Cogn Process (2012) 13 (Suppl 2):S533–S540
DOI 10.1007/s10339-011-0417-9

RESEARCH REPORT

Conversation analysis at work: detection of conflict in competitive discussions through semi-automatic turn-organization analysis

Anna Pesarin · Marco Cristani · Vittorio Murino · Alessandro Vinciarelli

Received: 5 March 2011 / Accepted: 19 September 2011 / Published online: 19 October 2011

Abstract This study proposes a semi-automatic approach aimed at detecting conflict in conversations. The approach is based on statistical techniques capable of identifying turn-organization regularities associated with conflict. The only manual step of the process is the segmentation of the conversations into turns (time intervals during which only one person talks) and overlapping speech segments (time intervals during which several persons talk at the same time). The rest of the process takes place automatically and the results show that conflictual exchanges can be detected with Precision and Recall around 70% (the experiments have been performed over 6 h of political debates). The approach brings two main benefits: the first is the possibility of analyzing potentially large amounts of conversational data with a limited effort, the second is that the model parameters provide indications on what turn-regularities are most likely to account for the presence of conflict.

Keywords Social signal processing · Generative score space · Conflict detection turn-organization

This article is part of the Supplement Issue on "Social Signals. From Theory to Applications," guest-edited by Isabella Poggi, Francesca D'Errico, and Alessandro Vinciarelli.

A. Pesarin · M. Cristani · V. Murino
University of Verona, Verona, Italy
e-mail: anna.pesarin@univr.it

M. Cristani
e-mail: marco.cristani@univr.it

V. Murino
e-mail: vittorio.murino@univr.it

M. Cristani · V. Murino
Italian Institute of Technology, Genoa, Italy

A. Vinciarelli (✉)
University of Glasgow, Glasgow, UK
e-mail: Alessandro.Vinciarelli@glasgow.ac.uk

A. Vinciarelli
Idiap Research Institute, Martigny, Switzerland

Introduction

Conflict is a mode of interaction that takes place whenever interacting agents do not share a common goal, but pursue individual goals, possibly incompatible with one another [2]. The main subjects of conflict are typically finite resources or attitude differences with respect to an issue of interest [19]. In both cases, conflicts might result into attempts of damaging or limiting the opportunities of others, with potentially disruptive effects on the life of any group where conflict takes place [21]. For this reason, this study proposes a semi-automatic approach for the detection of conflict in conversations.

Following the *Social Signal Processing* (SSP) framework [37], the proposed approach relies on extraction and analysis of non-verbal behavioral cues expected to carry conflict relevant information (e.g., overlapping speech and regularities in speaker sequences). The cues are represented in terms of *Steady Conversational Periods* (SCP), i.e., time intervals during which the conversation has a stable configuration (e.g., only one person talks, two persons talk at the same time, there is silence, etc.). The use of SCPs is motivated by findings in *Conversation Analysis* (CA) showing that turn-organization (who speaks when and how much) carries information about social aspects of a conversation [28], possibly including conflicts [6, 30].

The analysis approach is based on *Generative Score Spaces* (GSS), a recent technique combining both generative and discriminative models, the two main classification paradigms applied in machine learning [25]. Both generative and discriminative models are specified by probability distributions and are aimed at mapping objects of interest (in the case of this work, sequences of conversational turns) into classes belonging to a predefined set (*conflict* and *non-conflict* in this case).

Generative models are typically characterized by a set of intelligible parameters describing explicitly different classes of data. In the generative modeling, each class receives its own model, whose learning is class-independent, i.e., considers only those elements belonging to that class. The generative classification fits the test data on each available model and the one which realizes the best fit determines the class. The discriminative modeling learns how different classes can be separated, inferring a interclass separation hyperplane. The discriminative classification works by looking where a sample is located with respect to the separation hyperplane. Discriminative models have, on average, higher classification performance than generative approaches, but they fail in providing insights about the data. GSS approaches try to profit from the advantages of both paradigms and achieve performances superior to both (see Sect. 3 for more details).

Automatic analysis of conversations has addressed a large number of social phenomena, including the recognition of roles [34], the identification of dominant individuals [18], the analysis of personality [27], etc. (see Sect. 2 for a brief state-of-the-art). However, to the best of our knowledge, conflict has been so far largely neglected and the few available works focus on the detection of disagreement [8] or on the analysis of conflict structures (composition of the coalitions involved in a conflict) [35]. The application potential for automatic conflict detection extends beyond SSP and includes indexing and retrieval of data portraying social interactions (movies, meetings, etc.), supporting human–human communication in both co-located and computer-mediated scenarios, detecting unusual behaviors in surveillance applications, etc.

The experiments have been performed over a collection of 13 political debates for a total of 6 h and 27 min of material. Each SCP in the corpus has been manually labeled as conflictual or non-conflictual using a physically based annotation manual [4]. The accuracy, percentage of time correctly labeled in terms of conflict or lack of it, is around 80%. This corresponds to precision and recall of 71.7 and 74.0%, respectively, for conflictual SCPs. The corresponding figures are 86.8 and 85.5% when the SCPs are non-conflictual.

The rest of the study is organized as follows: Sect. 2 proposes a brief survey of related works, Sect. 3 presents the conflict detection approach, Sect. 4 reports on experiments and results, and Sect. 5 draws conclusions and outlines future perspectives.

State-of-the-art

While human sciences have investigated conflict extensively, no major works about the subject are available, to the best of our knowledge, in the computing community [37]. On the other hand, automatic analysis of social interactions attracts increasingly wider attention among machine intelligence researchers (see [38] for an extensive survey). Hence, many techniques have been developed for automatic understanding of non-verbal communication in social interactions. The rest of this section will provide first an introduction to the notion of conflict proposed by human sciences (with particular attention to the effect of conflict on non-verbal behavior) and then will provide a brief survey of the major works on automatic CA presented so far in the computing literature.

Conflict in human sciences

Conflict has been studied extensively in a wide spectrum of disciplines, including Sociology (e.g., see [24] for social conflict), Social Psychology (e.g., see [33] for intergroup conflicts), Psychology (e.g., see [15] for conflict in marriage), Political Sciences (e.g., see [23] for organizational conflict and [20] for international conflict), and Anthropology (e.g., see [9] for the role of language in conflict). A full survey of the subject is out of the scope of this study and the rest of this section will focus on those aspects of the problem that are most relevant to this work, namely interpersonal conflict in face-to-face interactions and its effect on non-verbal behavior.

When it comes to face-to-face communication, conflict is a "*mode of interaction*" [2] and it takes place over *resources* or *attitudes* [19]. In the former case, involved parties aim at maximizing their access to a given resource at the expense of the others. In the latter case, involved parties aim at imposing their attitudes, i.e., their beliefs and value orientations, toward an issue of interest. In both cases, conflict results from situations where, to a certain extent, the attainment of the goals of one party precludes the attainment of the goals of the other parties [1, 19].

Like every other phenomenon related to social interactions, conflict leaves traces in non-verbal communication, one of the main channels through which people form their impressions about others and social situations [40]. This applies in particular to turn-organization, i.e., the way people share and distribute the opportunity of speaking. During conflictual conversations, overlapping speech becomes both

longer and more frequent [10, 30], the consequence of a competition for holding the floor and preventing others from speaking. In the same vein, higher frequency is observed for interruptions [32] and speaker changes [10].

Other works have investigated the effect of conflict on non-verbal cues non-directly related to speech. In particular, the study in [17] shows that certain facial expressions (fear, disgust, and anger), arms akimbo and lack of symmetry in postures tend to be more frequent during a conflict. In parallel, the work in [31] shows that eye glances, direct gaze, self-touching and illustrating gestures are less frequent during a conflict.

Automatic conversation analysis

Automatic analysis of social phenomena in conversations has attracted attention only recently in the computing community. However, the related literature proposes a large number of approaches aimed at detecting social phenomena (see [37] for an extensive survey). Several works have proposed the use of non-verbal cues detectable through cameras (e.g., gestures, fidgeting, head movements, etc.), but turn-organization typically appears to be the most reliable source of information whether the goal is to recognize people personality (see, e.g., [27]), identify dominant individuals (see, e.g., [18]) predict the outcome of negotiations (see, e.g., [12]), or recognize the roles interaction participants play (see, e.g., [29]).

As turn-organization cannot be fully understood without taking into account its sequential aspects [6], the application of probabilistic sequential models is widespread. Generative models such as Markov models and extensions [3, 7, 41] are the main technique for exploiting vocal behavior cues for social signaling. Turn-taking dynamics may be effectively modeled as conditional dependencies among states of one or more stochastic processes [35]. The common idea is to sample a dialog at fixed time intervals, to learn a representative model, and to infer over the model parameters for detecting social aspects of that dialog. In [41], a two-layer hidden Markov model was employed to model individual and group actions (e.g., discussions, presentations, etc.). In [3, 5], the purpose was to detect the dominant interlocutor through social cues of mimicking. The authors employed an *Observed Influence Model* (OIM), i.e., an aggregate of first-order Markov processes, each one addressing an interlocutor.

More recently [11, 26], a generative framework has been proposed aimed at classifying conversation intervals of variable length (from a few minutes to hours), considering the nature of the people involved within (children, adults) and the main mood. The framework is basically an OIM, fed by low-level auditory non-verbal cues, dubbed SCPs (see Sect. 1). These are built on duration of continuous slots of silence or speech, and, in addition, they take into account conversational turn-taking. In practice, SCPs allow one to capture the attitude of self-selecting for turn-taking even though the interlocutor has not yet completed his own turn. Further, they also indirectly model speech planning by characterizing the tendency to utter short sentences instead of longer propositions.

Conflict detection process

Figure 1 depicts the overall conflict detection process. The first step is the extraction of the SCPs, i.e. the segmentation of the conversation into time intervals during which the configuration of the conversation is stable (Module 1). The second step is the generative model learning, i.e. the training of an ensemble of generative models (Markov chains in this work) over labeled sequences of SCPs (Module 2). The third is the representation of the Markov chains obtained at step 2 in an appropriate GSS (Module 3), and the fourth is the actual segmetation into conflict and non-conflict time intervals.

Module 1: SCP extraction

Initially conceived for dyadic conversations [11], the SCP extraction process has been extended here to an arbitrary number of speakers as follows: if there are M persons participating in a conversation, the original audio data is split into M synchronized sources (one per speaker) that are segmented into speech and silence intervals (the task has been performed manually in this work). As a result, the conversation is represented with M synchronized binary signals (see Module 1 in Fig. 1) or *single processes* ${}^{1}D,\ldots,{}^{c}D,\ldots,{}^{M}D$.

The SCP is a turn-organization-based feature built on the durations of continuous speech or silence intervals. The SCP extraction process assumes that there is a *global* transition of the conversation state whenever a single process changes its state. The segmentation caused by global state transitions creates $T \times M$ different SCPs ${}^{c}O_t$, where the apex $c \in 1,\ldots, M$ indexes the single process and $t = 1,\ldots, T$ enumerates the different time slots where SCPs have been defined. An SCP ${}^{c}O_t$ can be formally encoded as a pair $<{}^{c}I_t, L_t>$ where the first term tells whether the subject c is talking (= 1) or not (= 0), and the second term is the temporal duration of the $t-$th SCP in seconds (shared across the c processes). Summing L_t over t gives the duration of the conversation.

For an example of SCP extraction, see Module 1 in Fig. 1, where the black dots represent speech samples, while the white ones account for the silence ones.

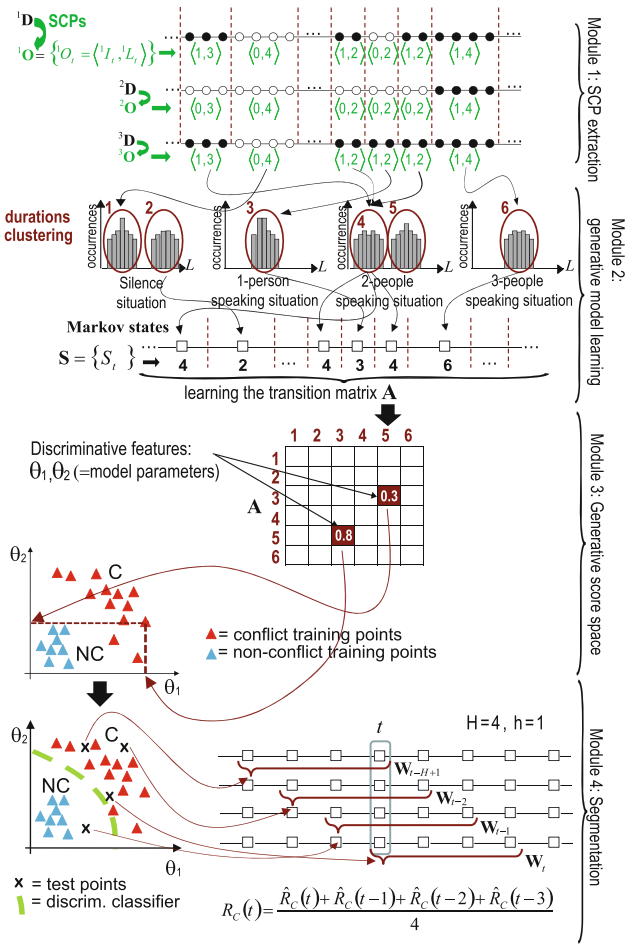

Fig. 1 The conflict detection approach for a three-person dialog

Module 2: generative model learning

The aim of Module 2 is to exploit the SCP sequence for capturing the dynamics of the conversation as if it was a unique joint Markov process.[1]

The *rationale* is to represent at time index t with a certain state S_t all of the SCPs where the number of speakers is P, with $P \in [0, M]$, and the duration is in a given range. While identifying P simply requires to count the speakers for which ${}^{c}I = 1$ (see above), finding suitable duration ranges requires the application of a Gaussian clustering following the approach proposed in [11]. In this way, the states account not only for a particular configuration of the conversation (mainly corresponding to the number of people talking at the same time), but also for how stable these configurations actually are. For a given number P of people talking, the number of duration clusters must be decided arbitrarily.

Once the states have been identified, it is possible to train a Markov model and the parameters provide an intuitive description of conversation dynamics like, e.g., the tendency of listeners to interrupt someone that speaks for too long or the occurrence of pauses during discussions.

Module 3: generative score space

The idea underlying the GSS is to project the parameter set of an ensemble of generative models onto an Euclidean vectorial space. In this work, the generative models are Markov chains trained over conversation segments labeled as conflict (*C*) or non-conflict (*NC*). In the resulting space, standard data analysis tools can be applied. In our case, we want to highlight the discriminative power of some subsets of features in a classification context, and therefore, we apply a feature selection (or ranking) strategy.

The rationale under the choice of this score space is that, by using parameters as discriminative features, we can understand what portions of a model differ from the other models at hand. For example, capturing the fact that a particular state transition is strongly discriminant for a certain class, means that such transition is peculiar for that model. This property cannot be mimicked by Fisher score based approaches, where the basic tool is the differentiation with respect to particular quantities (i.e., the log-likelihood in the Fisher score), that can suffer of the so-called "wrap-around" problem, where very different data points may map to the same derivative (see [25] for an example).

Module 4: segmentation

This step assigns each conversation interval addressed by S_t to the *C* or *NC* classes. Test sequences $\{S_t\}$ are first split into overlapping subsequences W_t (where t indicates the first state S_t) of length H, starting at steps of h time slots. A Markov model is trained over each W_t and the parameters that have been shown to be more discriminant at Module 3 are used as features.

The feature vectors are then classified (see next section for more details about the classifiers) as conflictual or non-conflictual so that each subsequence is assigned two scores $\hat{R}_C$ and $\hat{R}_{NC}$ accounting for how well they fit each of the two classes, respectively. As the subsequences are overlapping, each state S_t is included in several subsequences and it is thus assigned several scores. This allows one to assign each

[1] A first-order ergodic Markov model captures stochastic processes described by a set of N states $S_t \in \{1, \ldots, N\}$, that occur following a transition probability $P(S_t|S_{t-1})$. It is formally defined as a couple $\lambda = <A, \pi>$. A is the $N \times N$ time-invariant transition probability matrix. The initial state probability distribution $\pi = \{\pi_i\}$ represents the probability of the first state $\pi_i = P(S_1 = i)$. The parameters of a Markov chain can be easily estimated by frequency counts directly from training state sequences.

conversation interval corresponding at state S_t at time t two scores ($R_C(t)$ and $R_{NC}(t)$) corresponding to the average of the $\hat{R}_C(\cdot)$ and $\hat{R}_{NC}(\cdot)$ scores assigned to each of the subsequences they belong to (In Fig. 1 an example for R_C).

The experiments of this work are based on a general kernel-based classifier implemented in the *kernelc* function of the software package PRTOOLS [14]. The output of the classifier is normalized so that the R_C (and R_{NC}) value is bound between 0 and 1. Thus, it is possible to reject those S_ts for which the scores are lower than a predefined threshold γ that can be interpreted as a level of confidence.

Experiments and results

The experiments have two main goals: the first is to identify non-verbal cues most likely to account for conflict in the data at hand, the second is to measure the conflict detection accuracy of the proposed approach. The first goal is addressed during the model and feature selection step, when the approach configuration most likely to have high accuracy is found (see Sect. 4.2). The second goal is addressed during the conflict detection experiments, when the effectiveness of the approach in spotting conflict is measured over a dataset of political debates (see Sect. 4.3).

Hence, the experiments are beneficial under two main respects: the first is that it is possible to verify whether the approach is sensitive to the same cues as those proposed in the psychological literature (see Sect. 2). The second is that the approach can be used to analyze automatically large amounts of data and, if the accuracy is sufficiently high, it might become a tool useful to extend the observations of psychologists over datasets much larger than those used today.

The data

The experiments have been performed over a subset of the Canal9 corpus, a collection of television debates broadcast in Switzerland between 2005 and 2007 [36]. The selected subset includes the debates with three participants, one moderator and two guests defending opposite positions about an issue proposed at the beginning of the emission (e.g., *Are you favorable to the new tourism law?*). Overall, the selected subset includes 13 debates, for a total of 6 h and 27 min of material.

The debates have been manually segmented into *conflictual* and *non-conflictual* intervals using a physically based annotation manual [4]. An annotator has assigned the label C to two consecutive turns when at least one of the following phenomena is observed: there is overlapping speech, the speakers talk faster, or the loudness increases. In total, 2 h and 5 min have been labeled as *Conflict* (32.3%) and the remaining time as *Non-Conflict* (67.7%).

Since the approach relies on purely non-verbal behavioral cues, the annotators do not understand the language of the debates (French). This has two main advantages: the first is that both annotators and approach are sensitive to the same information (non-verbal cues in this case). Hence, it is possible to isolate the effect of non-verbal behavior on both perception and detection of conflict. The second is that the approach becomes language independent and the technical adaptation effort required to process data from different countries can be limited to cultural effects [38].

Model and feature selection

The first part of the experiments was designed to study the best model setting for the segmentation. This means to assess the model topology (i.e., the number of durations N_P for each number P of speakers talking at the same time) and the subset of features in the GSS. For this aim, we exploited the data labeling of all the conversations but one that we left out as test sequence. For each labeled conflictual (non-conflictual) segment, we learned a Markov model, initially using 2 durations (short, long) for $P = 0, 1, 2$. Thus, each model produced 36 parameters (due to the transition matrix[2]). We projected all the parameters in the GSS, and we applied *Forward Feature Selection* (FFS) based on the 1-nearest neighbor classification criterion. We set an observation window length $H = 90$ states S_t, with overlap $h = 85$ states, and confidence threshold $\gamma = 0.5$.

We iterated this process in a Leave-One-Out (LOO) sense, i.e., inserting the test sequence in the training dataset and exploiting another sequence of the pool as test. In this way, we captured in general the best features from a classification point of view. As a measure of accuracy, we evaluated both the percentage of states and the amount of time (in seconds) correctly classified. The latter accounts more effectively for the quality of the segmentation in terms of conflictual and non-conflictual exchanges.

We replicated the LOO validation changing the number of durations per type of state, ranging from 1 (no duration modeled) to 3 (three different types of duration: short, medium, and long), exploring all the possible configurations. We obtained the best result considering the following parametrization: 2 durations for $P = 0$, 1 duration for $P = 1$, and 2 for $P = 2$.

[2] Using only one sequence to train a model, the initial probability parameter array is meaningless.

With the best duration parametrization, the feature selection algorithm picked up two features, namely the probability of going from a *long* state of two-person speech to the state of a single person speech and the probability of the opposite transition. The feature selection strategy gives us important insights on the nature of the two conversation classes: actually, it emerges that the speech overlapping is discriminant and, in particular, it is discriminating to go/arrive in a state of *long* 2-person overlapping speech, in line with the findings of the psychological literature presented in Sect. 2. In Fig. 2, we show the values of such transition probabilities, for a particular run of the LOO validation. As one can see, the selected transitions have probability 0 for the non-conflictual dialogs. This indicates that the above transitions are characteristic (have probability >0) for the conflictual class.

Conflict detection results

The second part of the experiments was designed to measure the effectiveness of the approach in actually detecting conflicts in conversations. Model topology and features identified in the selection phase (see previous section) have been retained for these experiments.

The performance is measured in terms of *Precision* π and *Recall* ρ, where π corresponds to the percentage of time assigned a given class that actually belongs to that class, while ρ corresponds to percentage of time belonging to a given class that is assigned automatically to that same class. In both cases, the closer the performance measure to 100%, the better the performance.

In order to validate the classification results, the experiments have been performed using a Leave-One-Out approach: the classifier is trained over the whole dataset except one debate and then tested over the debate left out.

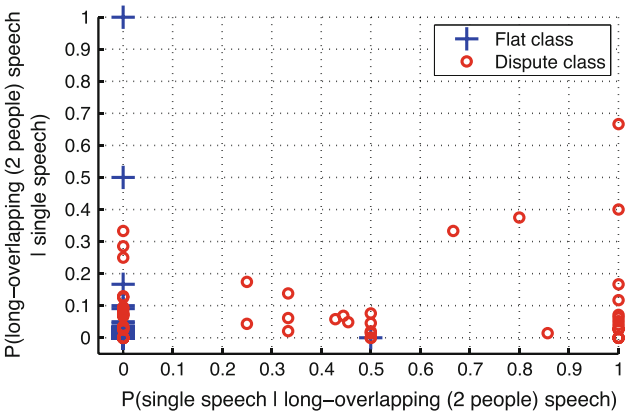

Fig. 2 The values of the selected P(long-overlapping (2 people) speech—single speech) and P(single speech—long-overlapping (2 people) speech) of all the training sequences of both the classes are shown as 2D points

This procedure is repeated and each time a different debate is left out for test. In this way, the whole database can be used as a test set while still keeping a rigorous separation between training and test material.

As for the discriminative classifier, we tested different options:

- *kernelc* [13]: a classifier based on a kernel or dissimilarity representation defined by Fisher approach;
- *knnc* [16]: a classifier based on k-nearest neighbor rule;
- *parzenc* [22]: a parzen classifier, using the best smoothing parameter of the kernel;

The best performances were reached by the *kernelc* classifier, and we report only these results for brevity.

Table 1 reports the π and ρ values for $\gamma = 0.5$ (no rejection mechanism) and $\gamma = 0.75$ (states S_t for which the classification score is lower or equal than γ are rejected). In this second case, the classifier does not make a decision for 16% of the data time. In accuracy terms (percentage of time correctly labeled in terms of Conflict and Non-Conflict), these values correspond to 78.1% ($\gamma = 0.5$) and 81.6% ($\gamma = 0.75$). The accuracy when assigning each conversation portion S_t to the class with highest prior is 67.7%, the difference with respect to the proposed approach is statistically significant. The results seem to suggest that a high γ does not change significantly the performance, thus the classifier tends to make decisions with high confidence even when they are not correct.

The results reported above are the result of an average over the whole dataset. Figure 3 shows the results for each debate separately. Each point corresponds to a pair (π, ρ) measured over a specific debate, the value of the coordinates is the average between the values of π and ρ measured for Conflict and Non-Conflict. In the plot are considered 9 out of 13 debates. This because in the remaining four debates (05-10-19, 05-11-02, 05-11-16, 06-01-11) the approach fails completely in segmenting the conflict part, producing a degenerate precision. Anyway, considering that in such cases the conflict segments hold altogether less than the 3% of the entire dialog length, this lack is negligible. In this respect, the approach seems to fail simply because conflicts are too short to be detected.

Table 1 The table reports precision and recall for different values of γ (0.5 and 0.75) and for the two classes separately (conflict and non-conflict)

	π (%)	ρ (%)
Conflict ($\gamma = 0.5$)	68.3	72.0
Non-conflict ($\gamma = 0.5$)	83.9	81.4
Conflict ($\gamma = 0.75$)	71.7	74.0
Non-conflict ($\gamma = 0.75$)	86.8	85.5

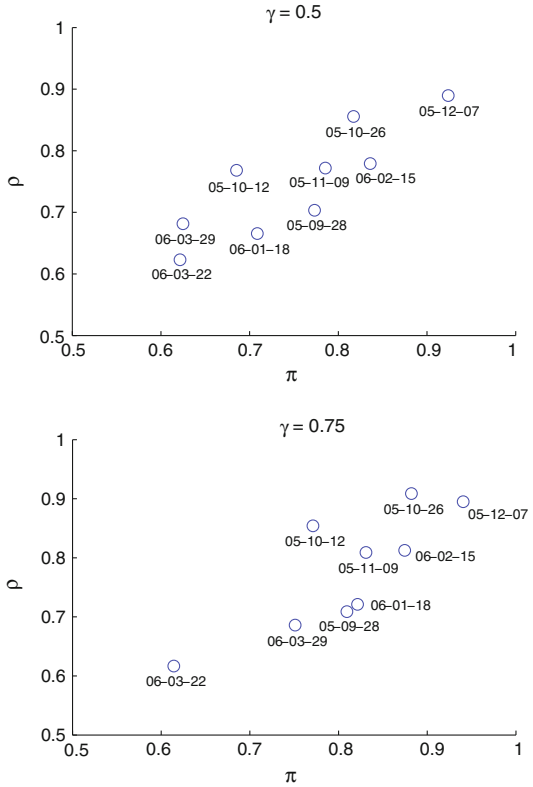

Fig. 3 Averaged precision and accuracy of the single debates, shown for two different confidence values. Under each point, it is reported the name of the debate file considered in the database

Conclusions

This study has presented an approach for the detection of conflicts in competitive discussions, i.e., conversations where people try to impose their point of view rather than to find a compromise [19]. The experiments have been performed over political debates as these are one of the most evident and accessible sources of competitive discussions. Furthermore, television data are known to provide ecologically valid samples of natural behavior [39]. The proposed approach is based on the analysis of turn-organization, one of the most salient aspects of conversations. Turn-organization, in simple terms, means *who talks when, to whom and how much*, i.e., how people share the speaking time at disposition, the dynamics of the exchange between speakers, etc. [28].

The experiments show that the approach detects conflicts with Precision and Recall around 70%. In other words, a conflict is detected roughly 7 times out of 10 when it actually takes place. On the other hand, the results have been obtained over manually extracted turn-organization (speaker segmentation is the result of a manual process) and it is not evident how much the use of an automatic process for the same task can affect the performance. However, conversations are often captured in settings where each participant is equipped with a lapel microphone and, in these conditions, the turn-organization can be extracted with virtually no error [29].

The results have been obtained over a collection of 13 debates and this limits the amount of material at disposition to train the classifiers used in the proposed approach. Thus, the extension of the corpus used in the experiments is one of the first directions that will be explored to obtain a more reliable assessment of the performance. Furthermore, the experiments will be performed no longer over manual speaker segmentations, but over the turn-organization as it is extracted automatically from raw conversation recordings. This requires the investigation of preprocessing steps capable of identifying interruptions and overlapping speech, of distinguishing between participants and moderators, etc. Last, but not least, the annotation is physically based, while it should be more appropriate to ask assessors when they actually perceive an interaction being conflictual [4].

To the best of our knowledge, this is one of the very first works to consider the conflict detection problem, in particular when it comes to situations where the conflict is not expressed intensely (e.g., by shouting), but through more subtle behavioral cues like in the case of moderated debates.

Acknowledgments This work is partially supported by the European Communitys Seventh Framework Programme (FP7/2007-2013), under grant agreement no. 231287 (SSPNet).

References

1. Allwood J (1992) The academic seminar as an arena of conflict and conflict resolution. Gothenbg Pap Theor Linguist 67:1–35
2. Allwood J (2007) Cooperation, competition, conflict and communication. Gothenbg Pap Theor Linguist 94:1–14
3. Asavathiratham C (2000) A tractable representation for the dynamics of networked Markov chain. PhD thesis, Department of ECS, MIT
4. Bakeman R, Gottman J (1986) Observing interaction. Cambridge University Press, New York
5. Basu S, Choudhury T, Clarkson B, Pentland A (2001) Learning human interaction with the influence model. Technical Report 539, MIT MediaLab
6. Bilmes J (1988) The concept of preference in conversation analysis. Lang Soc 17(2):161–181
7. Bishop C, Tipping M (1998) A hierarchical latent variable model for data visualization. IEEE Trans Pattern Anal Mach Intell 20(3):281–293
8. Bousmalis K, Mehu M, Pantic M (2009) Spotting agreement and disagreement: a survey of nonverbal audiovisual cues and tools. In: Proceedings of the international conference on affective computing and intelligent interaction, vol II. pp 121–129
9. Brenneis D (1998) Language and dispute. Annu Rev Anthropol 17:221–237
10. Cooper V (1986) Participant and observer attribution of affect in interpersonal conflict: an examination of noncontent verbal behavior. J Nonverbal Behav 10(2):134–144

11. Cristani M, Pesarin A, Drioli C, Tavano A, Perina A, Murino V (2011) Generative modeling and classification of dialogs by a low-level turn-taking feature. Pattern Recogn 44(8):1785–1800
12. Curhan J, Pentland A (2007) Thin slices of negotiation: predicting outcomes from conversational dynamics within the first five minutes. J Appl Psychol 92(3):802–811
13. Duda R, Hart P, Stork D (2001) Pattern classification. Wiley, New York
14. Duin R, Juszczak P, Paclek P, Pekalska E, De Ridder D, Tax D (2004) Prtools version 4.1: a matlab toolbox for pattern recognition. Internet http://www.prtools.org
15. Fincham F, Beach S (1999) Conflict in marriage: implications for working with couples. Annu Rev Psychol 50:47–77
16. Fukunaga K (1990) Introduction to statistical pattern recognition. Academic Press, New York
17. Gottman J, Markman H, Notarius C (1977) The topography of marital conflict: a sequential analysis of verbal and nonverbal behavior. J Marriage Fam 39(3):461–477
18. Jayagopi D, Hung H, Yeo C, Gatica-Perez D (2009) Modeling dominance in group conversations from non-verbal activity cues. IEEE Trans Audio Speech Lang Process 17(3):501–513
19. Judd C (1978) Cognitive effects of attitude conflict resolution. J Conflict Resolut 22(3):483–498
20. Kydd A (2010) Rationalist approaches to conflict prevention and resolution. Annu Rev Polit Sci 13:101–121
21. Levine J, Moreland R (1998) Small groups. In: Gilbert D, Lindzey G (eds) The handbook of social psychology, vol 2. Oxford University Press, Oxford, pp 415–469
22. Lissack T, Fu K (1976) Error estimation in pattern recognition via l-distance between posterior density functions. IEEE Trans Inf Theory 22(1):34–35
23. Morrill C, Rudes D (2010) Conflict resolution in organizations. Annu Rev Law Soc Sci 6:627–651
24. Oberschall A (1978) Theories of social conflict. Annu Rev Sociol 4:291–315
25. Perina A, Cristani M, Castellani U, Murino V, Jojic N (2009) Free energy score space. In: Advances in neural information processing systems. pp 1428–1436
26. Pesarin A, Calanca P, Murino V, Cristani M (2010) A generative score space for statistical dialog characterization in social signaling. In: Structural, syntactic, and statistical pattern recognition, Lecture Notes in Computer Science 6218. pp 630–639
27. Pianesi F, Mana N, Cappelletti A, Lepri B, Zancanaro M (2008) Multimodal recognition of personality traits in social interactions. In: Proceedings of international conference on multimodal interfaces. pp 53–60
28. Sacks H, Schegloff E, Jefferson G (1974) A simplest systematics for the organization of turn-taking for conversation. Language 50(4):696–735
29. Salamin H, Favre S, Vinciarelli A (2009) Automatic role recognition in multiparty recordings: using social affiliation networks for feature extraction. IEEE Trans Multimedia 11(7): 1373–1380
30. Schegloff E (2000) Overlapping talk and the organisation of turn-taking for conversation. Lang Soc 29(1):1–63
31. Sillars AL, Coletti SF, Parry D, Rogers MA (1982) Coding verbal conflict tactics: nonverbal and perceptual correlates of the "avoidance-distributive-integrative" distinction. Hum Commun Res 9(1):83–95
32. Smith-Lovin L, Brody C (1989) Interruptions in group discussions: the effects of gender and group composition. Am Sociol Rev 54(3):424–435
33. Tajfel H (1982) Social psychology of intergroup relations. Annu Rev Psychol 33:1–39
34. Vinciarelli A (2007) Speakers role recognition in multiparty audio recordings using social network analysis and duration distribution modeling. IEEE Trans Multimedia 9(9):1215–1226
35. Vinciarelli A (2009) Capturing order in social interactions. IEEE Signal Process Mag 26(5):133–137
36. Vinciarelli A, Dielmann A, Favre S, Salamin H (2009) Canal9: a database of political debates for analysis of social interactions. In: Proceedings of IEEE workshop on social signal processing. pp 17–20
37. Vinciarelli A, Pantic M, Bourlard H (2009) Social signal processing: survey of an emerging domain. Image Vis Comput 27(12):1743–1759
38. Vinciarelli A, Pantic M, Heylen D, Pelachaud C, Poggi I, D'Errico F, Schröder M (2011) Bridging the gap between social animal and unsocial machine: a survey of social signal processing. IEEE Trans Affect Comput (to appear)
39. Waxer P (1985) Video ethology: television as a data base for cross-cultural studies in nonverbal displays. J Nonverbal Behav 9(2):111–120
40. Wharton T (2009) Pragmatics and nonverbal communication. Cambridge University Press, Cambridge
41. Zhang D, Gatica-Perez D, Bengio S, McCowan I, Lathoud G (2006) Modeling individual and group actions in meetings with layered HMMs. IEEE Trans Multimedia 8(3):509–520