

COLLECTING DATA FOR SOCIALLY INTELLIGENT SURVEILLANCE AND MONITORING APPROACHES: THE CASE OF CONFLICT IN COMPETITIVE CONVERSATIONS

A.Vinciarelli^{1,2}, *S.Kim*², *F.Valente*² and *H.Salamin*^{1*}

¹University of Glasgow, Sir A.Williams Bldg., G12 8QQ, Glasgow (UK)

²Idiap Research Institute, CP592, 1920 Martigny (Switzerland)
{vincia,hsalamin}@dcs.gla.ac.uk,{fvalente,samuel.kim}@idiap.ch

ABSTRACT

One of the most recent trends in surveillance research is the application of socially aware approaches, i.e. approaches that integrate human sciences findings in order to better understand, model and predict the behaviour of people under observation. One of the key requirements for the development of such approaches is the collection of corpora that provide sufficient and reliable information about social phenomena of interest. However, the computing community still pays relatively little attention to the application of methodologies suitable for observational data collection, possibly inspired by human sciences experimental work. This paper tries to fill, at least partially, such a gap by providing an introduction to data collection techniques applied in nonverbal behaviour research. The collection of a corpus aimed at the study of conflict in conversations is used as a case study and example.

Index Terms— Social Signal Processing, Conflict, Non-verbal behaviour

1. INTRODUCTION

The ultimate goal of surveillance and monitoring technologies is to understand the behaviour of people in a number of spontaneous and naturalistic settings. Hence, it is not surprising to observe that the surveillance community shows increasingly more interest for domains like psychology and anthropology that have accumulated an extensive body of knowledge about human behaviour [1]. In particular, such a trend has become evident after that other technological fields like, e.g., Social Signal Processing and Affective Computing, have shown that the integration between technology and human sciences can be of major benefit [2].

Two lessons learned from human sciences are of particular interest in a surveillance and monitoring perspective: the

first is that human behaviour is not random, but follows regular patterns that are a source of both predictability and *a-priori* information [3]. The second is that social and affective phenomena leave physical traces under the form of nonverbal behavioural cues [4] (e.g., facial expressions, vocalizations, gestures, postures, etc.) and these can be detected and analyzed with sensors like microphones and cameras. In other words, social and affective phenomena are not immaterial, like typically believed, but belong to the range of physical, machine detectable phenomena accessible to any machine equipped with appropriate sensing devices [2].

The collection of large corpora of behavioural observations is one of the most important steps towards an effective cross-pollination between technology and human sciences. The reason is that no understanding of social and affective phenomena is possible without observation and quantitative analysis [5, 6]. Furthermore, the application of data driven approaches requires a large number of examples. However, even though the data plays such a crucial role, the literature still neglects to a large extent, if not at all, the application of correct methodologies for the collection of behavioural samples. Hence, the goal of this paper is to fill, at least partially, such a gap and to introduce some of the most important aspects of data collection when it comes to automatic understanding of human behaviour. The collection of a corpus aimed at automatic conflict detection will be used as a case study.

2. DATA COLLECTION METHODOLOGY

The domains traditionally interested in human behaviour (psychology, sociology, anthropology, ethnography, etc.) have developed rigorous methodologies for collection and annotation of behavioural observations, including compact notations for the description of behaviour, annotation protocols aimed at testing experimental hypotheses, formalization of observations, etc [4]. This section focuses on data collection techniques aimed at quantitative analysis of human behaviour, i.e. at the identification of measurable relationships between observable behaviour and its interpretation [5, 6].

*The research that has led to this work has been supported in part by the European Communitys Seventh Framework Programme (FP7/2007-2013), under grant agreement no. 231287 (SSPNet), in part by the Swiss National Science Foundation under the National Centre for Competence in Research IM2 (Interactive Multimodal Information Management), and in part by the Scottish Research Council via the Scottish Information and Computer Science Alliance (SICSA).

Such a perspective corresponds to the problems most commonly addressed in automatic understanding of human behaviour [2] (e.g., role recognition, automatic personality perception, etc.). Hence, it appears particularly suitable in surveillance and monitoring domains [1].

2.1. Quantitative Analysis of Human Behaviour

There are a few main questions that characterize every attempt to investigate human behaviour in quantitative terms [4]:

- Is there a relationship between a certain observable behaviour and a social/affective phenomena as perceived by human observers? (e.g., are people that smile perceived as happy?)
- Is there a relationship between multiple observable behaviours? (e.g., do people that smile more tend to fidget more as well?)
- Does observable behaviour take place in different ways? If yes, do differences have a meaning? (e.g., are there different ways of smiling? If yes, what do the differences mean?)
- Do social and affective phenomena influence observable behaviour? (e.g., does good quality of rapport makes smiling more likely?)

While being related to one another, the questions can result in different data collection approaches. In particular, questions addressing perception problems (i.e., how observers interpret behaviour) require the intervention of a number of human raters as large as possible (at least ten according to the thumb-rule commonly applied in practice). The goal is to smooth the effect of individual differences and non-collaborative raters. In contrast, questions related to self-perception or purely objective observations (no interpretation involved) can be addressed, at least in principle, with one rater only (in particular when it comes to self-perception). Furthermore, while self-perception questions require the subjects to work as raters, the other questions can be addressed by involving raters that are not subjects. The main advantage of the latter case is that it is possible to use widely available material such as television programs or web data.

The case study considered in this article (data collection for conflict detection) considers the first of the four questions listed above. In particular, the goal of the data collection is to identify conflict *markers* in nonverbal behaviour, i.e. the non-verbal behavioural cues most likely to account for the presence of conflict.

2.2. Observation

Ideally, observations should be conducted where and when phenomena of interest are likely to occur frequently. In some

cases, this is possible by simply observing real-world situations, in others it is necessary to create appropriate conditions. In this respect, behavioural observations are characterized by two main aspects, namely *setting* and *manipulations*. The first is the context where observations are made and it includes a continuum ranging from *naturalistic* (real-world situations where the experimenter cannot control anything) to *laboratory* (artificial contexts where the experimenters can control everything) settings. The manipulations are interventions aimed at eliciting desired effects and the resulting scenarios range between spontaneous (no manipulation at all) and controlled (behaviour is influenced by experimenter's interventions). As a result, observations can take four main forms [4]:

- *Laboratory experiments*: the setting is artificial and the scenario is manipulated (e.g., subjects are exposed to fixed stimuli and respond by picking one out of a few predefined alternatives)
- *Controlled observations*: the setting is artificial, but the scenario is spontaneous (e.g., subjects must perform a collaborative task, but do not receive indications on how to do it)
- *Field experiments*: the setting is naturalistic, but the scenario is manipulated (e.g., workers in a real company where different operative routines are tested)
- *Naturalistic observations*: the setting is naturalistic and the scenario is spontaneous (e.g., people participating in a demonstration)

In the case of conflict, television political debates appear to be particularly suitable for observation and data collection. The reason is that debates are typically built around the competition between two or more parties that try to impose their view on the issue of the day. Each party tries to acquire a share as high as possible of a finite resource (the consensus of the audience) at the expense of the others. In this respect, the interaction fits the definition of conflict proposed in [7, 8]: a mode of interaction where the attainment of the goals of one participant precludes the attainment of the goals of the other participants. Furthermore, television data has been shown to be a source of reliable and ecologically valid behavioural data [9].

Following the taxonomy proposed above, the use of political debates corresponds to a *naturalistic observation*. The debate is part of the real life of the participants and it takes place in a real television studio. The subjects tend to respect social norms, but they have real motivations and behave spontaneously to achieve their goals. The observers cannot do anything to manipulate the scenario.

#	Question	Layer
1	The atmosphere is relaxed	I
2	People wait for their turn before speaking	P
3	One or more people talk fast	P
4	One or more people fidget	P
5	People argue	I
6	One or more people raise their voice	P
7	One or more people shake their heads and nod	P
8	People show mutual respect	I
9	People interrupt one another	P
10	One or more people gesture with their hands	P
11	One or more people are aggressive	I
12	The ambience is tense	I
13	One or more people compete to talk	P
14	People are actively engaged	I
15	One or more people frown	P

Table 1. The table shows the questionnaire used to annotate the conflict database of the case study. The first column reports the question ID, the second column shows the question and the third column says whether the question belongs to the Inferential (I) or Physical (P) layer.

2.3. The Annotation

In most cases, the observation step results into recordings portraying social and affective phenomena of interest. However, the information necessary for quantitative analysis of behaviour is rarely available as such in the raw data and it must be extracted through a process, typically manual, called *annotation*. In other words, an appropriate number of *assessors* or *raters* (see Section 2.2 for what is meant by “appropriate”) must manually analyze the recordings and produce metadata that help to make sense of the information contained in the data.

Since it involves a degree of subjectivity, the annotation must be performed according to a protocol (the *annotation manual*) that typically takes one of the following forms:

- *Detailed description of the phenomena of interest:* in this case, the raters are expected to identify the points of the recordings where observable behaviour matches the descriptions of the manual
- *Multiple choice questionnaire:* the raters are expected to fill the questionnaire for specified segments of the recordings
- *Notation:* the raters are expected to describe what they observe in terms of the notation

In the conflict example, the annotation has been performed over a corpus of 1430 non-overlapping 30 seconds long clips (11 hours and 15 minutes). The clips have been extracted from the Canal9 Corpus, a collection of 45 political debates [10], and include all Canal9 segments where at least

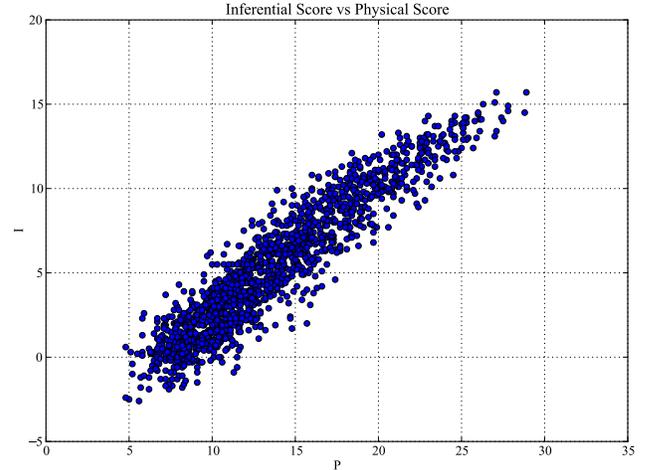


Fig. 1. Each point corresponds to a plot and the coordinates are physical (P) and inferential (I) scores.

two participants talk. Each clip has been annotated by 10 independent assessors that have filled the questionnaire of Table 1. The questions are split into two *layers*, the *physical* and the *inferential* one [5], but the assessors are not aware of the distinction. The physical layer includes questions about objectively measurable, possibly machine detectable non-verbal cues. The inferential layer includes questions about the way the assessors interpret their observations in terms of presence or absence of conflict.

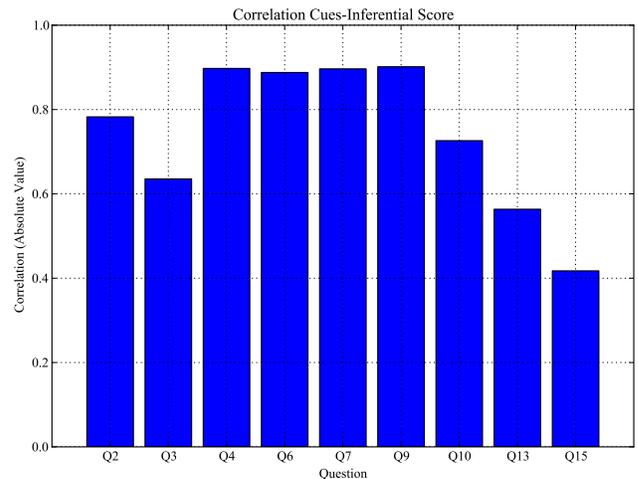


Fig. 2. Absolute value of the correlation between the score of individual physical questions (average over raters) and aggregate score of the inferential layer (average over raters).

Since the work aims at identifying nonverbal markers of conflict, the questions of the physical layer focus on how frequent are a number of nonverbal behavioural cues (interruptions, fidgeting, frown, etc.). In particular, the questionnaire

has focused on those cues that are most typically associated with competition and conflict such as the attempts of preventing others from speaking, the degree of motor activation and facial expressions conveying disappointment and aggressiveness [11].

The possible answers for the inferential layers are “*strongly disagree*”, “*disagree*”, “*nor agree neither disagree*”, “*agree*”, and “*strongly agree*”, those for the physical layer are “*never*”, “*once or twice*”, “*sometimes*”, “*often*”, “*always*”. In both cases, the scale is mapped into scores between -2 and 2 . For a given assessor, the scores of the two layers are aggregated as follows:

$$P = -Q_2 + Q_3 + Q_4 + Q_6 + Q_7 + Q_9 + Q_{10} + Q_{13} + Q_{15}$$

$$I = -Q_1 + Q_5 - Q_8 + Q_{11} + Q_{12} + Q_{14}$$

where the negative terms correspond to questions posed in negative form. Each clip is rated by 10 different assessors and the scores assigned to a clip are the average of the individual scores.

The goal of such a representation is to verify whether there is a correlation between how frequently certain behavioural cues are observed and how intense the conflict is perceived to be. Figure 1 shows a scatter plot where each point is a clip and the coordinates correspond to P and I . Since the correlation is 0.93 ($p < 10^{-16}$), the nonverbal behavioural cues mentioned in the physical layer appear to be, on average, good conflict markers. Figure 2 shows the correlation between scores of individual questions in the physical layer and the aggregated score of the inferential layer. All correlations are statistically significant ($p < 10^{-20}$), but some cues appear to be more effective than others as physical traces of conflict. Hence, any automatic approach should focus on these as a priority. From this point of view, the annotation provides good indications not only on the possibility of automatically detecting conflict, but also on the way an approach should be developed.

3. CONCLUSIONS

This article has presented some of the methodologies that human sciences apply to collect data for quantitative analysis of human behaviour. In particular, the article has shown that the investigation of quantitative relationships between observable behaviours can be articulated into few main questions (see Section 2.1), that behavioural observations take different forms depending on how tightly the experimenters control setting and scenario (see Section 2.2) and, finally, how to annotate behavioural observations to extract information necessary for quantitative analysis (see Section 2.3).

The collection of a corpus aimed at automatic conflict detection has been used as an example of how the points above can be addressed in a practical case. In particular, Section 2.3 has shown how the correct application of data collection

methodologies can show not only whether automatic conflict detection is possible, but also how an automatic approach should be developed. Since human behaviour is the main subject of surveillance and monitoring techniques, the application of appropriate data collection methodologies is likely to improve the current state-of-the-art in the domain.

4. REFERENCES

- [1] V. Murino, M. Cristani, and A. Vinciarelli, “Socially Intelligent Surveillance and Monitoring: Analysing Social Dimensions of Physical Space,” in *Proceedings of International Workshop on Socially Intelligent Surveillance and Monitoring*, 2010, pp. 51–58.
- [2] A. Vinciarelli, M. Pantic, D. Heylen, C. Pelachaud, I. Poggi, F. D’Errico, and M. Schroeder, “Bridging the Gap Between Social Animal and Unsocial Machine: A Survey of Social Signal Processing,” *IEEE Transactions on Affective Computing*, 2012.
- [3] H L Tischler, *Introduction to Sociology*, Harcourt Brace College Publishers, 1990.
- [4] M.L. Hecht and L.K. Guerrero, “Perspectives on Nonverbal Research Methods,” in *The Nonverbal Communication Reader*, K.L. Guerrero, J.A. De Vito, and M.L. Hecht, Eds., pp. 24–41. 1999.
- [5] R. Bakeman and J.M. Gottman, *Observing Interaction. An Introduction to Sequential Analysis*, Cambridge University Press, 1986.
- [6] P. Martin and P. Bateson, *Measuring Behaviour. An Introductory Guide*, Cambridge University Press, 2007.
- [7] J. Allwood, “The Academic Seminar as an Arena of Conflict and Conflict Resolution,” *Gothenburg Papers in Theoretical Linguistics*, vol. 67, pp. 1–35, 1992.
- [8] J. Allwood, “Cooperation, Competition, Conflict and Communication,” *Gothenburg Papers in Theoretical Linguistics*, vol. 94, pp. 1–14, 2007.
- [9] P.H. Waxer, “Video ethology: television as a data base for cross-cultural studies in nonverbal displays,” *Journal of Nonverbal Behavior*, vol. 9, no. 2, pp. 111–120, 1985.
- [10] A. Vinciarelli, A. Dielmann, S. Favre, and H. Salamin, “Canal9: A database of political debates for analysis of social interactions,” in *Proceedings of International Workshop on Social Signal Processing*, 2009, pp. 1–4.
- [11] V.W. Cooper, “Participant and observer attribution of affect in interpersonal conflict: an examination of non-content verbal behavior,” *Journal of Nonverbal Behavior*, vol. 10, no. 2, pp. 134–144, 1986.