Assessing Group Performance from Collective Behavior

Daniel Olguín Olguín and Alex (Sandy) Pentland
MIT Media Laboratory
Human Dynamics Group
{dolguin, pentland}@media.mit.edu

Introduction

The Human Dynamics research group at the MIT Media Laboratory has demonstrated that wearable technology can be used to characterize face-to-face interactions, measure individual and collective patterns of human behavior, and automatically map out a company's de facto organizational chart (Choudhury & Pentland, 2003; Pentland, 2006; Olguin-Olguin et al., 2009a, 2009b). This capability can be an extraordinary resource for studying group behavior, group performance and team formation processes.

With that goal in mind we developed the *Sociometric* badges, wearable electronic sensors capable of detecting face-to-face interactions, conversations, body movement, and proximity to others (Olguin-Olguin, 2007). The *Sociometric* badges are capable of extracting speech features without recording the content of conversations in order to maintain privacy, and of wirelessly transferring data to a central server. We have used them in several organizations to capture face-to-face communication patterns and study the relationship between collective behavior and performance outcomes, such as productivity and job satisfaction (Olguin-Olguin et al., 2009a, 2009b; Wu et al., 2008).

The design of the *Sociometric* badges was motivated by the fact that a large number of organizations already require employees to wear RFID name tags that identify them and grant them access to several locations and resources. These traditional RFID name tags are usually worn around the neck or clipped to the user's clothing. With the rapid miniaturization of electronics, it is now possible to augment RFID badges with more sensors and computational power that allow us to capture human behavior without requiring any additional effort on the user's side. By capturing individual and collective patterns of human behavior with *Sociometric* badges and correlating these behaviors with individual and group performance, it is possible to identify successful vs. unsuccessful teams, high performing teams, and predict group outcomes. The added value for the users is the feedback that they can receive about their daily behaviors and interactions with others, and how these behaviors affect their individual and group performance.

Collective intelligence is sometimes defined as the ability of a group to solve problems more effectively than any of its individual members (Heylighen, 1999). The term is also used to describe several web tools aimed at improving group performance, such as wikis, social networking sites, and other software programs that facilitate group collaboration. *Sociometric* badges are measurement tools that facilitate the study of collective behavior and help organizations maximize their groups' collective intelligence through specialized software that analyzes behavioral patterns and generates automatic feedback reports and dynamic visualizations. We can design organizational interventions based on these measurements and feedback mechanisms.

In this paper we present two studies in which we used *Sociometric* badges to capture individual behavior and assess group performance from aggregated behavioral features across members of the same group. We focus on behavioral features such as: the amount of face-to-face interaction, the number of different people with whom a person interacts (degree), physical activity, speech activity, and time in close proximity to others, as measured by the *Sociometric* badges. These behavioral features have been used in several of our studies, and have been repeatedly correlated with performance outcomes.

Background: Collective behavior and group performance

There is evidence that aggregating individual performance characteristics across members of a group is indicative of the group's overall performance. Pirola-Merlo and Mann (2004) investigated how the creativity of individual team members is related to team creativity. They found that team creativity scores could be explained statistically by aggregation processes both across people and time.

Increased face-to-face interaction time improves overall group communication and provides opportunities for social networking and relationship building. Driskell and Salas (1992) studied collectively oriented and egocentric team members and found that collectively oriented team members benefit from group interaction to enhance their own performance. Team-oriented groups interacted more, solved problems faster, and were more accurate. They conclude that collective behavior is particularly important for tasks marked by high levels of uncertainty or unpredictability, such as those encountered by teams performing under stressful conditions.

One important aspect of social behavior is the study of non-linguistic signals during face-to-face interactions. According to Pentland (2008), these signals can be measured by analyzing the timing, energy, and variability of speech and body movement patterns. He describes four different types of "honest" signals in humans: *influence* (the extent to which one person causes the other person's pattern of speaking to match their own pattern), *mimicry* (the reflexive copying of one person by another during a conversation), *activity* (speaking time and physical activity level), and *consistency* (low variability in the speech signal or physical activity).

In the studies that we describe next, we looked at two of these signals: **activity** (speaking time and physical activity level), and **consistency** (of speech and body movement). We measure these characteristics at the individual level and aggregated them at the group level in order to assess group performance.

Study 1: Identifying successful teams from collective behavior

We used *Sociometric badges* during the Entrepreneurship Development Program (EDP) at MIT in January of 2009 to capture the participants' social interactions and predict team performance. The EDP is a one-week program for aspiring entrepreneurs, corporate venturing officers, academics, and regional development officers that introduces participants to MIT's entrepreneurial education programs, technology transfer system, and global entrepreneurial network. Participants in the program come from all over the world and get to talk to each other during the first day of the program with the goal of forming a team that will work on a business plan during the following three days. During the last day of the program each team has to present their business plan and their elevator pitches are judged (MIT Entrepreneurship Center, 2009).

During the EDP of 2009 there were 139 participants out of which 109 used a sociometric badge during the first day of the program. On average they wore a badge for 147.44 minutes (std = 70.23 minutes), ranging between [26 and 263] minutes. There were 17 different teams and 3 of them were judged as "winners" in the elevator pitch contest. Several features were calculated from each person's *Sociometric* badge data during the first day of the program:

F1: Number of different people met (degree)
F2: Face-to-face interaction time
F3: Physical activity level
F4: Variation in physical activity level (higher variation means lower consistency and vice versa)
F5: Speech energy
F6: Percentage of speaking time
F7: Time in close proximity to others (or to a specific location)

Results

In this study we used a binary group performance metric for each team depending on whether their business plan was one of the three winners or not. We used logistic regression and bootstrapping to predict the winning teams from the badge features averaged across members of each team. The average accuracy achieved at predicting the winning teams after 100 bootstrapping iterations was 90%. The regression coefficients are:

β_0 = -125.74
β_1 = 662.19(*)
β_2 = 1.6272
β_3 = 497.71(*)
β_4 = -4315.6(*)
β_5 = -427.16(*)
β_6 = 155.84(**)
β_7 = 22(*)

With significance values: $p < 0.05$(*) and $p < 0.01$(**).

Clearly, the best predictor was the average percentage of speaking time (activity) across members of the team (F6). The next best predictors were the number of different people met (F1), physical activity level (F3), consistency or variation in physical activity (F4), speech energy (F5) and time in close proximity to others (F7).

Figure 1 shows the average probability of winning for each of the 17 participating teams after 100 bootstrap iterations. Teams 1, 3, and 13 were the actual winning teams.

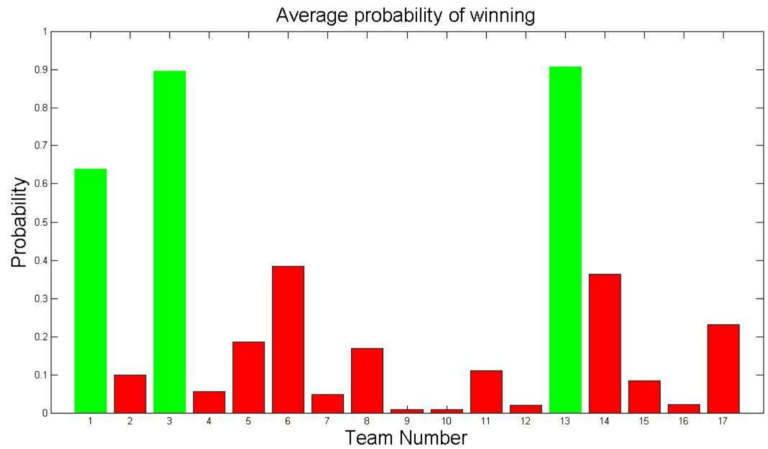

Figure 1. Average probability of being a winning team after 100 bootstrap iterations.

We can see from the regression coefficients that members of successful teams tended to speak more, be more energetic, while at the same time maintaining high consistency (little variation) in their physical activity levels, have lower speech energy, and spend more time in close proximity to others.

Study 2: Predicting patient length of stay from collective nurse behavior

We instrumented a group of 67 nurses working in the Post Anesthesia Care Unit (PACU) of a Boston area hospital with *Sociometric* badges. Using the data collected with the badges we were able to estimate the daily average patient length of stay (LOS) and number of delays (Olguin-Olguin, Gloor, & Pentland, 2009a). In this case we considered all nurses in the PACU to be members of one team, and the team's performance was measured by the daily average LOS and the daily number of patient transfer delays.

The study sample was composed of 67 nurses who worked in the PACU of a Boston-area hospital. Each nurse wore a sociometric badge every day for a period of 27 days. In total we collected 3,906 hours of data. The mean number of hours each participant wore a badge was 7.18 hours per day (±4.17). During this period a total of 1128 patients were admitted to the PACU, with an average LOS of 235.66 (±261.76) minutes.

The daily average and standard deviation of the same badge features described in the previous section were calculated for each participant. We used stepwise multiple linear regression analysis to predict the daily average LOS and number of delays from the daily features aggregated across subjects.

Results

We found that it is possible to predict the variation in the daily average LOS in minutes (R^2 = 0.79) and the daily average number of outgoing delays (R^2 = 0.56) from the aggregated features across nurses. In the case of LOS, the variation in physical activity (F4), and face-to-face interaction time (F2), across nurses played an important role. Low variation across the nurses' level of physical activity (either all nurses having high levels of activity or low levels of activity) and high variation across the nurses' face-to-face interaction time was an indication of extended LOS. In the context of the PACU these results can be interpreted as either most PACU nurses being busy (high activity levels) or waiting for bed availability (low activity levels). The variation across the nurses' face-to-face interaction time could be an indicator of poor communication among nurses.

When estimating the daily number of delays the variation across subjects in their individual physical activity variation (F4) throughout the day and the average time they are in close proximity to a phone (F7) were the most predictive features. This means that a high variation across the nurses' daily activity levels (having alternate periods of high activity and low activity during the day), coupled with the variation in the time they spend in close proximity to a phone, is an indication of increased number of delays in the PACU.

Conclusions

The use of pervasive sensors has allowed us to study human behavior with unprecedented levels of detail. By capturing individual behaviors such as the amount of face-to-face interaction, speaking patterns, and non-linguistic social signals; and aggregating them at the group level, it is possible to assess group performance and find optimum team configurations.

We presented results from one study in which we were able to predict the top-performing teams from their interaction patterns within the first day of the team formation process. Aggregated features such as the amount of time spent in face-to-face conversations and non-linguistic signals such as activity and consistency were predictive of team performance.

The results from our second study show that it was possible to assess the overall performance of a post-anesthesia care unit by analyzing aggregated behavioral features across all nurses working in the unit as a group. Further applications of our work include automatic clustering of people to maximize team performance as well as dynamic visualizations of team processes. Some implications that this work has for future studies and future technologies for collective intelligence in organizations are:

- Studies confirming social science theories based on human observation can be corroborated using automatic measurement tools on larger populations.
- Human behavior will soon be automatically captured and analyzed using electronic tools (sensors, websites, software, etc.) in organizations. Most of the required infrastructure and sensors are already in place.
- Further collaboration tools and technologies that make use of sensor data to promote collective intelligence in organizations will emerge.
- Users should have the right to set their privacy settings, have access to their data, know what kind of data is being collected, and decide how their data will be managed.
- Privacy concerns will be overthrown by the potential benefits for the users and the organizations.

References

Choudhury, T., & Pentland, A., 2003. Sensing and Modeling Human Networks Using the Sociometer. In: *7th International Symposium on Wearable Computers.* October 21-23, 2003.

Driskell, J. E., & Salas, E., 1992. Collective Behavior and Team Performance. *Human Factors*, 34 (3), pp. 277-288.

Gloor, P. A. et al., 2007. *Studying Microscopic Peer-to-Peer Communication Patterns*. AMCIS 2007: Americas Conference on Information Systems. Keystone, Colorado. August 9-12, 2007.

Heylighen F., 1999. Collective Intelligence and its Implementation on the Web: algorithms to develop a collective mental map. *Computational and Mathematical Organization Theory,* 5(3), pp. 253-280.

MIT Entrepreneurship Center. *MIT Entrepreneurship Development Program*. Retrieved on December 12, 2009 from: <http://entrepreneurship.mit.edu/edp.php>

Olguin-Olguin, D., 2007. Sociometric Badges: Wearable Technology for Measuring Human Behavior. *Master's Thesis*. Cambridge, MA, USA: Massachusetts Institute of Technology.

Olguin-Olguin, D., Gloor, P. A., & Pentland, A., 2009a. Wearable Sensors for Pervasive Healthcare Management. In: *3rd International Conference on Pervasive Computing Technologies for Healthcare*, pp. 1-4. April 1-3, 2009. London, UK.

Olguin-Olguin, D., Waber, B., Kim, T., Mohan, A., Ara, K., & Pentland, A., 2009b. Sensible Organizations: Technology and Methodology for Automatically Measuring Organizational Behavior. *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, 39 (1), pp. 43-55.

Pentland, A., 2006. Automatic Mapping and Modeling of Human Networks. *Physica A: Statistical Mechanics and its Applications*, 378 (1), pp. 59-67.

Pentland, A., 2008. *Honest Signals: How they Shape our World*. The MIT Press: Cambridge, MA.

Pirola-Merlo, A. & Mann, L., 2004. The relationship between individual creativity and team creativity: aggregating across people and time. *Journal of Organizational Behavior*, 25, pp. 235-257.

Wu, L., Waber, B. N., Aral, S., Brynjolfsson, E., & Pentland, A., 2008. Mining Face-to-Face Interaction Networks Using Sociometric Badges: Predicting Productivity in an IT Configuration Task. In: *International Conference on Information Systems.* December 14-17, 2008. Paris, France.