Learning your Life: Wearables and Familiars

Brian Clarkson, Sumit Basu, Nathan Eagle, Tanzeem Choudhury, Alex Pentland
M.I.T. Media Laboratory Human Design Group
clarkson,sbaus,nathan,tanzeem,pentland@media.mit.edu

Abstract

The wearable platform is an important perspective from which to collect developmental data. While not an agent of its own accord, it is like the mythical "familiar" that sits perched on a user's shoulder, seeing what he sees, with the opportunity to learn what he learns. This gives a developmental agent the ability to experience many rich aspects of human behavior, locomotion, interaction, and social structure without knowing how to actively participate in these activities. In the I Sensed project, we combine natural sensor modalities (camera, microphone, gyros) in a wearable framework to build a first prototype of such an agent. We have also taken the next step to build robust statistical models with a massive data collection experiment: 100 days of full surround video, audio, and orientation, amounting to over 500 Gigabytes of data. The first challenge with this data is the discovery and prediction of daily patterns - can we automatically infer the typical paths through someone's day and their daily activities, predicting what they will do next and detecting anomalies.

Armed with this kind of omnivideo sensor data, we can also apply our learning work tools to conversational scene analysis to help the agent develop a rudimentary understanding of social interactions. This work is aimed towards understanding the structure of face-to-face conversations - not speech recognition, but instead parsing the states of the interaction. For instance, we would like to know if where people are arguing, if they are making small talk, is something being explained to others, etc. This begins at the lowest levels, building hierarchical dynamic models for speech production, and continues with the learning of a state space of interactions and the dynamics of that state space. Learning this structure helps the agent decide when to speak, to interrupt, or back down in an interaction.

The agent can also understand some of the content of these interactions by leveraging our recent work on topic spotting. The OverHear project has resulted in a classifier that takes the noisy output of a commercial speech recognizer and extracts the conversation context. The principle behind this is simple - much like text classification, a particular context implies a variety of words that are unique to it, and by amassing this information over time, we can reliably decide which context the words are coming from.

At a larger scale, the agent can learn how the communication structure evolves within a group of people. In the Shortcuts project, we are developing methods to automatically and unobtrusively learn the social network structure that arises within a group based on wearable sensors. Computational models of group interaction dynamics are derived from data gathered using wearable sensors. The questions we are exploring are: Can we tell who influences whom? Who are the key players or connectors in the community? How does information diffuse within the community? How can we modify group interactions to promote better information diffusion? The goal is real-time learning modification of social network relationships by applying statistical machine learning techniques to data obtained from unobtrusive wearable sensors.