ST-One	0.15	ST-NL	0.12	SI-One	0.16
SI-Silence	0.13	SI-Two	0.10	SL-One	0.11
ST-Two	0.06	ST-Rest	0.06	ST-One	0.07

Table 2: LDA based discovery at 2-minute scale.

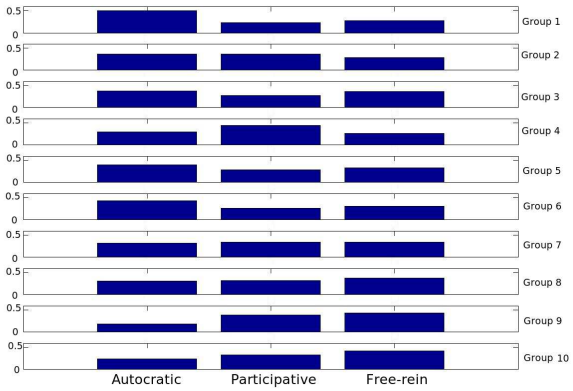

Figure 7: Average topic distribution over groups at 2-minute scale.

of speaking length, speaker turns, and successful interruptions among participants, and added words in the bag to describe the position of the leader with respect to the other members. Using an LDA model, we infered the topic distributions and word distributions for each topic using meeting slices of 5 minute and 2 minute duration. For a small number of topics, our model discovers patterns that seem to mimic three classic leadership styles - autocratic, participative and free-rein. In the future, we propose to validate our findings with manual annotation on a subset of meetings. We would also like to expand our bag-of-NVP to other nonverbal cues and to investigate other topic models capable of jointly modeling topics and the groups themselves as variables.

Acknowledgments: This research was partly supported by the EU project AMIDA and the Swiss NCCR IM2.

5. REFERENCES

[1] S. Basu et al. Towards measuring human interactions in conversational settings. In *Proc. IEEE CVPR Workshop on Cues in Communication*, 2001.

[2] D. Blei et al. Latent Dirichlet Allocation. *J. Machine Learning Research*, 3:993–1022, January 2003.

[3] J. Carletta et al. The AMI meeting corpus: A pre-announcement. In *Proc. Workshop on Machine Learning for Multimodal Interaction (MLMI)*, Edinburgh,UK, Jul. 2005.

[4] A. Dielmann et al. Automatic meeting segmentation using dynamic bayesian networks. *IEEE Transactions on Multimedia*, 9(1):25, 2007.

[5] J. Hall et al. Nonverbal behavior and the vertical dimension of social relations: A meta-analysis. *Psychological Bulletin*, 131(6):898–924, 2005.

[6] D. Jayagopi et al. Modeling dominance in group conversations using nonverbal activity cues. *IEEE Transactions on Audio, Speech and Language Processing, Mar 2009.*

[7] D. Jayagopi et al. Predicting two facets of social verticality in meetings from five-minute time slices and nonverbal cues. In *Proc. ICMI*, Chania, Greece, Oct. 2008.

[8] K. Lewin et al. Patterns of aggressive behavior in experimentally created social climates. *Twentieth Century Psychology: Recent Developments in Psychology*, 1946.

[9] J. McGrath. *Groups: Interaction and Performance*. 1984.

[10] K. Otsuka et al. Automatic inference of cross-modal nonverbal interactions in multiparty conversations.. In *Proceedings of ICMI*, pages 255–262. ACM New York, NY, USA, 2007.

[11] B. Raducanu et al. You are fired! Nonverbal role analysis in competitive meetings. In *Proc. ICASSP, Taiwan*, 2009.

[12] D. Zhang et al. Modeling individual and group actions in meetings with layered hmms. In *IEEE Transactions on Multimedia*, volume 8, pages 509–520, June 2006.

[13] J. Dines et al. The segmentation of multi-channel meeting recordings for automatic speech recognition. In *Proc. Interspeech*, 2006.