Mining for Motivation: Using a Single Wearable Accelerometer to Detect People's Interests

Gwenn Englebienne
University of Amsterdam
Science Park 904
1098XH Amsterdam
G.Englebienne@uva.nl

Hayley Hung
University of Amsterdam
Science Park 904
1098XH Amsterdam
hayleyhung@gmail.com

ABSTRACT

This paper presents a novel investigation of how motion as measured with just a single wearable accelerometer is informative of people's interests and motivation during crowded social events. We collected accelerometer readings on a large number of people (32 and 46 people in two crowded social events involving up to hundreds of people). In our experiments, we demonstrate how people's movements are informative of their particular interests: during talks, their interests in particular topics, and during networking events, their interest to participate successfully to make new contacts and foster existing ones. To our knowledge, using a single body worn accelerometer to measure and automatically infer these aspects of social behaviour has never been attempted before. Our experiments show that despite the challenge of the proposed task, useful automated predictions are possible and demonstrate the potential for further research in this area.

Categories and Subject Descriptors

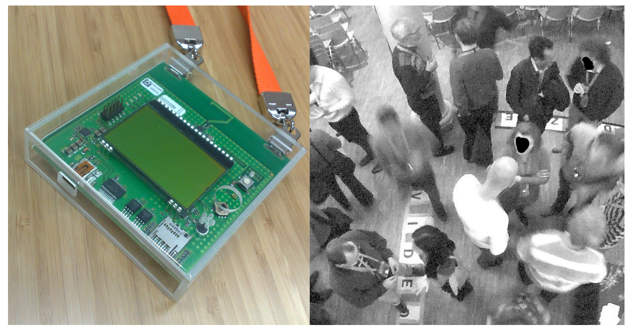
H.1.2 [**Models and Principles**]: User/Machine Systems—*Human Information Processing*; H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing—*Indexing Methods*; G.3 [**Probability and Statistics**]: Time Series Analysis

Keywords

Human behavior, human factors, wearable sensors, algorithms, data mining

1. BACKGROUND

The last several decades have seen a growing realisation that body language and gesturing reflect many different aspects of people's personality [1, 4, 13, 14]. More recently, 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 [3, 5, 6, 7] and outside [8, 12]. However, up till now, these effects have been measured in relatively clean conditions and typically on a small scale [3, 13]. Large-scale experiments have been conducted over extended periods of time. One example includes predicting the "Big Five" personality traits [12] but it relied on features, such as speech detection and proximity which still require relatively noiseless and uncrowded environments. In this paper, we present novel work on analysing how measuring body motion in crowded situations, with large groups of people, allows the discovery of people's interests and, as a consequence, related information such as their status or affiliation. In contrast to earlier work, this work focuses on crowded, noisy social situations such as that shown in Fig. 1(b). In such scenarios, extracted prosodic or proximity features are too noisy to be useful. Moreover, practical aspects such as battery life of the sensor node become important. Relying on few, simple and power-efficient sensors is therefore important so we focus exclusively on body motion as measured by a single wearable accelerometer, hung around the neck.



(a) (b)

Figure 1: (a) The sensor badge. Note that only the accelerometer was used on the sensor badge for our experiments. (b) Scene from the second scenario.

The novel contributions of this paper are the experiments that demonstrate the viability of using only accelerometers to measure aspects of social behaviour and affiliation. To our knowledge, this is the first study that has attempted to use a single body worn accelerometer, to measure and analyse aspects of instantaneous social behaviour in crowded social gatherings. The advantage of considering only a single accelerometer is that it is much less obtrusive than sensors such as video cameras or microphones. Secondly, we have collected and annotated accelerometer data from two large crowded social events; one event containing over 50 people, and another with around 300 (see Fig. 1(b)).

In the past, accelerometers have often been used for the

IMMPD'12, November 2, 2012, Nara, Japan.

recognition of simple "activities" such as household tasks [2], or to detect the mode of transport a person is using [9]. Other related work has concentrated on using proximity sensors to detect interactions and audio sensors to predict detailed information about the quality of face to face interactions [11]. Olguin et al. [10] have also combined sensor readings from microphones and accelerometers to monitor non-verbal cues related to conversation.

In this paper, we propose to analyse behaviour in crowded social events where there is much ambient noise from others talking, or music. Therefore, identifying the voice of the speaker and extracting prosodic features robustly becomes challenging. Moreover, some wearers may still consider audio recordings to be an invasion or privacy, regardless of whether privacy preserving features are used. We suggest that in crowded and noisy environments, exploiting other cues related to being involved in social events such as speaking, listening, or mingling, is beneficial.

Our work demonstrates the potential for far-reaching applications. It could allow the organisers of conferences or events to evaluate participants' interest, and marketeers to detect in how far their efforts are successful at reaching their target audience. It could let lecturers evaluate the quality of their lecturing and scientists to predict, while presenting their work, whether their conference paper has any chance of being cited (and by whom). In addition, this paper supports the use of cheap, disposeable wearable sensors. By using just a single XYZ accelerometer, we reduce the cost per unit, while still allowing the analysis of subtleties of social aspects of behaviour. Companies designing such devices expect the cost to fall easily below 1$ per unit with mass manufacture, making it an extremeley viable solution for analysing social behaviour in crowds.

2. DATA

The data for our experiments was collected during two separate events of real social gatherings, using wearable sensor badges (Fig. 1(a)) which were hung around the volunteers' necks using a lanyard. Accelerometer readings were stored locally at a recording frequency of 20Hz. The recording badges are similar in capabilities to the MIT Sociometric Badges [12], and are equipped with accelerometers along three axes, an LCD screen, and 4 MB of flash storage. The badges are additionally equipped with a microphone, but this was disabled during our experiments. The badges are not equipped with a Digital Signal Processor, so the audio-signal could not be processed on-line and no privacy-preserving features could be extracted. Purely accelerometer-based analysis has an important practical advantage in terms of modest hardware requirements.

The first of the social events, referred to as the "inaugural" in the remainder of this paper, occured during a professorial inaugural lecture event. This was attended by friends, colleagues, and family of the professor. The event consisted of an inaugural lecture (60 minutes), followed by a drinks reception for all attendees (90 minutes). We asked a number of volunteers from both groups to wear the nodes and collected accelerometer data during the talk. The talk was presented in such a way as to both challenge the views of the professor's colleagues, and to appeal to the more personal relationship of the speaker with his friends and family. Accelerometer data from 32 people were collected, for a total of 75 hours and 20 minutes.

The second social event, referred to as the "symposium" below (Fig. 1(b)), was on a much larger scale. It consisted of 300 people from different institutions, who gathered at an afternoon symposium on computer science research issues. The event consisted of two sessions of scientific presentations (80 and 100 minutes resp.), interrupted by a short break and followed by a drinks reception (100 minutes). For the purposes of our experiment, accelerometer readings from 46 people were collected for a total of 156 hours and 24 minutes.

3. EXPERIMENTS

We conducted three experiments to investigate our claims:

1. **Discriminate between the groups of "family and friends" and "colleagues"** at the inaugural event. In this setting, people from the different groups react at different times during the talk, based on their affinity with the subject at hand.

2. To confirm the results of the first experiment in a less obvious setting, we attempted to **discriminate between the different research group affiliations** based on people's reaction during talks at the symposium event. Different talks referred to different topics, and we hypothesise that people with different backgrounds react differently to those.

3. **Discriminate between professors and non-professors during networking**. In the above two experiments, the speaker is, in effect, a centralised source of events that synchronises the reactions of different groups of participants. However, our hypothesis is that, even in the absence of such a central source, motion behaviour is still informative of people's interests. To confirm this, we discriminate between professors and non-professors during drinks after the talks.

3.1 Feature extraction

The badges recorded acceleration along the three axes at a frequency of 20Hz. For our experiments, we computed the magnitude of the acceleration vector and discarded the orientation. We then computed statistics of the distribution of the magnitude during windows of varying length: the first, second, third and fourth central moments (mean, variance, skewness and kurtosis), as well as the minimum, maximum and entropy of the signal.

3.2 Distinguishing Work from Leisure

In this scenario, the speaker addresses an audience that consists of professors and non-professors. These are two clearly distinct groups, as the non-professors consist of family, friends and colleagues, who all know the speaker well personally. Both groups are specifically targeted during the talk, where jokes are alternated with insightful scientific comments. As a consequence, we can expect members of the different groups to react differently during these targeted moments, for example by laughter or apathy, respectively, when the speaker makes an inside joke, or by boredom or excitement, when an interesting scientific factoid is mentioned. We capture this effect by discretizing time and extracting features from each time window independently.

Visualisation As a first exploratory experiment, we used a non-linear dimensionality reduction technique (t-SNE, [15]) to visualise the data, using the variance of the acceleration

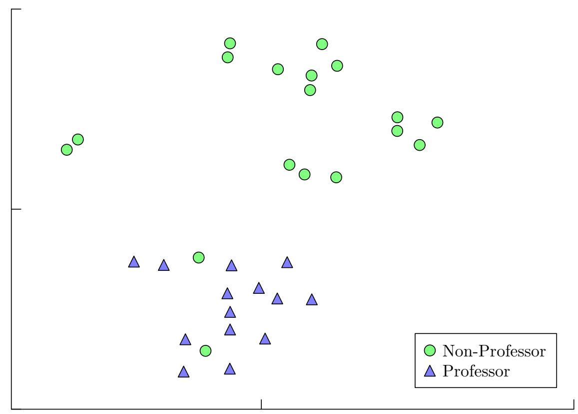


Figure 2: Visualisation of the motion data after non-linear dimensionality reduction. The labels were not used to obtain the low-dimensional representation, but were added afterwards for illustration.

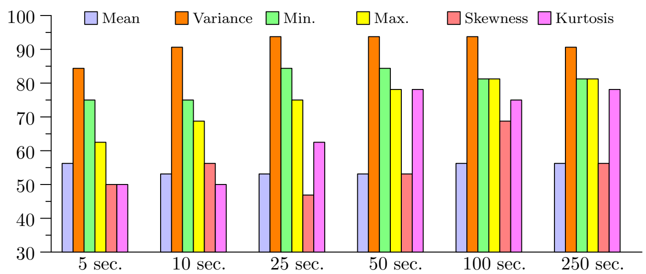


Figure 3: Classification accuracy using different features and different window lengths, for the inaugural

in 100-second time windows. The results are depicted in Fig. 2. Notice that the group labels were not used to obtain the 2D representation, they were added after dimensionality reduction to make the actual plot. We can observe that both groups are quite well differentiated, with few exceptions.

Classification This result is confirmed by classifying the people as professors or non-professors based on their motion. We used a Support Vector Machine (SVM) with Radial Basis Function kernel to separate the two classes. We used leave-one-out cross-validation on the same dataset, training on 31 datapoints and testing on one. Given the high dimensionality of the data and the limited size of the dataset, we used Principal Component Analysis (PCA) to first project the data to a lower-dimensional space (the three principal components were kept), capturing correlations between behaviours in different time windows. Fig. 3 shows the results for different window lengths, using the different features. We can see that the performance is best for intermediate window lengths, which seems to indicate that too short window lengths are sensitive to the different reaction speeds of different attendees, while long windows discard the timing information. Overall, we obtain an accuracy of 93.75% using variance, for window lengths ranging from 25 to 100 seconds.

In terms of features used, the variance of the acceleration is most informative for all window lengths we tested for, minimum and maximum acceleration being slightly worse. This is not so surprising, as they are basically more noisy representations of the same information: *viz.*, how much motion there was in the time slice. The skewness is not informative, but one would expect the kurtosis to complement the variance. However, more data is required for an accurate estimation of this feature, which may explain why it becomes more informative as the window length is increased.

3.3 Recognising affiliation

The affiliation of the participants to the symposium is more challenging to detect than the two groups that we investigated in the inaugural. People routinely move between research institutes, and although a person's affiliation is certainly informative of their interests, the reverse is not necessarily true. Two people doing very similar research can easily be in different groups, at different institutes. For privacy reasons, we could only ask people for very generic affiliation information, so that the results reported here should be seen as a lower bound on the classification accuracy of people's interests, rather than an accurate estimate.

By gathering information about each participant's research group, we obtained 11 different affiliations. Since SVMs are inherently a 2-class classification method, and extensions to multiclass are problematic, we used the standard k-nearest-neighbours, $k = 3$, to predict the affiliation of a person. Keeping the best feature found in experiment 1, 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.

3.4 Detecting Professors by how they move

We reasoned that people of varying status would have different motivations for socialising during professional social events. We asked people to report on their professional status to see if their behaviour would distinguish professors from non-professors. We expected that since professors would tend to have a larger social network, and know more people, they would be more likely to move between groups during a social occasion. On the other hand, students, for example, would tend to circulate mostly within a much smaller group of people, or indeed, stay with the same group for the entire event. We also reasoned that professors would tend to have more expansive gesturing and movement compared to non-professors and extracted appropriate features accordingly.

Classification and Cross Validation We used a linear SVM classifier. In the inaugural event, the behaviour of 14 professors and 17 non-professors were available during the post talk mingling. For the symposium event, data from 6 professors and 28 non-professors were recovered. Here, the non-professor category consisted of mostly PhD students, assistant, or associate professors. Since the class sizes were significantly imbalanced and also relatively small, we carried out leave-one-out cross validation with additional subsampling of the larger class.

Analysis of the Results The classification accuracy of this task are summarised in Table 1, which were generated from accumulating the statistics of the raw magnitude (Sig) or energy (En) of the sensor readings over the entire period of mingling at each event. The best performing feature had a classification accuracy of 62% for the inaugural event was the entropy of the energy (EnEnt), while for the symposium event, a classification accuracy of 61.82 % was achieved when using the mean of the signal magnitude (SigMean).

	Inaugural	Symposium
SigMean	19.49	**61.82**
SigVar	41.03	33.22
SigKurtosis	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: Recognition accuracy for professors and non-professors, based on the motion observed during the post-inaugural mingling. Sig: features computed on the magnitude of the raw acceleration signal. En: features computed on the square of the raw acceleration signal.

On inspecting the signals for professors compared to non-professors during the inaugural event, we observed that the signal tended to have low energy interspersed occasionally with high energy while for non-professors, the signal tended to be more continuous and noisy. This supports the result that the entropy of the energy signal performed best for this scenario as the sensor data of non-professors tended to be noisier and therefore have a higher entropy. For the symposium event, the mean signal value performed best probably because there would have been more movement between different groups as professors went about socialising.

The difference in features for the two events could be explained by the size and number of participants for each event. For the inaugural, around 50 people were mostly limited to a single room where people queued to congratulate the professor and mingled with others. This may explain why SigMean performs badly. However, in the symposium event, the space was much larger. The participants could move between two large rooms and the crowd density was extremely high in some places. Different conversing groups were standing extremely close together and jostling past each other. This may explain the extremely poor performance of the entropy of the energy signal (EnEnt) as a lot of movement would be due to people moving to let someone pass, while they were talking to someone else.

The contrast between the best two performing features in both events is quite interesting. The entropy feature for the inaugural event represents both small and large motions, which capture the gestural activities while people are standing, as well as how often they moved between groups. However for the mean (SigMean) feature, larger movement patterns will easily outweigh smaller movement patterns, putting more emphasis on movement between groups.

4. CONCLUSIONS

In this paper, we have shown how to extract complex and high-level information from body motion measured by wearable accelerometers. We have shown that in situations where a "global" source of events is synchronising multiple participants, different participants will exhibit motion behaviour that is indicative of different levels of interest about the various events. This is reflected in body gesturing, and makes it possible, to an extent, to identify the background and status of the different participants. Even when no such events exist, different people still have different motivations, resulting in different physical behaviours. Despite the more free nature of these behaviours, we have shown that, complex information such as the profession of a person can be identified with substantially better than random accuracy. In the future, we will investigate how much information can be gleaned from unstructured body gesturing with unsupervised techniques.

5. ACKNOWLEDGEMENTS

This research was supported in part by a Marie Curie Research Training Network fellowship in the project "AnaSID" (PIEF-GA-2009-255609) and by the Health-Lab project.

Thanks to Matthew Dobson for his help in preparing the sensor badges and providing code for parsing the acceleration data. Also thanks to Ben Kröse and Maarten van Steen for their support during the data collection.

References

[1] N. Ambady and R. Rosenthal. Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, 111(2):256–274, 1992.

[2] L. Bao and S. Intille. Activity recognition from user-annotated acceleration data. *Pervasive Computing*, pages 1–17, 2004.

[3] G. Castellano, L. Kessous, and G. Caridakis. Emotion recognition through multiple modalities: face, body gesture, speech. *Affect and emotion in human-computer interaction*, pages 92–103, 2008.

[4] B. De Gelder. Towards the neurobiology of emotional body language. *Nature Reviews Neuroscience*, 7(3):242–249, 2006.

[5] S. D'Mello and A. Graesser. Automatic detection of learner's affect from gross body language. *Applied Artificial Intelligence*, 23(2):123–150, Feb. 2009.

[6] W. Dong, B. Lepri, T. Kim, F. Pianesi, and A. P. Pentland. Modeling conversational dynamics and performance in a social dilemma task. In *Proceedings of ISCCSP 2012*, Rome, Italy, May 2012.

[7] T. Kim, E. McFee, D. O. Olguin, B. Waber, and A. â. Pentland. Sociometric badges: Using sensor technology to capture new forms of collaboration. *Journal of Organizational Behavior*, 33(3):412–427, Apr. 2012.

[8] M. Laibowitz and J. Paradiso. The UbER-Badge, a versatile platform at the juncture between wearable and social computing. *Advances in Pervasive Computing, Oesterreichische Computer Gesellschaft*, 363:368–373, 2004.

[9] A. Mannini and A. M. Sabatini. Machine learning methods for classifying human physical activity from on-body accelerometers. *Sensors*, 10(2):1154–1175, Feb. 2010.

[10] D. Olguin and A. Pentland. Assessing group performance from collective behavior. In *Proc. CSCW*, volume 10, 2010.

[11] D. Olguin, B. Waber, T. Kim, A. Mohan, K. Ara, and A. Pentland. Sensible organizations: Technology and methodology for automatically measuring organizational behavior. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(1):43–55, Feb. 2009.

[12] D. Olguin Olguin, P. A. Gloor, and A. Pentland. Wearable sensors for pervasive healthcare management. In *Proc. Pervasive Computing Tech. for Healthcare*, pages 1–4, 2009.

[13] M. Pantic, A. Pentland, A. Nijholt, and T. Huang. Human computing and machine understanding of human behavior: a survey. *Artifical Intelligence for Human Computing*, pages 47–71, 2007.

[14] J. H. Sterrett. The job interview: Body language and perceptions of potential effectiveness. *Journal of applied psychology*, 63(3):388, 1978.

[15] L. Van der Maaten and G. Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(2579-2605):85, 2008.