Recognizing Expertise

Benjamin N. Waber and Alex (Sandy) Pentland

MIT Media Laboratory
20 Ames St. Cambridge, MA 02139 USA
{bwaber, sandy}@media.mit.edu

Abstract

A difficult issue with understanding face-to-face interactions is the lack of context information. Badge-like wearable sensors, for instance, can now tell us who talks to whom, but the context of these interactions (work? social?) likely trumps any content or conversational dynamics information. In this paper we detail a study where we have automatic annotations of work context information. We find that this leads to different indications of effective social structure in corporate contexts, showing that behavior during tasks drives the predictive strength of aggregated data. Finally we describe how behavioral features collected with the Sociometric Badge are being used to create a real time feedback system.

Introduction

Organizations are not sensible. Rather than manage how people actually talk to each other, formal reporting relationships are altered without a true appreciation for how this meshes with organizational social structure. Employee training is inherently a continuous process, yet modern organizations usually train employees in fits and starts. Finally, the physical structure of organizations is often unresponsive to the immediate needs of the organization.

What if we could use sensors and real-time data to create a Sensible Organization? We could determine what behaviors lead to better performance, continuously training employees with feedback systems. We could enable employees to collaboratively define the social structure of the organization using introduction programs, or use actuators to transform the physical face of the organization.

To accomplish all of this, we first must understand how real-world organizations function, and what types of data are important for enabling these applications. What patterns of behavior are actually effective during work? How does this differ with when employees are not working? Particularly important is detecting employees that are sources of expertise within the organizations and how interacting with experts helps or hinders productivity.

Recently wearable sensors - essentially smart ID badges - have enabled us to study human behavior at an unprecedented scale. Not only can we use accelerometers and microphones to understand subsecond behavioral changes, but we can study behavior continuously for months and even years. In addition, sensors allow for the measurement of thousands of individuals. One of the most fertile application grounds for such technology is organizational behavior.

A difficult issue with much of the current work in real world environments is the lack of context information. Sensors can tell us who talks to whom, but the context of these interactions likely trumps any content or conversational dynamics information. For example, if two people are talking while they are working on a task it has very different connotations than when they are talking on a break.

Previously, most experiments have simply aggregated all interaction and behavioral information irrespective of context. Experimental results have proved extremely promising in some cases, although in other cases a surprising lack of predictive power emerged. In this paper we detail a study where we have automatic annotations of work context information, allowing us to compare the predictions of contextual and non-contextual methodologies. We find that this leads to different indications of effective social structure in corporate contexts, showing that behavior during tasks drives the predictive strength of aggregated data. Since the goal of this project is to enable real-time feedback systems for organizational change, we detail the amount of data that is required to create such systems with the prescriptive power that we describe.

Background

Expertise

At its core, what is expertise? Informally, we believe that someone is an expert when, for a particular type of question, that person will have the best answer. Al Gore is an expert on climate change. Warren Buffet is an expert on the stock market. Within organizations this definition doesn't change, only the scale does. People who are authorities on specific types of tasks become important.

Still, the context of interactions with these experts is extremely important. If one of Warren Buffet's friends is talking to him in the gym, chances are they are not discussing the stock market. Similarly, coworkers who are not working and talking over coffee are probably not discussing how to solve certain tasks. To recognize expertise, then, it is important to recognize the context of interactions as well.

In the context of this paper, we define expertise as *the centrality of an individual in the social network formed by examining only interactions that occur during work*. This definition is similar to those presented in (Barahona & Pentland, 2006), (Ibarra & Andrews, 1993), (Pujol, Sanguesa, & Delgado, 2002), and so is line with previous work. Here we define centrality as betweenness centrality, although we did not find any difference in predictive ability when we used other network position-based centrality measures.

Sociometric Badges

In order to recognize expertise, we first need to sense the underlying interactions that are occurring in the workplace. For this purpose we have created a wearable *Sociometric* badge that has advanced sensing, processing, and feedback capabilities (Olguin Olguin, 2007). In particular, the badge is capable of:

- Recognizing common daily human activities (such as sitting, standing, walking, and running) in real time using a 3-axis accelerometer (Olguin Olguin & Pentland, 2006).
- Extracting speech features in real time to capture nonlinguistic social signals such as interest and excitement, the amount of influence each person has on another in a social interaction, and unconscious back-and-forth interjections, while ignoring the words themselves in order to assuage privacy concerns (Pentland, 2005).
- Performing indoor user localization by measuring received signal strength and using triangulation algorithms that can achieve position estimation errors as low as 1.5 meters, which also allows for detection of people in close physical proximity (Sugano, Kawazoe, Ohta, & Murata, 2006), (Gwon, Jain, & Kawahara, 2004).
- Communicating with Bluetooth enabled cell phones, PDAs, and other devices to study user behavior and detect people in close proximity (Eagle & Pentland, 2006).
- Capturing face-to-face interaction time using an IR sensor that can detect when two people wearing badges are facing each other within a 30°-cone and one meter distance. Choudhury (Choudhury, 2004) showed that it was possible to detect face-to-face conversations of more than one minute using an earlier version of the Sociometric badge with 87% accuracy.

This represents a fundamental shift from earlier work in organizational behavior, since with this technology we are able to objectively quantify behavior at a level of detail unimaginable just a few years ago. In addition, we can examine radically different behavioral features than is possible using traditional observational and survey methods.

Detecting Expertise

Surveys are good at soliciting individual impressions of relationships, but they are not objective or at a very fine scale. They are still the most widely used method for collecting relationship data, partially because it's possible to elicit context information within the questions. For example Barahona et al. (Barahona & Pentland, 2006) used surveys to discover who the experts within an agricultural community were.

The advent of Sociometric badges, however, has made surveys a far less attractive option. They are objective and offer a level of detail unsurpassed by even e-mail and IM data mining methods. The issue with Sociometric badges is that by themselves they give very little indication of context, making it difficult distinguish when people are giving advice versus talking about the weather. That is, from sensor data it is difficult to determine in *general* situations exactly what high level activity the individual is engaged in. By supplementing Sociometric badges with survey and other types of data, however, this limitation can be overcome. In our study, in addition to badge data we obtained timestamped company records that indicated exactly when people were working on a task. From this, we were able to operationally define expertise.

Given our definition of expertise, it seems most relevant to use betweenness centrality of an individual within the advice network as a measure of expertise since these people give advice that has the highest overall impact. One could also choose closeness and other network measures that are more dependent on the strength of advice ties however in our data this choice did not substantially change our results. Other arguments may be made for using degree centrality, since these people are giving the highest volume of advice. On an organizational level, however, these people are less important than those with higher centrality, since people with high betweenness are often ultimately the source of advice dispensed by others.

Experiment

We deployed our Sociometric badge platform for a period of one month (20 working days) at a Chicago-area data server configuration firm that consisted of 28 employees, with 23 participating in the study. Each employee was instructed to wear a Sociometric badge every day from the moment they arrived at work until they left their office. In total we collected 1,900 hours of data, with a median of 80 hours per employee. All of these employees were male, and since this was a recently formed department none had been employed for over a year. Still, there were five recognized experts, and in our analysis we controlled for skill level differences. Electronic communication was not extensively utilized in this firm for task-related communication, so we did not collect this data. Now we will explain the actual task structure for these employees, and in the analysis below we examine employee behavior at the task level rather than at the individual level. This allows for a much finer-grained analysis than would otherwise be possible, as well as uncovers some startling results.

Task Structure and Productivity Data

Salesman in the field used an automated program to request a computer system configuration for a potential customer. These configurations are automatically assigned a difficulty (basic, complex, or advanced, in ascending order of difficulty) based on the configuration characteristics. Employees in the department are then assigned a configuration task in a first come first served fashion. This configuration task may require them to use a computer aided design (CAD) program in order to satisfy the customer's needs. Finally, the employee submits the completed configuration as well as price back to the salesman, and the employee is placed at the back of the queue for task assignment. The exact start and end time of the task is logged, and the number of follow-ups that are required after the configuration is completed is also recorded in the database. We were able to obtain this data in addition to the badge data, although in our analysis we only examined tasks where the employee was wearing the Sociometric badge for the entire task duration.

We omitted tasks that took no time to complete, as this was due to preconfigured systems being purchased rather than actual work by the employee. We also omitted instances where an employee only completed one task in the available data, since then we were unable to control for behavioral differences in our analysis. In our final dataset we have 170 basic tasks, 16 complex tasks, and 34 advanced tasks. In our analysis we only consider basic tasks, although we note that both complex and advanced tasks exhibited similar trends with the basic tasks.

We used (negative) completion time as our measure of productivity, since shorter completion times are more desirable, and in this organization employees are rewarded based on their throughput. Another important measure are follow-ups, where the employee was contacted again by the customer. This mostly occurred because the original job was insufficiently accurate or complete. We controlled for individual productivity and behavioral differences in our analyses at the task level.

Previous Work on This Dataset

In previous work (Waber, Olguin Olguin, Kim, & Pentland, 2008) we found that productivity had a high significant negative correlation with the standard deviation of activity ($r = -0.50$, $p < 0.001$). This relationship between productivity and variation in activity is also shared by variation in speaking behavior, as cases in which the subject spoke to others had a much higher negative correlation with productivity ($r = -0.59$, $p < 0.00005$) than cases in which the subject did not speak to others ($r = -0.39$, $p < 0.001$).

We also discovered that overall the number of follow-ups is highly negatively correlated with productivity ($r = -0.57$, $p < 0.001$). This effect is much stronger in non-speaking cases ($r = -0.67$, $p < 0.00005$) than in speaking cases ($r = -0.45$, $p < 0.001$). This has important practical implications, since intuitively predicting follow-ups is extremely important in this organization for understanding how employee time will be allocated in the near term.

For our work here, it is important to note that interacting with others mediated the predictive power of the features we collected.

Methods

Recognizing Interactions

We are able to recognize face-to-face interactions by combining IR and microphone information. When two individuals are standing facing each other, there is an IR detection. Combining this with speaking information allows us to determine that the people were actually in a conversation. This method is elaborated in (Waber, Olguin Olguin, Kim, & Pentland, 2008). Both frequency and length of interactions have no correlation with productivity, and thus we ignore them below.

Measuring Expertise

In this study we contextualize the interactions by differentiating circumstances where employees are working from those when they are not. We then look at the centrality of each person during work to measure expertise. During tasks, we use the maximum expertise value of all interaction partners to measure the expertise that was accessed, although this had extremely high correlations with both the summed expertise values and average expertise values ($r = 0.93$, $p < 0.0001$ and $r = 0.82$, $p < 0.0001$, respectively).

Recognizing Work

In general, the behavior across all activities is statistically extremely similar to behavior during tasks. While interaction with an expert has a strong indication that an individual is working, this does not necessarily occur at the start of a task, and thus it is difficult to determine the exact time extents of working regions. In our case we have this data given to us automatically by a corporate logging system, and these results imply that such an electronic logging system is necessary to obtain similar data in other studies.

Results

We found that expertise aggregated over the length of the study had a moderate positive correlation with individual productivity ($r = 0.4$, $p < 0.001$). This is not unexpected since results such as this have been frequently reported in previous studies using surveys (Baldwin, Bedell, & Johnson, 1997), (Cross & Parker, 2004), (Ibarra & Andrews, 1993), (Mehra, Kilduff, & Brass, 2001), (Roberts & O'Reilly, 1979). What is interesting is when we contrast this with the measures of expertise that aggregate data across all activities. In this case, betweenness was not significantly correlated with productivity, demonstrating how essential this context information is.

But what happens at the task level? That is, does interacting with higher level experts enhance or hamper completion rate? Again, we controlled for behavioral and productivity differences by dividing all variables by their averages at the individual level, and we only examined tasks where people interacted with others ($n = 35$). In this case, interacting with higher level experts had a high *negative* correlation with productivity ($r = -0.67$, $p < 0.0001$).

The most likely explanation for this result is that people were assigned a task that was beyond their skill level, and therefore they had to talk to other employees with greater knowledge, and the people that are most likely to have that knowledge are those who are central.

Combining this measure of expertise accessed during a task with the standard deviation of movement energy, which we found to be predictive of productivity in a previous analysis of this data (Waber, Olguin Olguin, Kim, & Pentland, 2008), we performed a multi-linear regression. The results of this regression are shown in table 1 below. This model had extremely high predictive power with $r = 0.81$, showing that contextualized sensor data and using simple behavioral features may be sufficient for developing feedback interfaces as well as to suggest changes to the organization. As we showed above, interacting with others did have a strong effect on the other features that we collected, explaining the comparatively lower importance of the movement variable. In future work we hope to tease apart these variables to determine the causal effect of each.

Total Model: $r = -0.81$, $F = 22.65$, $p < 0.0001$		
Feature	β	p
Intercept	0.31	0.16
Movement σ	-0.29	0.002
Expertise	-0.81	<0.0001

Table 1. Productivity prediction using behavioral features, with the regression coefficients for each feature and their significance.

Discussion

These results are crucial for understanding how to apply sensing data to organizational management. This organization could change its methods of talent identification, looking not only at individual productivity but examining the level of expertise an individual has. In future experiments it is crucial that the causality of this feature be examined (i.e. does position in the contextualized interaction network cause higher productivity, or vice versa).

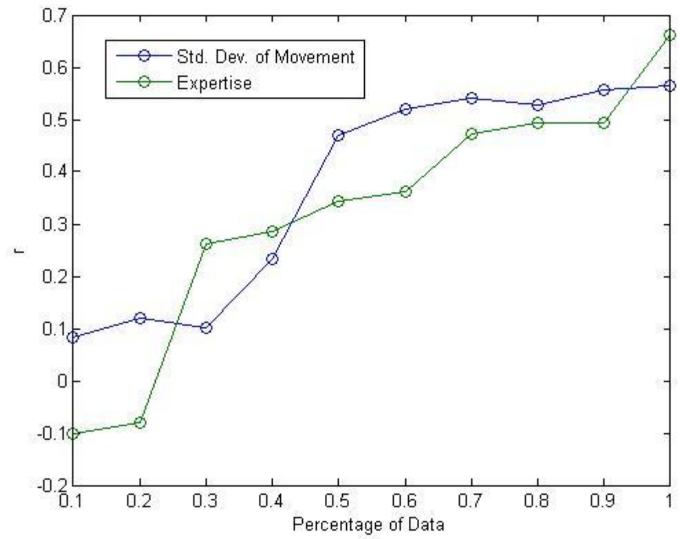

Figure 1. Predictive power of features as a function of the first X% of the data is used. The standard deviation of movement energy is significantly predictive when at least 40% of the data is used, while maximum expertise accessed is significantly predictive when at least 60% of the data is used.

Our findings at the task level could also fundamentally change the way that this organization trains their employees. Employees should be trained to recognize their own limits and given directions about who they should contact when they have problems, taking care that people with high levels of expertise are not overloaded.

Real-Time Feedback Systems

The obvious application of these results is to build feedback systems, and it important to ask under what conditions could a system make use of this data. That is, how long would a system have to collect data in order to make an accurate prediction about productivity on that task?

To investigate this problem we extracted behavior data during tasks at 10% increments. That is, we took the first 10, 20, 30,… 100 percent of data for each task and compared the computed features with those calculated when 100 percent of the data was available. We show the results below in figure 1. We broke up the data into two groups: movement data and interaction data, since the characteristics of these two types of data are quite different. From these results it is possible to derive the optimum threshold for collecting data for a real-time feedback system.

Future Work

We believe that this work represents a significant step towards not only understanding organizations, but also understanding how sensing technology can drive real world feedback systems. It is crucial that designers of feedback systems understand what the limits of predictive methods are and how this can be used to drive data collection.

We are currently implementing feedback systems based on these principles, in particular examining if we can influence not only physical behavior, but interaction patterns as well. We hope to enable a Sensible Organization, an organization system that takes account of sensor, electronic, and survey data to collaboratively change the social structure, training, and even the physical architecture of the organization. This will enable a new definition of management, training, and what it means to go to work.

Acknowledgements

The authors would like to thank Daniel Olguín Olguín and Taemie Kim for their help with creating the Sociometric Badge platform, Inna Koyrach for her help with the interaction detection algorithm, and Koji Ara for assisting with data collection. This work was made possible with support from Hitachi, Ltd. and the Things That Think consortium at the MIT Media Laboratory.

References

Baldwin, T., Bedell, M., and Johnson, J. (1997). The social fabric of a team-based M.B.A. program: Network effects on student satisfaction and performance. *The Academy of Management Journal* , 1369-1397.

Barahona, J. C., and Pentland, A. (2006). Advice and Influence: The Flow of Advice and the Diffusion of Innovation. *The XXVI International Sunbelt Social Network Converence.* Vancouver, British Columbia, Canada.

Choudhury, T. (2004). *Sensing and Modeling Human Networks.* Cambridge, MA USA: PhD Thesis, MIT Media Laboratory.

Cross, R., and Parker, A. (2004). *The Hidden Power of Social Networks.* Boston, MA USA: Harvard Business School Publishing.

Eagle, N., and Pentland, A. (2006). Eigenbehaviors: Identifying structure in routine. *Ubicomp 2006: The 8th International Conference on Ubiquitous Computing.*

Gwon, Y., Jain, R., and Kawahara, T. (2004). Robust indoor location estimation of stationary and mobile users. *Proceedings of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies*, (pp. 1032-1043).

Ibarra, H., and Andrews, S. B. (1993). Power, Social Influence, and Sense Making: Effects of Network Centrality and Proximity on Employee Perceptions. *Administrative Science Quarterly* , 277-303.

Mehra, A., Kilduff, M., and Brass, D. (2001). The social networks of high and low self-monitors: Implications for workplace performance. *Administrative Science Quarterly* , 121-146.

Olguin Olguin, D. (2007). *Sociometric Badges: Wearable Techonology for Measuring Human Behavior.* Cambridge, MA USA: Massachusetts Institute of Technology.

Olguin Olguin, D., and Pentland, A. (2006). Human activity recognition: Accuracy across common locations for wearable sensors. *Proceedings of the 10th International Symposium on Wearable Computers (Student Colloquium)*, (pp. 11-13).

Pentland, A. (2005). Socially Aware Computation and Communication. *IEEE Computer* , 33-40.

Pujol, J. M., Sanguesa, R., and Delgado, J. (2002). Extracting Reputation in Multi Agent Systems. *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1* (pp. 467-474). Bologna, Italy: ACM.

Roberts, K., and O'Reilly, C. (1979). Some correlates of communication roles in organizations. *The Academy of Management Journal* , 42-57.

Sugano, M., Kawazoe, T., Ohta, Y., and Murata, M. (2006). Indoor localization system using RSSI measurement of wireless sensor network based on zigbee standard. *Proceedings of the IASTED International Conference on Wireless Sensor Networks*, (pp. 1-6).

Waber, B. N., Olguin Olguin, D., Kim, T., and Pentland, A. (2008). Understanding Organizational Behavior with Wearable Sensing Technology. *Acadmey of Mangement Annual Conference.* Anaheim, CA, USA.