Abstract

Understanding personality traits and group behavior of employees are essential for productivity, efficiency and long term health of human resources for any enterprise. Psychologists studying enterprises, especially large geographically diverse organizations continuously, is hard to scale. Regular surveys become exhausting and domain specific context is harder to capture for outsiders. Enterprise social media engagement provides an unobtrusive passive stream of the cognitive state of the employees. In this thesis our goal is to computationally mine the same and infer employee behavior in an unsupervised manner as obtaining labels run into regulatory and scale issues.

Our first contribution is mining text processing features that relate to the author's state of mind and topical interests on social media, such as sentiment analysis, named entity extraction and linking. We address the challenges due to the noisy nature of social media text. Our second contribution is an extension of the dictionary of personality describing words to increase coverage of a linguistically diverse data set such as ours, thereby identifying individual personality traits. The seed words are derived from well established organizational behavior studies. We find low coverage from linguistics alone, so in our third contribution we develop a hierarchical time series clustering that groups behaviorally similar employee's together using only their temporal activity in terms of number of posts each month as the input. We define a consistency score as well that indicates organizational relevance. In our final contribution we develop a multigraph edge attributed node embedding algorithm to identify behaviorally similar groups of employees across the organization who may be recommended to each other to increase collaboration and efficiency.

We present experimental results on four enterprise social network data sets. Among them the Enron email data set represents a commercial organization. The StackExchange Q&A and the Linux Kernel Mailing List represents open source communities. The Internet Engineering Task Force mailing list represents a standard driven organization.

Keywords: *Enterprise Social Media, Personality Trait Identification, Group Behavior, Time-Series, Clustering, Sentiment Analysis, Psycholinguistics*