Multimed Tools Appl (2012) 56:207–226
DOI 10.1007/s11042-010-0545-8

Inferring competitive role patterns in reality TV show through nonverbal analysis

Bogdan Raducanu · Daniel Gatica-Perez

Published online: 8 June 2010

Abstract This paper introduces a new facet of social media, namely that depicting social interaction. More concretely, we address this problem from the perspective of nonverbal behavior-based analysis of competitive meetings. For our study, we made use of "The Apprentice" reality TV show, which features a competition for a real, highly paid corporate job. Our analysis is centered around two tasks regarding a person's role in a meeting: predicting the person with the highest status, and predicting the fired candidates. We address this problem by adopting both supervised and unsupervised strategies. The current study was carried out using nonverbal audio cues. Our approach is based only on the nonverbal interaction dynamics during the meeting without relying on the spoken words. The analysis is based on two types of data: individual and relational measures. Results obtained from the analysis of a full season of the show are promising (up to 85.7% of accuracy in the first case and up to 92.8% in the second case). Our approach has been conveniently compared with the Influence Model, demonstrating its superiority.

Keywords Social interaction · Competitive meetings · Role analysis · Nonverbal cues

B. Raducanu (✉)
Computer Vision Center, Edifici "O"—Campus UAB, 08193 Bellaterra, Barcelona, Spain
e-mail: bogdan@cvc.uab.es

D. Gatica-Perez
IDIAP Research Institute, Rue Marconi 19, 1920 Martigny, Switzerland
e-mail: gatica@idiap.ch

D. Gatica-Perez
École Polytechnique Fédérale de Lausanne (EPFL), 1015, Lausanne, Switzeland

1 Introduction

While social media usually refers to web-based systems involving content and relationships, we believe that another type of social media is the one which depicts social interaction in professional, entertainment, or amateur settings. The interest in analyzing social constructs appearing in media is relatively recent, and in particular algorithms to automatically recognize, predict, or discover behavioral traits or outcomes from social interaction have begun to appear [15, 31]. The understanding of fundamental principles that govern a person's status in groups is of primary relevance for several social sciences and would pave the way to create automatic tools to support research in social and organizational psychology [2, 28].

Social interaction can be defined as a dynamic sequence of social actions between individuals who modify and adapt their behavior according to those of their partners. Social action is a concept that refers to the interaction between individuals in society and it is used to observe how certain behaviors are modified in certain conditions [43]. Our behavioral patterns have neither universal nor isolated character, but a circumstantial and relational one. On one hand, our behavior has an individual component, specific to each person, characteristic of one's personality. On the other hand, behavior has a relational component, defined by the interaction with other people. A very important variable used in psychology and sociology to characterize social interaction is the role. The term role is associated with a person's position in a group (status), with the obligations and rights it implies [22]. Some roles are ascribed and others achieved. In this latter case, role can be seen as an emergent property of the interaction with other people. It is worth clarifying that role is not synonym with behavior, although both are interrelated. There are several other variables that influence and define our role in a group meeting: the type of the meeting (informal or competitive), our position in the group, the structure of the group (if the hierarchy is well-defined or if the group is more homogenous), the degree of familiarity between the people in the group, the emotional load (reflected in the mood) of each participant, etc. In consequence, the same person can play different roles in different situations.

As stated in [32], social interaction can be addressed in two frameworks. One of them comes from linguistics and addresses the problem of social interaction from the perspective of dialog understanding. The other one comes from nonverbal communication. Within this framework, nonverbal communication is used to get hints about personal behavior. Facial expressions, gaze, voice prosody, body gestures provide powerful cues to display and perceive engagement, persuasion, mirroring, dominance, etc.

In this paper we add a novel dimension to the automatic analysis of social interactions by studying the openly competitive scenario. More concretely, we address the problem of role analysis in competitive meetings using nonverbal behavior. "The Apprentice" US TV series offers an attractive scenario for our purpose. "The Apprentice" is a reality television show hosted by Donald Trump [39]. Dubbed as "The Ultimate Job Interview", the show features a set of business people participating in an elimination-style competition for a one-year, $250,000 salary to run one of Trump's companies. The winner of the competition is called "The Apprentice". The show represents a unique data set for the study of social interaction in competitive meetings. Being a reality-show, the behavior and reactions of the participants are

naturalistic displaying a high degree of involvement. Participants are real people (not actors) competing for a real goal. In the elimination process, the outcome is not known a priori.

The main contributions of this paper are: (i) for the first time, it addresses the automatic nonverbal analysis in the context of competitive meetings; (ii) the analysis is based on a novel dataset, namely "The Apprentice" TV show; (iii) although we followed simple cues to characterize the interactional process, these proved to give good results for the prediction of different tasks.

The paper is organized as follows. In Section 2 we present an overview of the literature on social interaction from two perspectives: psychological and computational. In Section 3, we present the data set used in our study. In Section 4 we define the research tasks and the experimental results followed by their discussion. Finally, in Section 5, we draw our conclusions.

2 Literature overview

This section provides an overview of the concept of role from two perspectives: psychological and computational.

2.1 Psychological perspective

The concept of role is a classical one in social sciences. Goffman saw roles from a dramaturgical perspective [18] . In his opinion, we are all 'actors', playing a role in our everyday life. Besides this metaphorical interpretation, in social psychology there are three major views regarding roles:

1. as the expectation of a specific behavior a person is supposed to perform;
2. as a characteristic associated with a person's position (status) in a group;
3. as the enacted behavior of individuals in a particular situation.

One of the exponents of the first view is Bormann. In [8] he claims that "when the other members know what part a person will play, and that person knows what part they expect of her or him, that person assumed a role" (p. 161). This statement is known as the "trait framework". However, according to Salazar [35], this view suffers of some limitations. Firstly, because it might happen that a person might play no role at all in a group. Secondly, since roles are more of a result of the expectation of others, the individual's actual behavior might play a minor part in determining the individual's role. Finally, the third limitation concerns how the expected outcome of the role is perceived. At this point, the following question arises: Do individuals share similar expectations? In the most general case, the expectations might vary [35].

The second view is due to Katz et al. [25] and McGrath [28], according to whom the roles are equivalent with position (status) in a group or organization. According to [25], a role is "a set of expected activities associated with the occupancy of a given position" (p. 200). Similarly, in [28] the same idea is expressed: "role is not a characteristic of a particular person, but rather is a characteristic of the behavior of the incumbent of a particular position" (p. 249). In this light, Katz et al. state that a role is defined by a four-stage sequence: expectations, sending expectations, receiving expectations and behaving [25]. But this view also has its limitations. The

first limitation is based on the assumption that roles are ascribed. In other words, the role corresponds to a position to which a person is associated with and who is expected to show the corresponding behavior. The second limitation is motivated by the observation according to which a person's role should be the result of an interaction process among all the people in the group, not just associated with his/her position. These limitations were expressed by Ellis and Fisher in [12] where they claimed that the role has to be defined "in terms of the communicative behaviors engaged in by the member occupying that role. The definition of a role solely as some preordained position that exists apart from the identity of the person occupying the position is incomplete" (p. 115).

The opposition encountered by the previous two views, paved the way for a third one. Biddle [7] viewed the role as "behaviors characteristic of one or more persons in a context" (p. 393). In other words, roles are directly related with individuals, they occur in a given context, and they are limited by contextual specifications. These observations lead to the conclusion that roles consists of "modal characteristics" or "modal behaviors": they define a role category for a particular person or persons [35]. Modal behaviors may change in response to the temporal and spatial context changes. According to Giddens [17], group members may be said to be positioned interactionally relative to one another in time-space. Because contexts are, generally speaking, interactionally created and negotiated by group members, they have an influence on behavior, the same way previous interaction and personal experience have an impact on present interaction. This idea emphasizes the role of interaction in the positioning of individuals and the dynamic nature of roles. In conclusion, we could claim that roles often arise as an emergent property from the interactional process.

2.2 Computational perspective

Although psychological research on role analysis dates back to the late 40s [6], the computational approach for social interaction has only recently started to address this problem. Pentland was one of the first proponents [30]. The computational approach tries to quantify nonverbal behavior including voice prosody, facial, hand and body gestures. Several studies proved that there is a strong correlation between all these elements in communication [19, 40]. According to [40], this is known as the 'excitatory hypothesis', namely, that a single excitatory impulse gives triggers unconscious reactions in the modality of speech prosody, face, hand and body gestures. A comprehensive overview about different group activities that have been automatically analyzed during social interaction can be found in [14, 15].

The automatic analysis of group interactions has mainly focused on informal meetings [5, 27, 29, 38]. In some cases [27], meetings follow a scenario and so people behave in a somewhat controlled manner. In other cases, meetings are task-oriented [34] or driven by a topic of discussion [29], but the implicit degree of antagonism or controversy is not very high, thus resulting in essentially non-competitive conversations. Political debates [20] are examples of competitive discussions, under fairly controlled conditions.

Roles can be seen as an information channel, providing valuable information about the meeting structure. For example, summarization and retrieval systems can

use roles as a new cue to segment and index audio-visual conversational databases [16]. As a consequence, actual search engines can be enhanced with content based role recognition capabilities.

Despite the potential it represents, the effort devoted to this problem has been limited. In [3], Banerjee et al. proposed a taxonomy based on a decision tree (created using simple speech-based features) to recognize the states of a meeting and the role of each participant. The meeting states are categorized as 'discussion' (debate) and 'information flow' (presentation). The roles of the participants are defined (depending on the meeting type) as: 'discussion participants', 'presenter', 'information provider' and 'information consumers'. The same idea of characterizing the role of different participants in a meeting (following a pre-defined task), has been followed in [44] and [10]. These papers share the same objectives and strategies. Their approach is based on the 'Functional Role Coding Scheme', which consists of ten labels that identify the behavior of each participant in two complementary areas: the 'Task Area' and the 'Socio Emotional Area'. The first area refers to functional roles related to facilitation and coordination tasks, meanwhile the second area is concerned with the relationship between group members. The observations are generated based on speaking/non-speaking segments, body and hand movement, and number of overlapping speakers. Only the methodology is different: a Support Vector Machine (SVM) is used in the former, and the Influence Model (IM) [4] in the latter. One of the claimed advantages of the IM resides in its generalization ability, i.e. this methodology could be easily adapted for groups with a variable number of participants. In [13] and [41], speaker role recognition is performed through a simple statistical Social Network Analysis (SNA) model. For their study, audio-only features (speaking length segments) were used which were extracted through unsupervised speaker diarization. The approach was tested on two kinds of data: radio broadcasts and meeting data (AMI corpus [26]). Several labels were assigned for the different roles they tried to recognize. For the first radio program dataset, these are: the 'anchorman', the 'guest', the 'interview participants', the 'abstract', and the 'meteo'. For the second meeting dataset, the labels corresponding to the roles are: 'project manager', the 'marketing expert', the 'user interface expert', and the 'industrial designer'.

The above mentioned examples could be considered as multi-class role analysis. Opposite to them, there is another group of computational studies whose aim is to perform a single-class role analysis: identifying the person with the highest-status role. There is a whole social theory regarding the relationship between status and dominance [11]. Perhaps the most complete study on dominance and status has been reported in [23] and [24]. Jayagopi et al. present a systematic study on automatic dominance modelling in small group meetings from nonverbal activity cues [24]. The analysis is based on a number of audio and visual activity cues for the characterization of behavior, including audio-only, visual-only and audio-visual cases to understand the relative power of each of the modalities and the benefits of using them jointly. The experiments were carried out on the AMI meeting corpus [26]. The work in [23] examined the problem of predicting both dominance and role-based status. The participants were assigned the same roles as in [13]. The 'project manager' has the highest status. The results showed that somehow simply nonverbal cues can be relatively effective in identifying the 'project manager' role, without relying on the verbal information.

The relationship between status and other characteristics (like dominance, for instance) has been studied in-depth in psychology [9]. Status is seen as a quality which implies respect and privilege, but not necessarily the ability to control others. On the other hand, dominance represents the quality to exert power and influence [9]. Although they are different concepts, they are intertwined: dominant people usually occupy higher status in a group; the other way around, people with higher status tend to make use of their power over their subordinates. A number of nonverbal cues have been found to be correlated with both status and dominance and some of them have been used to predict dominant people [23] and [34].

In [33], we reported some preliminary results of nonverbal analysis applied to role recognition.

3 Analyzing roles in "The Apprentice"

3.1 Overview and data

The TV show has a season-based periodicity, with the first season being broadcasted in 2004, and the last one finished in spring 2009. Each season starts with a group of job candidates having different backgrounds, including real estate, political consulting, sales, management, and marketing aspiring to work for Donald Trump. People are placed in two teams, and each week (i.e. in each episode of the show) they are assigned a task to be performed and asked to select a project manager for the task. The decision of what team wins/loses is made based on the teams' performance with respect to the task assigned. The winning team receives a reward, while the losing team faces a "boardroom showdown" in order to determine which team member should be fired (eliminated from the show). Elimination proceeds in two stages. In the first one, all of the losing team's members are confronted. The project manager of the losing team is asked to select some of the team members who are believed to be most responsible for the loss. In the second stage, which takes place in the boardroom meeting, the rest of the team is dismissed, and the project manager and the selected members face a final confrontation in which at least one of the members is fired by Trump at the end of the meeting. In this meeting, on one side we have the 'candidates board' and on the other side we have the 'executive board'. The 'executive board' is formed by Trump together with other persons (usually two) which will help him make the decision of what member of the team gets fired. Figure 1 presents a sketch of this scenario and Fig. 2 shows some snapshots from a boardroom meeting.

The data collected for our study correspond to the 6th season of the show, which was broadcasted during 14 weeks between January 7th–April 22nd, 2007. The number of initial candidates is 18. The following assumptions have been taken based on the exceptions mentioned below:

- In episode 3 (third week), one of the candidates resigns. Although this is a voluntary act, we consider it as a firing;
- In episode 13, Trump made it clear from the beginning that there would be no winners or losers for the assigned task; for this reason, we removed it from our study;
- Episode 14, the final one, consists of two stages: the 'semi-final' and 'final'. In the 'semi-final', two persons are chosen (from a total of 4) to become the finalists; in

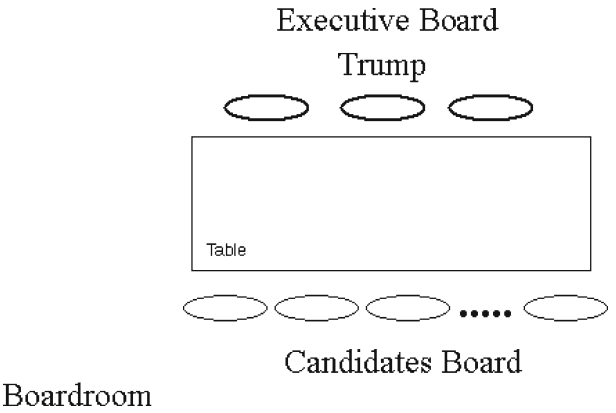

Fig. 1 "The Apprentice" boardroom meeting scenario

the final, the hired person will be declared. For this reason, we treat episode 14 as two separate meetings.

In conclusion, our data set is formed of 14 meetings. From each episode, our analysis has been focused towards, the second stage of the elimination process (i.e. the boardroom meeting). The meetings have an average duration of 6 mins and the

Fig. 2 Snapshots from the boardroom meeting; *upper left*: overview of a meeting; *upper right image* corresponds to the moment when Trump announces the fired person; *bottom row*: two hot-spot instances during the debate between candidates

number of participants varies between 5 and 11. Overall, we processed around 90 mins of audio data.

3.2 Prediction tasks

The particular character of this dataset offers the possibility to study and model roles in a competitive group interaction using nonverbal cues.

A quick overview of "The Apprentice" dataset provides a very interesting insight. On one hand, we could consider the overall meeting (all the participants) and on the other hand we could consider a subset of people taking into account only the participants from the 'candidates board'. This observation suggested two cases for our automatic analysis (see Fig. 3):

(i) *A 'supervised' case*, corresponding to the overall meeting (it is considered 'supervised' because we know that Trump is the person with the highest status). This is an example of a meeting where the hierarchical structure is very well established. As a consequence, we can find for this case two types of roles: first, the high status person, as an ascribed property as it was established by psychological research [35] and second, the fired person, as an emergent property from the interactional process. This results in two computation tasks:

- Task T1: predict the person with highest status (Trump);
- Task T2: predict the person who is fired.

(ii) *An 'unsupervised' case*, in which we eliminate Trump from the analysis. This is 'unsupervised' because without Trump, the remaining group of participants lacks of a pre-established high-status person (remains without a well-defined hierarchy). In this case, the new task we try to predict is the following:

- Task T3: 'is the most dominant person, the person who is going to be fired'?

3.3 Tasks prediction protocol

We adopt a common protocol for the prediction of T1, T2 and T3. For each of the tasks previously defined, we perform a rank-based classification, taking into account the first two persons with highest measure values. This convention was motivated by the assumption that the person with the highest status and the fired person are the ones who interact the most. These measure values are based on a series of descriptors which will be introduced and explained in more detail in the next subsection. The

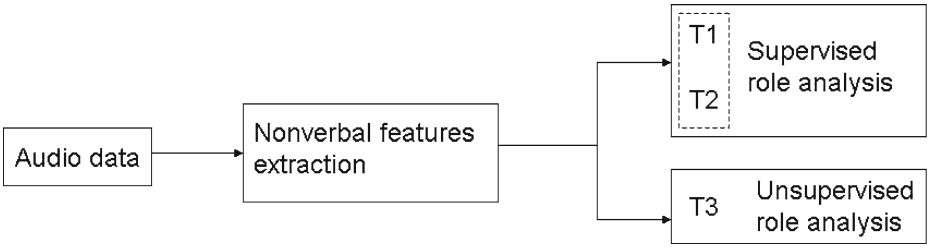

Fig. 3 A graphical representation of role analysis

general idea of rank-based classification is that it assigns a measure value (also called 'strength score' or 'confidence') to each data of a set, such that these values induce a specific ordering over the set. The rank-based classification provides information that is usually not conveyed by traditional classification methods: which data is more relevant than the other.

In case of the task T1, the following rule is established: "the person with the highest status is the first ranked person".

Regarding T2, a similar rule is applied: "consider the second-ranked position, except for those situations when this position is occupied by Trump, in which cases consider the first-ranked person". This rule arises from the obvious fact that Trump cannot be fired. Although more than one person can be fired in the boardroom meeting, in our study we consider to have made a good prediction, if we are able to make at least one positive identification.

Regarding T3, we have issued the following rule: "is the most dominant candidate (first ranks) the one who is going to be fired?". This approach is motivated by the psychologically-supported evidence stating that, in general, a person who feels his position threatened tends to display a more outgoing attitude, become more involved in the debate, trying to defend him/herself. In other words, he/she tries to become the most dominant one.

3.4 Feature extraction and representation

The data we had access to was the TV broadcast, so we had only one audio channel available. Due to the recording conditions (strong background music for the whole duration of each meeting), for our study we decided to manually produce the speaker segmentation in order to assure an optimal analysis of the data. Speaking segments are a binary vector (0—speaking, 1—non-speaking) indicating the status of a person (speaking/non-speaking). Based on the speaking segments, we define two types of data that were used as meeting-wise descriptors.

The first type is the class of individual nonverbal descriptors, that are person specific. For a more formal definition of these features, we will refer to Fig. 4, which depicts the speaking vectors for two hypothetical persons A and B. The speaking vector for person A could be decomposed in the form of n sequential, non-overlapping list of speech segments: $S_A = \{(s^i_{A_1}, \tau_{A_1}), (s^i_{A_2}, \tau_{A_2}), ..., (s^i_{A_n}, \tau_{A_n})\}$, where τ_{A_j} represents the length (in seconds) of the speech segment $s^i_{A_j}$, n represents the number of segments and $\sum_{j=1}^{n} \tau_{A_j} = T$ (the total duration of the meeting). In case when $i = 0$, we interpret it as a non-speaking (silence) segment; otherwise we consider it as a speaking segment.

- *TST—total number of speaking turns*: how many times a person takes the speaking turn during the meeting;

$$TST_A = \#\left(s^{i=1}_{A_j}\right), \text{ for } j = 1..n$$

  where $\#(Z)$ represents the cardinality of set Z.
- *TSI—total number of successful interruptions*: how many times a person successfully interrupts the others. 'A' interrupts successfully 'B' if 'A' was talking when 'B' started talking and 'A' stopped talking before 'B' does.

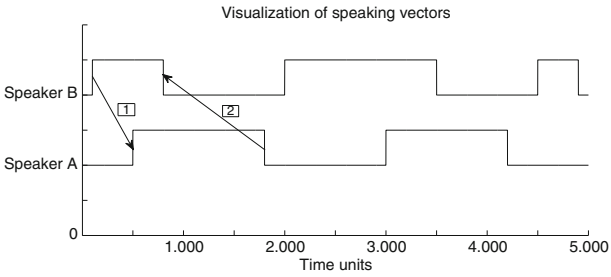

Fig. 4 Speech signals for two hypothetical persons A and B

We must first introduce the definition of an overlapping speech fragment as:

$$O_{jk} = s_{A_j}^{i=1} s_{B_k}^{i=1}$$

such that $O_{jk} \neq 0$. Additionally, the flow indicated by arrows must satisfy the temporal constraint as shown in Fig. 4. When all these conditions are fulfilled, then TSI can be expressed as follows:

$$TSI_{AB} = \#(O_{jk}), \text{ for } j = 1..n \text{ and } k = 1..m$$

– *TSL—total speaking length*: the total time a person speaks during the meeting.

$$TSL_A = \sum_{s_{A_j}^{i=1}} \tau_j, \text{ for } j = 1..n$$

In Fig. 5, we depict the histogram of these descriptors corresponding to the first episode. The number of persons who participated in this meeting is 5. These nonverbal descriptors have been used with relative success in [23]. They are known to be related to status from social psychology [36, 37].

The second type of cues is the class of relational descriptors, which characterize the interaction between people:

– *IM—interruption matrix*. It contains the information regarding 'who interrupts who' (column 'j' interrupts line 'i'). Its size is $N \times N$, N being the number of participants in the meeting;
– *TTM—turn taking matrix*. It contains the information regarding 'who is talking after who' (column 'j' talks after line 'i'). It can be also roughly interpreted as 'who answers to who'. Its size is the same as IM.

In the case of individual descriptors, each of them represents also a measure for personal characterization. Following the same approach, we define an equivalent

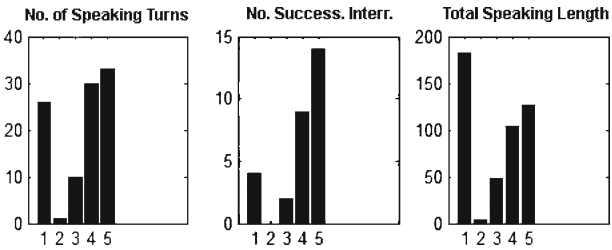

Fig. 5 Histogram of individual descriptors: TST, TSI and TSL, respectively in a meeting consisting of 5 participants. TSL is expressed in seconds

measure for the relational descriptors. In social network analysis, a common approach to assess a person's position in a group is centrality [42]. From graph theory's perspective, if we consider that the nodes correspond to people, then the links represents the relations between persons (who relates to whom). Each of these links carries a weight whose meaning is related to the type of relational descriptors we are referring to. In the case of IM, the weight represents the number of times a person successfully interrupts the other. In the case of TTM, the weight represents the number of times a person is talking after the other. We decided to use this measure in order to predict the 'fired' person because a person who feels his/her position is 'threatened' might try to become more involved in the discussion. In other words, he/she might manifest a high degree of engagement in the meeting, by trying to persuade the others. Intuitively, we also expect that the person with the highest status tends to occupy a central position in the group.

Centrality can be expressed in several ways. We chose for our study the following definitions [21]:

- *Degree centrality*: it is defined as the number of links incident upon a node (i.e., the number of links that a node has). If the network is directed (meaning that links have direction), then we usually define two separate measures of degree centrality, namely indegree (IC) and outdegree (OC). Indegree is a count of the number of links directed to the node, and outdegree is the number of links that the node directs to others;
- *Closeness centrality (CC)*: it is a centrality measure of a node within a graph. Nodes that have short geodesic distances to other nodes in the graph have higher closeness. In the context of group meetings, we can say that the smaller the distance between people corresponds to higher interaction between them.

4 Experimental results

In this section, we present the experimental results obtained for the two cases discussed in Subsection 3.2: supervised and unsupervised. Furthermore, we compare our results with those obtained with the Influence Model [5], which we will briefly recall below. For the sake of completeness, we concluded our study by considering also the users' experiences, i.e. what impressions people have when watching the dataset.

4.1 The Influence Model

The Influence Model (InfModel) is a tool developed to quantitatively analyze a group of interacting agents. In particular, it can be used to model human behavior in a conversational setting. In this context, the participants and their corresponding interactions are modelled through a coupled Hidden Markov Model (HMM). In Fig. 6, we offer a visual representation of this architecture.

The model is completely defined by a parametrization scheme that represent the influence of one chain over the others. More concrete, the multi-process transition probability $P(S_t^i | S_{t-1}^1, ..., S_{t-1}^N)$ is approximated only by the transition probability

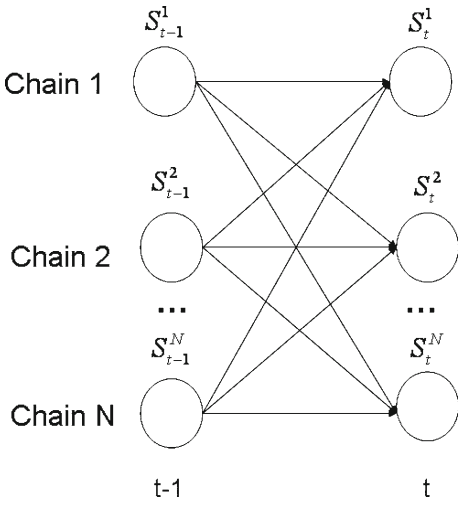

Fig. 6 The Influence Model architecture

$P(S^i_t|S^j_{t-1})$, where t represent the time stamp and N is the number of participants. With this convention, the multi-process transition could be expressed now as:

$$P\left(S^i_t|S^1_{t-1}, ..., S^N_{t-1}\right) = \sum_j \alpha_{ij} P\left(S^i_t|S^j_{t-1}\right) \quad (1)$$

In other words, the state of chain i at time t is conditioned only by the state of chain j at time $t-1$. The α-s parameters which appear in the equation above are referred as 'influences', because they are constant factors that tell us how much the state transitions of a given chain depend on a given neighbor. A more intuitive interpretation could be the following : the amount of influence from a neighbor is constant, but how this influence is reflected, depends on its state.

In its current implementation, the InfModel is able to model interactions between pairs of participants, but it is not able to model the joint effect of several chains together. The learning algorithm for the InfModel is based on constrained gradient descent. For our experiments, we estimated the InfModel based on speaking-non speaking features. The states represent the probability that a person is speaking or not at a given moment.

4.2 Supervised case

From the rank-based classification, and after applying our approach, the following tables present the prediction accuracy. In Table 1, we present the results based on the individual descriptors. We can appreciate that, in general, TSI and TSL are good cues for both T1 and T2. In change, TST is a good cue only for T2. We observe a large variation in the performance across measures, which suggests that some measures are more suitable than others to characterize the addressed tasks. It is interesting to notice that our approach outperforms InfModel in both tasks. The fact that TSI seems to be a better measure than TSL for predicting the person with the highest status is in line with our expectations, since nonverbal dynamics are higher

Table 1 Individual measures used to predict the highest-status person and the fired candidate

Tasks	TST (%)	TSI (%)	TSL (%)	InfModel (%)
T1	50.0 (7/14)	85.7 (12/14)	64.2 (9/14)	42.8 (6/14)
T2	92.8 (13/14)	85.7 (12/14)	78.5 (11/14)	64.2 (9/14)

in competitive meetings (note that the average percentage of overlapping speaking time is about 14.3%, which shows that interruptions seems to play an important role). This finding comes in contrast to previous research on non-competitive meetings [23], which found TSL to be the best measure to characterize high status.

In Table 2, we present the results for the relational descriptors IM and TTM, respectively. From these results, we could see that some of the centrality measures are considerably better in predicting T2 than T1. Between them, predicting T2 based on TTM is more reliable than on IM (for two centrality measures). From the results obtained so far, we could remark that centrality measures based on successful interruptions and turn taking descriptors seem to provide a good characterization of interaction dynamics. They contain implicit information that is useful for role analysis. The results we obtain seem to support the evidence according to which 'thin-slices' of behavioral data, based exclusively on nonverbal cues, are discriminant to predict the outcome of an interaction [1].

We performed an online analysis of predicting the person with the highest status based on the TSI measure, to see from which point during predictions were correct. The results of this analysis are depicted in Fig. 7. The horizontal axis corresponds to the percentage of accumulated meeting duration (this way we have a normalized representation of the data, irrespective to each meeting duration). We started our analysis at 30%. The vertical axis corresponds to the prediction accuracy (in percentage) at a given stage. When 100% of the data has been processed, the curve converges towards the result shown in Table 1. From this figure we appreciate that the person with the highest status can be identified in the early and late stages of the meeting (when the conclusions are made and the final decision is announced). The drop suffered by the curve (corresponding more or less to the middle of the meeting) might be explained by the fact that at some point during the debate, the person with the highest status 'passes on' the protagonism to the other participants and withdraws a bit. Starting with 60% of the meeting time, the curve recovers its ascending trend.

In Fig. 8 we depict a snapshot from our automatic analysis system. Video demos are available online at: http://www.cvc.uab.es/~bogdan/Videos/DemoBogdan_S06E01.wmv or http://www.cvc.uab.es/~bogdan/Videos/DemoBogdan_S06E05.wmv.

Table 2 Relational measures based on the successful interruption matrix (*IM*) and the turn taking matrix (*TTM*), respectively, used to predict the highest-status person and the fired candidate

Tasks	Meas. type	IC (%)	OC (%)	CC (%)
T1	IM	21.4 (3/14)	85.7 (12/14)	42.8 (6/14)
T2	IM	78.5 (11/14)	71.4 (10/14)	64.2 (9/14)
T1	TTM	57.1 (8/14)	64.2 (9/14)	42.8 (6/14)
T2	TTM	92.8 (13/14)	85.7 (12/14)	64.2 (9/14)

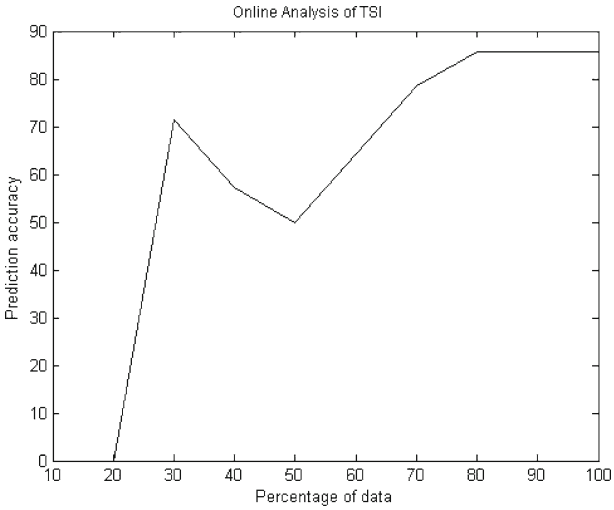

Fig. 7 Online analysis based on TSI for predicting the person with the highest status. See text for more details

4.3 Unsupervised case

As in the 'supervised' case, our prediction for the 'unsupervised' case is based on both the individual and relational measures. In Table 3, we present the prediction accuracy based on the individual descriptors. We can appreciate that, in general, TSI and TSL are good measures. The same as in the supervised case, our approach (using simple measures) outperforms InfModel.

In Table 4, we present the prediction accuracy based on the relational measures, which again show good performance.

4.4 Users' experiences

In order to complete the experimental part, we wanted to study also what user perception is when addressing the problem of behavioral pattern analysis in meetings. We want to stress from the very beginning that our study has not been conducted in a scientific manner (i.e. it hasn't been done or supervised by a psychologist familiarized with this kind of surveys), but it was intended as an ad-hoc, first impression study for informative purposes only. For this reason, a comparison between the results obtained from this study and the ones obtained using the automatic analysis is not pertinent. The subjects who took part in this experiment were not fluent in english (in order to avoid the effect content understanding might have upon their judgement). At the same time, they were not familiarized with The Apprentice series either. They were told to watch each episode up to a given point[1] which we considered it doesn't

[1]This has been decided by the authors. It was set up between 80% and 95% from the length of an episode, depending on the case.

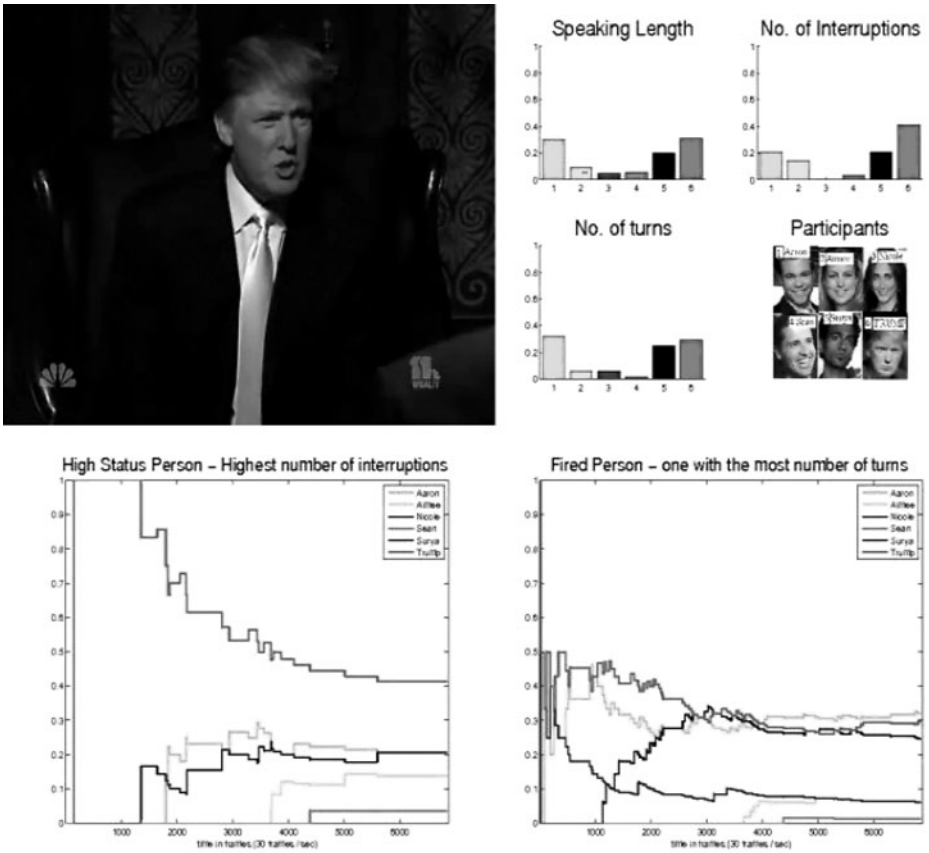

Fig. 8 Screen shot of our automatic prediction system we developed: *upper left* is a snapshot from an episode; *upper right* we depicted the histogram corresponding to the following measures: speaking length, number of interruptions and number of turns; *bottom left* and *bottom right* depict the evolution in time of our confidence measure for T1 and T2, respectively; the *curves* correspond to each of the meeting participants

offer a strong hint on the real outcome of the meeting. Afterwards, they were asked to specify which person, in their opinion, is going to be fired. They were asked not only to guess the to-be-fired candidate but also to justify their choice based on the people's behavioral patterns. We used in total 2 groups of people consisting of 4–5 persons each. Each person has been shown 2 or 3 episodes of the show, depending on its length. Surprisingly, the people in the first group had a very low correct guessing rate of only 27.2%, meanwhile, people in the second group had a higher correct guess rate: 54.5%. The low score in guessing the correct candidate to be fired

Table 3 Individual measures used to predict the most dominant person as the fired candidate

Task	TST (%)	TSI (%)	TSL (%)	InfModel (%)
T3	85.7 (12/14)	78.5 (11/14)	78.5 (11/14)	50.0 (7/14)

Table 4 Relational measures based on the successful interruption matrix (*IM*) and turn taking matrix (*TTM*), respectively, used to predict the fired candidate

Tasks	Meas. type	IC (%)	OC (%)	CC (%)
T3	IM	78.5 (11/14)	71.4 (10/14)	71.4 (10/14)
T3	TTM	85.7 (12/14)	78.5 (11/14)	71.4 (10/14)

could be explained through the lack of knowledge regarding candidates' personality. Psychological research revealed that persons behave in different manner under external stress factors, some of them showing a more reserved (introvert) attitude, meanwhile others are more outspoken, displaying a more affective and theatrical side. The subjects who watched the show fell in this pitfall. In some cases, the error occurred was due to the fact that they considered the candidate who was talking most of the time, capturing the attention of all the other participants couldn't be fired because he/she displayed a very self-confident attitude. Opposite to this, in other situations, a person who was too quiet, who seemed unable to defend him/herself was declared as the fired candidate because of him/her lack of involvement in the debate. These two cases are part of the same complex property called human personality.

As we mentioned before, this was a non-scientific survey. We identified some aspects which can be considered responsible for the low correct guessing rate and which were not taken into account from the beginning. This means that the understanding of social interaction, based solely on non-verbal behavior analysis, although relevant, could offer only a limited hint about the outcome. In conclusion, a more solid, psychologically-supported survey, conducted by persons with competence in this area is required in order to limit the pitfalls mentioned above. Another benefit for having a scientific-conducted survey would reside in the fact that will allow us to establish a bivocal correspondence between results obtained from user experiences and the results obtained via our automatic analysis. Last, but not least, such a survey could have a feed-back effect, in the sense that it might allow us to incorporate some intrinsec information in our automatic analysis, which could finally result in a more accurate prediction.

4.5 Limitations

We would like to make some general considerations regarding the challenges we confronted in our work. One of the main limitations is represented by the reduced size of the current data set, which limits the statistical validation of results.

Another limitation is represented by the recording conditions. Being a TV show, we had to adapt the analysis to the existing conditions. The only source of information that was consistent during the meeting and valid for analysis was the audio channel. It would have been very interesting to analyze also the visual channel. Unfortunately, the video data was not useful, since cameras were moving from one participant to the other and thus the information in the visual domain was affected by continuous interruptions. Extracting additional characteristics (like body motion or visual focus of attention) would have provided additional cues that cannot be studied for this data set.

5 Conclusions

In this paper we addressed the novel problem of role analysis in competitive meetings using nonverbal cues. Our study was based on the "The Apprentice" TV-reality show which offered an adequate data set. Our analysis was centered around two tasks regarding a person's role in a meeting: predicting the person with the highest status and predicting the fired candidates. For this purpose, we adopted a double strategy: supervised and unsupervised. The analysis, which uses only nonverbal features extracted from the audio modality was based on two types of nonverbal cues: individual and relational measures. Despite of the fact that our approach is a simple one, we showed that it serves as a good predictor for role analysis. Furthermore, it was compared with the Influence Model, a commonly used method for interaction modelling, demonstrating its superiority.

Although we performed our analysis on a small data set, the preliminary results obtained so far are promising. The methodology presented in this case-study would have to be validated in other types of competitive meetings to clarify whether the investigated features are good predictors of role-related behavioral outcomes'.

In the future, we are planning to extend the results presented here in several ways: firstly, to include elements of voice prosody in our analysis; secondly, to perform speaker diarization automatically; and finally, to create a larger corpus of data which will allow us to define a statistical framework for our analysis.

Acknowledgements This work was done while B. Raducanu visited IDIAP as an AMIDA project trainee. D. Gatica-Perez thanks the support of the AMIDA and IM2 projects. B. Raducanu is also supported by MEC Grants TIN2009-14404-C02-01 and CONSOLIDER-INGENIO CSD 2007-00018, Spain. We thank Dinesh Jayagopi (IDIAP) for providing the code for Influence Model.

References

1. Ambady N, Bernieri F, Richeson J (2000) Towards a histology of social behavior: judgmental accuracy from thin slices of behavior. In: Zanna P (ed) Advances in experimental social psychology, pp 201–272
2. Bales R (1951) Interaction process analysis: a method for the study of small groups. Addison–Wesley, New York
3. Banerjee S, Rudnicky A (2004) Using simple speech-based features to detect the state of a meeting and the roles of the meeting participants. In: Proc. of int'l. conf. on spoken language processing (ICSLP), p N/A. Jeju Island, Korea
4. Basu S, Choudhury T, Clarkson B, Pentland A (2001) Learning human interactions with the influence model. In: Tech report 539. MIT Media Lab
5. Basu S, Choudhury T, Clarkson B, Pentland A (2001) Towards measuring human interactions in conversational settings. In: Proc. IEEE int'l. conf. on computer vision, workshop on cues in communication (CVPR-CUES). Kauai, Hawaii, USA
6. Benne K, Sheats P (1948) Functional roles of group members. J Soc Issues 4(2):41–49
7. Biddle T (1979) Role theory: expectations, identities, and behaviors. Academic, New York
8. Bormann E (1990) Comunicating in small groups: theory and practice. Harper and Row, New York
9. Burgoon J, Dunbar N (2006) Nonverbal expressions of dominance and power in human relationships. In: Manusov Vea (ed) The Sage handbook of nonverbal communication. Sage, pp 279–297

10. Dong W, Lepri B, Capelletti A, Pentland A, Pianesi F, Zancanaro M (2007) Using the influence model to recognize functional roles in meetings. In: Proc. of int'l. conf. on multimodal interfaces (ICMI), pp 271–278. Nagoya, Japan
11. Dunbar N, Burgoon J (2005) Perceptions of power and interactional dominance in interpersonal relationships. J Soc Pers Relatsh 22(2):207–233
12. Ellis D, Fisher B (1994) Group decision-making: communication and the group process. McGraw-Hill, New York
13. Favre S, Salamin H, Dines J, Vinciarelli A (2008) Role recognition in multiparty recordings using social affiliation networks and discrete distributions. In: Proc. int. conf. on multimodal interfaces (ICMI), p N/A. Chania, Crete Island, Greece
14. Gatica-Perez D (2006) Analyzing group interactions in conversations: a review. In: Proc. of IEEE int'l. conf. on multisensor fusion and integration for intelligent systems. Heidelberg, Germany
15. Gatica-Perez D (2009) Automatic nonverbal analysis of social interaction in small groups: a review. Image Vis Comput (Special Issue on Human Spontaneous Behavior) 27(12):1775–1787
16. Gatica-Perez D, Zhang D, Bengio S (2005) Extracting information from multimedia meeting collections. In: Proc. of ACM int. conf. on multimedia, workshop on multimedia information retrieval (ACM MM MIR). Singapore
17. Giddens A (1984) The constitution of society: outline of the theory of structuration. University of California Press, Berkeley
18. Goffman E (1959) The presentation of self in everyday life. Doubleday, New York
19. Graf H, Cosatto E, Strom V, Huang F (2002) Visual prosody: facial movements accompanying speech. In: Fifth IEEE int'l. conf. on automatic face and gesture recognition. Washington, DC
20. Gregory Jr S, Gallagher T (2002) Spectral analysis of candidates' nonverbal vocal communication: predicting U.S. presidential election outcomes. Soc Psychol Q 65(3):298–308
21. Hanneman RA, Riddle M (2005) Introduction to social network methods. University of California (Riverside), Riverside, CA. Retrieved from http://faculty.ucr.edu/~hanneman/
22. Hare A (1976) Handbook of small group research. Free Press, New York
23. Jayagopi D, Ba S, Odobez JM, Gatica-Perez D (2008) Predicting two facets of social verticality in meetings from five-minute time slices and nonverbal cues. In: Proc. of int'l. conf. on multimodal interfaces (ICMI), p N/A. Chania, Greece
24. Jayagopi D, Hung H, Yeo C, Gatica-Perez D (2009) Modeling dominance in group conversations from nonverbal activity cues. IEEE Trans on Audio, Speech and Language Processing (Special Issue on Multimodal Processing for Speech-based Interactions) 17(3)
25. Katz D, Kahn R (1978) The social psychology of organization. Wiley, New York
26. McCowan I, Carletta J, Kraaij W, Ashby S, Bourban S, Flynn M, Guillemot M, Hain T, Kadlec J, Karaiskos V, Kronenthal M, Lathoud G, Lincoln M, Lisowska A, Post W, Reidsma D, Wellner P (2005) The ami meeting corpus. In: Proc. of the 5th int. conf. on methods and techniques in behavioral research, p N/A. Wageningen, The Netherlands
27. McCowan I, Gatica-Perez D, Bengio S, Lathoud G, Barnard M, Zhang D (2005) Automatic analysis of multimodal group actions in meetings. IEEE Trans Pattern Anal Mach Intell 27(3):305–317
28. McGrath J (1984) Groups: interaction and performance. Prentice Hall, New York
29. Otsuka K, Sawada H, Yamato J (2007) Automatic inference of cross-modal nonverbal interactions in multiparty conversations. In: Proc. ACM 9th int'l conf. on multimodal interfaces (ICMI), pp 255–262. Nagoya, Japan
30. Pentland A (2005) Socially aware computation and communication. Computer:63–70
31. Pentland A (2008) Honest signals. MIT Press, Cambridge, MA
32. Pentland A, Madan A (2005) Perception of social interest. In: Proc. IEEE intl. conf. on computer vision, workshop on modeling people and human interaction (ICCV-PHI). Beijing, China
33. Raducanu B, Vitrià J, Gatica-Perez D (2009) You are fired! nonverbal role analysis in competitive meetings. In: Proc. of int'l. conf. on audio, speech and signal processing (ICASSP), pp 1949–1952. Taipei, Taiwan
34. Rienks R, Zhang D, Gatica-Perez D, Post W (2006) Detection and application of influence rankings in small group meetings. In: Proc. ACM 8th int'l. conf. on multimodal interfaces (ICMI), pp 257–264. New York, US
35. Salazar A (1996) An analysis of the development and evolution of roles in the small group. Small Group Res 27(4):475–503
36. Schmid Mast M (2002) Dominance as expressed and inferred through speaking time: a meta-analysis. Human Commun Res 28(3):420–450

37. Smith-Lovin L, Brody C (1989) Interruptions in group discussions: the effects of gender and group composition. Am Sociol Rev 54(3):424–435
38. Stiefelhagen R, Chen S, Yang J (2005) Capturing interactions in meetings using omnidirectional cameras. International Journal of Distance Education Technologies 3(3):34–37
39. The Apprentice. http://www.nbc.com/The_Apprentice/
40. Valbonesi L, Ansari R, McNeill D, Quek F, Duncan S, McCullough KE, Bryll R (2002) Multimodal signal analysis of prosody and hand motion: temporal correlation of speech and gesture. In: EUSIPCO. Toulouse, France (2002)
41. Vinciarelli A (2007) Speakers role recognition in multiparty audio recordings using social network analysis and duration distribution modeling. IEEE Trans Multimedia 9(6):1215–1226
42. Wasserman S, Faust K (1994) Social network analysis. Cambridge University Press, Cambridge, UK
43. Weber M (2000) Basic concepts in sociology. Citadel, California
44. Zancanaro M, Lepri B, Pianesi F (2006) Automatic detection of group functional roles in face to face interactions. In: Proc. of int'l. conf. on multimodal interfaces (ICMI), p N/A. Banff, Canada

Bogdan Raducanu received the B.Sc. in Computer Science from the Politechnical University of Bucharest (PUB), Romania, in 1995 and a Ph.D. Cum Laude from the University of The Basque Country (UPV/EHU), Spain, in 2001. Currently, he is a senior researcher at the Computer Vision Center, Barcelona, Spain. His research interests include computer vision, pattern recognition, artificial intelligence and social robotics.

Daniel Gatica-Perez received the BS degree in Electronic Engineering from the University of Puebla, Mexico, in 1993, the MS degree in Electrical Engineering from the National University of Mexico in 1996, and the PhD degree in Electrical Engineering from the University of Washington, Seattle, in 2001. He joined the IDIAP Research Institute in 2002, where he is currently a senior researcher. He is currently an associate editor of the IEEE Transactions on Multimedia. His research interests include multimedia signal processing and information retrieval, computer vision, and statistical machine learning applied to these domains. He is a member of the IEEE.