Ego-Centric Graphlets for Personality and Affective States Recognition

Stefano Teso[1], Jacopo Staiano[1], Bruno Lepri[2,3], Andrea Passerini[1], Fabio Pianesi[3]

[1]DISI-University of Trento, [2]MIT Media Lab, [3]FBK

Introduction

Affect and personality permeate people's daily and working life and also the interdependent relationships they usually hold with bosses, colleagues, and subordinates. Several studies showed the relationships between personality and e.g. job performance [1], work motivation [2], job satisfaction [3].

In this work, we investigate how the details and the complexity of the social network structure of the interacting alters can play a significant role in influencing the ego's expression of a particular affective/emotional state or a specific personality state in a work environment in predicting the affective and personality states of the ego. The social psychology literature has recently coined the term personality states to refer to concrete behaviors that can be described as having a similar content to the corresponding personality traits. In other words, a personality state describes a specific behavioral episode wherein a person behaves more or less introvertly/extravertly, more or less neurotically, etc.

To this end, we represent people's interactions as graphlets, induced subgraphs representing specific patterns of interaction, and we design classification experiments with the target of predicting the subjects' self-reported personality and affective states. Graphlets have extensively been employed to study properties of biological networks, e.g. to discover invariant patterns characterizing specific properties of enzymes and small molecules. Being able to capture the local structure of interactions, graphlets represent a promising methodology to study interactions between humans in the online and offline worlds.

We investigate graphlets centered on the reference node (the ego), embedding information on the state of the alters and their interactions in order to recognize the affective/personality state of the ego. We explore how interaction patterns, encoded as graphlets, gathered from two distinct sensory channels, Bluetooth (BT) and Infrared (IR), affect recognition of personality and affective states.

Our Data

For this study we exploited the SocioMetric Badges Corpus [4], a multimodal corpus specifically designed to capture the psychological and situational aspects of the daily lives of employees in an organizational structure. The data were collected in a research institute for six weeks and involved a sample of 54 subjects (46 subjects that belong to four computer science research groups and 7 subjects of the IT department) during their working hours. Males predominated (90.8%) while the average age was 36.83 years with a standard deviation of 8.61 years. Out of the 54 subjects, 37 subjects were researchers and employees, 7 had a leading role while 10 were doctoral students.
The data about subjects' activities were collected using wearable sensors called Sociometric Badges. These sensors are equipped with accelerometers, audio, Bluetooth and Infrared to respectively capture: body movements, prosodic speech features, co-location with other individuals and face-to-face interactions [5]. In this work, we exploit information from the IR and the BT sensors.
An Experience Sampling Methodology (ESM) was employed to collect information about transient psychological states (personality and affectivity states). Participants completed a short Internet-based survey three times a day during the morning (11:00 AM), the afternoon (2:00 PM) and the evening (5:00 PM). Links to the surveys were automatically administered via email at fixed times and users were granted a temporal window of 2.5 hours to fill the survey before its expiration. Participants were asked to confirm their presence in the institute during the 30 minutes before starting the questionnaire; only if confirmed, their responses would be included in the database.The questions in the experience sampling surveys referred to affectivity and personality states experienced over the 30 minutes prior to the survey.
The ten-item personality inventory TIPI [6] was used to assess personality states. A 7-point scale ranging from 1=Strongly Disagree to 7=Strongly Agree was used for re- sponses. The scores for each state were calculated by summingthe raw scores of the two corresponding items, with proper inversion when needed.
Similarly, respondents were asked to report on a scale from 1 to 5 (1=Very Slightly Or Not At All and 5=Extremely) to what extent they experienced High Positive Affect (HPA) and/or High Negative Affect (HNA) in the 30 minutes before starting to fill the survey. HPA and HNA were assessed by means of a 6-items shortened version of the Positive and Negative Affect Schedule (PANAS) [7].

Our Approach

We define a binary classification task for each subject and each personality and affective state. This is done by mapping the state of a given subject at a certain deadline from {1, . . . , 7} to {0, 1} using its median value for the subject as a threshold. Therefore, negative labels represent cases where the subject was found below his/her median.

One of the main contributions of this paper lies in the encoding of the subjects' interactions as graphlets, defined as induced subgraphs of a larger network, providing a succinct representation of social structure. In the Bioinformatics and Computational Biology domains, graphlets have been introduced for the study of large biological networks, for e.g. network alignment [8]. Recently, graphlet analysis has been applied to Facebook messaging and historical crime data [9]. We investigate their effectiveness in the context of a human interaction network, for the prediction of behavioral determinants such as personality and affective states.

Starting from the network of interactions between subjects, we extract for each subject the graphlets representing his/her local interactions. In this work, we consider all possible graphlets up to 4 nodes. Furthermore, the graphlets embed information on the current (binary) state of the alters (but not of the reference subject whose state is to be predicted), in order to account for possible influence and propagation effects.

For each deadline, we extract graphlet-based features from sensory data gathered over the previous 3 hours. We discretize each 3-hour window in 15-minutes slices in order to represent the evolution of the interaction patterns over time, taking into consideration the neighbours' states in order to account for situational influence effects. To do so, we count occurrences of graphlet configurations and build a histogram; then, we average the histograms obtained for each slice and obtain a feature vector representative of the 3-hours window under analysis. Finally, we use it as feature vector to predict the ego's state at the deadline under analysis.

In our setup, two kinds of missing data are possible: i) missing labels (i.e. surveys not filled by the subjects); ii) missing interactions, in which case the interaction graphs will be empty. In both cases we exclude the deadline under analysis from the training and testing stages.

Since we are interested in understanding influence and propagation effects on a given subject, to classify a specific target we only rely on features derived from labeled graphlets computed on the same target.

We build a linear Support Vector Machine (SVM) model [10] for each agent and each target state, evaluating its performance in a leave-1-week-out cross-validation procedure. We employed the LibLinear [11] library with L1-regularization, which tends to produce sparse models (with few non-zero weights). The learned models prioritize informative features, leading to robust handling of noise, and are simpler to interpret. To avoid any bias in the interpretation of the results, we discard all agent/target state pairs for which one class (i.e. positive or negative) covers at least 75% of the instances. This occurs when subjects exhibit very little variance on the labels and thus many instances fall on the median value itself. We build models of increasing complexity by considering graphlet-based features made of up to one, two and three alters respectively, and evaluate the performance of each model in order to assess the predictive power of different levels of interaction. We compare these models against each other and against a majority classifier (i.e. a classifier that always predicts the class with more instances in the training set), which we use as a baseline.

Results and Discussion

The major goals of this paper were to investigate new perspectives on affect and personality states recognition and move the first steps towards addressing it. In particular, we study the influence played by specific situational factors, the face-to-face interactions and the proximity interactions with alters, over the ego's expression of a particular affective/emotional state or a specific personality state, a concrete behavior that can be described as having a similar content to the corresponding personality trait.

To this end, we investigate graphlets centered on the reference node (the ego), embedding information on the state of the alters and their interactions in order to recognize the affective/personality state of the ego. We thus explore how interaction patterns, encoded as graphlets, gathered from two distinct sensory channels, Bluetooth (BT) and infrared (IR), affect recognition of personality and affective states. Graphlets are able to represent not only the size of the ego-network but also different levels of its structural complexity.

The results of our experiments show that feature-sets including graphlet-based features constantly contribute to performance improvements and more complex configurations of graphlets often correspond to higher improvements. These results show the feasibility of the proposed perspective and hopefully encourage further research.

Another interesting finding is that graphlet-based features derived by different sensory channels have different impacts on the recognition performance of distinct target states: e.g. more complex graphlet-based features have a big impact for Conscientiousness when derived from IR data and a small one when coming from BT data. An opposite trend is found for Low NA. These findings supports the intuition that these two channels are able to capture different concrete behaviors: BT captures proximity in a broad-cast manner (i.e. many-to-one), IR is restricted to face-to-face (thus one-to-one) interactions

References

[1] M. R. Barrick, M. K. Mount, and T. A. Judge, “Personality and Performance at the Beginning of the New Millennium: What Do We Know and Where Do We Go Next?” International Journal of Selection and Assessment, vol. 9, no. 1&2, pp. 9–30, 2001.

[2] R. Ilies and T. A. Judge, “Understanding the dynamic relationships among personality, mood, and job satisfaction: A field experience sampling study,” Organizational Behavior and Human Decision Processes, vol. 89, no. 2, pp. 1119–1139, 2002.

[3] T. A. Judge, D. Heller, and M. K. Mount, “Five-factor model of personality and job satisfaction: a meta-analysis.” Journal of Applied Psychology, vol. 87, no. 3, pp. 530–541, 2002.

[4] B. Lepri, J. Staiano, G. Rigato, K. Kalimeri, A. Finnerty, F. Pianesi, N. Sebe, and A. Pentland, “The sociometric badges corpus: A multilevel behavioral dataset for social behavior in complex organizations,” in SocialCom/PASSAT, 2012, pp. 623–628.

[5] K. Ara, N. Kanehira, D. O. Olgu ́ın, B. N. Waber, T. Kim, A. Mohan, P. A. Gloor, R. Laubacher, D. Oster, A. Pentland, and K. Yano, “Sensible organizations: Changing our businesses and work styles through sensor data,” Journal of Information Processing, vol. 16, pp. 1–12, 2008.

[6] S. D. Gosling, P. J. Rentfrow, and W. B. Swann, “A very brief measure of the big-five personality domains,” Journal of Research in Personality, vol. 37, pp. 504–528, 2003.

[7] D. Watson, L. A. Clark, and A. Tellegen, Development and validation of brief measures of positive and negative affect: The PANAS scales, Journal of Personality and Social Psychology, 54(6), pp. 1063-1070, 1988.

[8] N. Przulj, “Protein-protein interactions: Making sense of networks via graph-theoretic modeling,” Bioessays, vol. 33, no. 2, pp. 115–123, 2011.

[9] H. A. Soufiani and E. Airoldi, “Graphlet decomposition of a weighted network,” Journal of Machine Learning Research - Proceedings Track, vol. 22, pp. 54–63, 2012.

[10] C. Cortes and V. Vapnik, “Support-vector networks,” Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.

[11] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, “Liblinear: A library for large linear classification,” J. Mach. Learn. Res., vol. 9, pp. 1871–1874, Jun. 2008.