
Systematic Evaluation of Social Behaviour Modelling with a Single Accelerometer

Hayley Hung

Technical University of Delft
The Netherlands
h.hung@tudelft.nl

Gwenn Englebienne

University of Amsterdam
The Netherlands
englebienne@gmail.com

Abstract

We describe our ongoing research on systematically analysing what types of socially related attributes and behaviours can be estimated automatically in highly social and crowded situations. This is a challenging task because obtaining the true labels for social behaviours or attributes in practice is non-trivial. Here, individuals hang a sensing device around their neck that records their acceleration during a social event. We then devise models to estimate their social behaviour or attributes based on these measurements and systematically evaluate the feasibility of such a set-up. Since we only use a single triaxial accelerometer per person, our results are surprisingly accurate and suggest that further socially relevant information could also be extracted. Our systematic evaluations provide a deeper understanding of how to better model socially relevant information in the future.

Author Keywords

Human behavior, social actions, wearable sensors, data mining

ACM Classification Keywords

H.1.2 [Models and Principles]: User/Machine Systems—*Human Information Processing*; H.3.1 [Information Storage and Retrieval]: Content Analysis and Indexing—*Indexing Methods*

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).
UbiComp'13 Adjunct, September 8–12, 2013, Zurich, Switzerland.
ACM 978-1-4503-2215-7/13/09.

<http://dx.doi.org/10.1145/2494091.2494130>



Figure 1: Mingling experiment.



Figure 2: Symposium event.

Introduction

We present ongoing work [3, 4] on understanding how accurately we can automatically estimate social behaviour or attributes in crowded social settings (e.g. Figure 1 or Figure 2) using a single triaxial accelerometer, worn as a sensor badge (see Figure 3). In this context, we consider dense crowded social gatherings where their sheer size and density lead audio, video, ultra-wideband, etc. sensor data to be too noisy for robust processing.

Our premise is that the link between body motion and social behaviour has been well-documented by social psychologists [5]: so there is sufficient evidence to further study how a single accelerometer can capture this information and how effective algorithms can estimate social behaviours from this impoverished sensor setting.

Recent advances in sensor technology and signal processing have made it possible to automatically extract features reflecting body language and to relate these to the person's state of mind, both in the lab [1, 6] and outside [7, 8]. However, up till now, these effects have been measured in relatively clean conditions and typically on a small scale [1, 9]. Large-scale experiments have been conducted [2, 8] but it relied on features, such as speech activity and proximity which still require relatively noiseless and uncrowded environments. Other related work which has estimated different aspects of social behaviour using wearable sensing devices have also relied heavily on relatively clean audio data [10, 8].

Yet interesting social behaviour also occurs in crowded and noisy situations, where identifying the speaker and robustly extracting prosodic features becomes extremely challenging. In other situations, speaking might be considered socially inappropriate (e.g. in the audience during a public presentation). Moreover, some wearers

may still consider audio recordings to be an invasion of privacy, regardless of whether privacy-preserving features are used. In comparison, a single accelerometer, which could for example be worn as a conference badge, does not suffer these flaws and is easy to wear and use. Restricting the number of available sensors in each device is also appealing in terms of low battery consumption. Practically speaking, the hardware set up we use could scale to hundreds if not thousands of users and this is the type of scenario we target.

Using this set-up, how could we still obtain valuable and useful information about the social content of an event? How can we know if the acceleration signals generated contain truly socially relevant information? Until this point, this has been considered very challenging and existing work has relied on assumed proxies for social behaviour, without systematically evaluating their performance as proxies [8, 2]. Without an understanding of the failure and success modes of such proxies, we cannot go further to improve upon the automated models of these behaviours. This extended abstract summarises a number of experiments and ongoing work to try to address this gap in our understanding so more detailed and accurate social analyses can be enabled in the future.

Data

The accelerometer data was collected using sensor badges (Figure 3(a)) hung around the volunteers' neck using a lanyard. Each badge contains a triaxial accelerometer (recording at 20Hz onto the 4 MB of flash storage), an LCD, and are synchronised by low-power radio.

We collected data from various social events. The first event was an (**Inaugural**) lecture where family, friends, and colleagues of the professor attended a speech (60 minutes) and drinks reception (90 minutes). About 75



Figure 3: The sensor badge we used should become the size of a credit card in the future.

Feature	Inaug.	Symp.
SigMean	19.49	61.82
SigVar	41.03	33.22
SigKurt	19.11	35.76
SigSkew	22.38	39.35
SigEnt	61.27	35.63
EnMean	20.65	59.15
EnVar	47.83	34.69
EnKurt	51.61	40.02
EnSkew	42.08	38.74
EnEnt	62.00	2.99

Table 1: Classification accuracy for professors and non-professors. Features were computed based on the raw acceleration signal (Sig) or its square or energy (En). Best performing features are highlighted in bold.

hours of accelerometer data from 32 people were collected. The second event was a (**Symposium**) for 300 attendees (see [Figure 2](#)), comprising two scientific presentation sessions (80 and 100 minutes) with a short break in between, followed by a drinks reception (100 minutes). About 156 hours of accelerometer readings from 46 people were collected. Finally, a third, lab-based experiment (**Mingling**), (see [Figure 1](#)), was organised with 32 volunteers who were encouraged to mingle and learn about each other so that they could form quiz teams later. 10 minutes of this event was labelled for 9 people.

Experiments

We conducted experiments to evaluate our system's ability to recognise various social attributes or behaviours:

Recognising affiliation

At the **Symposium** event, we gathered 11 affiliations from the participants according to their research group. The premise was that people's interest in a given talk should be correlated with their affiliation, which should be reflected in their body language. We used the standard k-nearest-neighbours, $k = 3$, to predict the affiliation of a person. The variance of non-overlapping 10-second windows were concatenated into a single vector for each person. Leave-one-out cross-validation resulted in a classification accuracy of 39.3%, far better than random.

Detecting Professors by how they move

We reasoned that people of varying status would have different motivations for socialising during professional social events. Therefore, professors would tend to have a larger social network and would move more between groups, while students, for example, would tend to circulate mostly within a smaller group of people.

The classification accuracies are summarised in [Table 1](#), and were generated from the statistics of the raw

magnitude (Sig) or energy (En) of the sensor readings over the drinks reception period of each event. We used a linear SVM classifier and leave-one-out cross validation with additional sub-sampling of the larger class. In the **Inaugural** event, the behaviour of 14 professors and 17 non-professors were available during the post talk drinks reception. For the **Symposium** event, data from 6 professors and 28 non-professors were recovered.

There is a notable contrast between the best two performing features. The entropy feature for the inaugural event captures both small and large motions, which represent both the gestural activities of standing people and their motion between groups. However for the mean (SigMean) feature, larger movement patterns will easily outweigh smaller movement patterns, putting more emphasis on movement between groups.

Classifying Social Actions

In the **Mingling** data, the actions of 9 subjects was labelled every 2s for : *speaking*, *laughing*, *gesturing* (either hand or head), *stepping* (or walking) and *drinking*. Using the power spectral density (PSD) for each axis of the acceleration, we trained two (multi-state) Hidden Markov Models (HMMs) for each action : one on data annotated with the 'positive' action and one on a random sample of sequences not associated with the 'negative' action. During inference, the most likely label is assigned to the observed sequence by comparing the likelihood of the data under the positive and negative HMM.

Model selection was done beforehand using different data, which was used to select the number of states per HMM, the window length over which each PSDs was calculated, and the number of frequency bins per window. The distribution parameters of the HMMs were optimised by maximum likelihood on left-out data from the same event

	gesture	step	drink
Precision	0.59	1.00	1.00
Recall	0.24	0.21	0.21
F1	0.34	0.35	0.35
	laugh	speak	
Precision	1.00	0.64	
Recall	0.38	0.82	
F1	0.56	0.72	

Table 2: Average precision, recall and F-measure for the different action categories in our dataset over 10 repetitions of 10-fold cross validation.

as the test data. We performed 10-fold cross-validation ten times on random permutations of the sequences and the classification performance is summarised in Table 2. *speaking* was most detected most robustly while *gesturing* was the most difficult to detect. The remaining behaviours were detected with a very high precision but low recall.

We would like to exploit these behaviours to detect who is speaking with whom. Social scientists have found that people talking together have certain distinctive synchronous behaviour [5]. Preliminary analysis of our data showed that gesturing and stepping occurs synchronously more frequently, and speaking simultaneously occurs less often for people in the same group compared to different groups.

Conclusion and Future Work

We have presented systematic evaluations of the estimation of social attributes and behaviour from accelerometer data recorded during crowded social gatherings. Future work will be focused on understanding how fusing information from the social attributes and actions could help to improve the estimation performance.

Acknowledgements

We thank researchers at the VU University of Amsterdam (Matthew Dobson, Claudio Martella, and Maarten van Steen) for the use of their wearable sensors and help during the data collection, and the University of Amsterdam (Jeroen Kools and Ben Kröse) for their data collection and annotation help.

References

- [1] Castellano, G., Kessous, L., and Caridakis, G. Emotion recognition through multiple modalities: face, body gesture, speech. *Affect and emotion in human-computer interaction* (2008).

- [2] Cattuto, C., Van den Broeck, W., Barrat, A., Colizza, V., Pinton, J., and Vespignani, A. Dynamics of Person-to-Person Interactions from Distributed RFID Sensor Networks. *PLOS ONE* 5, 7 (07 2010).
- [3] Englebienne, G., and Hung, H. Mining for motivation: using a single wearable accelerometer to detect people's interests. In *ACM MM Workshops*, ACM (2012).
- [4] Hung, H., Englebienne, G., and Kools, J. Classifying Social Actions with a Single Accelerometer. In *UbiComp* (2013).
- [5] Kendon, A. *Conducting Interaction: Patterns of Behavior in Focused Encounters*. Cambridge University Press, 1990.
- [6] Kim, T., McFee, E., Olguin, D. O., Waber, B., and Pentland, A. S. Sociometric badges: Using sensor technology to capture new forms of collaboration. *Journal of Organizational Behavior* (2012).
- [7] Laibowitz, M., and Paradiso, J. The UBER-Badge, a versatile platform at the juncture between wearable and social computing. *Advances in Pervasive Computing* (2004).
- [8] Olguin, D. O., Waber, B. N., Kim, T., Mohan, A., Ara, K., and Pentland, A. Sensible Organizations: Technology and Methodology for Automatically Measuring Organizational Behavior. *IEEE Transactions on Systems, Man, and Cybernetics, Part B* (2009).
- [9] Pantic, M., Pentland, A., Nijholt, A., and Huang, T. Human computing and machine understanding of human behavior: a survey. *Artificial Intelligence for Human Computing* (2007).
- [10] Wyatt, D., Choudhury, T., Bilmes, J., and Kitts, J. A. Inferring colocation and conversation networks from privacy-sensitive audio with implications for computational social science. *ACM Trans. Intell. Syst. Technol.* 2, 1 (Jan. 2011).