Enhancing Social Interactions of Individuals with Visual Impairments: A Case Study for Assistive Machine Learning

Vineeth Balasubramanian, Shayok Chakraborty, Sreekar Krishna, Sethuraman Panchanathan
Center for Cognitive Ubiquitous Computing (CUbiC)
Arizona State University, Tempe, AZ 85287
{vineeth.nb,schakr10,sreekar.krishna,panch}@asu.edu

Abstract

Individuals with visual impairments face serious challenges in experiencing the fundamental privileges of social interactions. The realization of a Social Interaction Assistant (SIA) device for such individuals involves solving several challenging problems in pattern analysis and machine intelligence such as person recognition/tracking, head/body pose estimation, posture/gesture recognition, expression recognition, and human-object interaction recognition on a combination of wearable and ubiquitous computing platforms. This work presents sample machine learning contributions that have been made as part of the development of such a SIA device, including integrated face localization and detection, user-conformal confidence measures and online active learning.

More than 1.1 million individuals in the US who are legally blind (and 37 million worldwide) have a limited experience of the fundamental privilege of social interactions, since non-verbal cues (including prosody, elements of the physical environment, the appearance of communicators and physical movements) account for as much as 65% of the information communicated during such interactions [1]. The realization of a Social Interaction Assistant (SIA) device involves solving several challenging problems in pattern analysis and machine intelligence such as person recognition/tracking, head/body pose estimation, posture/gesture recognition, expression recognition, and human-object interaction recognition on a combination of wearable and ubiquitous computing platforms. While these problems are typically encountered in many other fields including robotics, the presence of the 'human in the loop' raises unique perspectives in the design of machine learning (ML) algorithms for recognition and learning in the SIA. This abstract presents a few ML contributions that have been made as part of the development of such a SIA for individuals with visual impairments.

In our current prototype, the Interaction Assistant device consists of a pair of glasses with a camera mounted on the nosebridge. When a person comes in the field of view of the camera, his/her face is captured using face detection algorithms. The identity of the person is then determined by a recognition engine through a similarity match with a database of stored images. Current work is focused on further analysis of the video stream in highly controlled settings (such as meetings and office rooms) to provide access to other visual nuances like gaze direction, expressions, hand gestures and attire of all the subjects in the scene, which allow the blind individual to interact socially. Many assistive devices for individuals who are blind provide only audio feedback, which however is inappropriate as visually impaired individuals use their ears to perceive the environment. To overcome this fundamental limitation, we have designed a vibrotactile haptic belt, which consists of a set of 7 vibrators, to be worn by the user around his waist. Information about the direction of an approaching individual is conveyed through the location of vibration, and the distance to the user is encoded in the duration of vibration [2]. Current work is focused on studying fundamental methods of communicating facial expressions through vibrotactile actuators.

-

Integrated Face Localization/Recognition: While building assistive systems for individuals with sensory disabilities, it is important to design user-centric ML systems that utilize the cognitive capabilities of the user to solve the problem at hand, and *support* the user in decision-making, rather than *provide* decisions. In our prototype, we have designed an intuitive system using the wearable camera and the haptic belt to address this issue. When the camera on the user's glasses detects a face in the vicinity of the user, our system localizes the individual into one of 7 regions around the user (such as far left, left, right, far right, etc). This is conveyed to the user through a vibrotactile cue in the haptic belt, where the location of the vibration indicates the person's direction and the duration of vibration indicates the distance between the person and the user. In our experiments, the users found the localization vibrations highly intuitive and automatically turned their head in the direction of the detected face, leading to eye contact between the individuals, which is desirable. Further, this process, in turn, facilitates frontal face recognition (reasonably achieved in controlled settings) and simplifies the problem at hand, rather than tackling the challenges of pose-invariant face recognition.

User-conformal Confidence Measures: An essential component of an ML algorithm in assistive systems is the computation of reliable confidence measures. We adopted a new framework called Conformal Predictions [3] to compute calibrated confidence measures that can be built on top of any existing classification/regression algorithm. The obtained confidence measures are well-calibrated in an online setting, i.e. the frequency of errors, ϵ, made by the system is exactly bounded according to the confidence level, $1 - \epsilon$, defined by the user. Depending on the context, the user can set a confidence threshold, and the system ensures that the number of errors made are exactly bounded by $1 - \epsilon$ fraction of the total number of data samples. For instance, in a party setting, the user can define a moderate level of confidence threshold; and in an official meeting, where accurate recognition is important, a high level of confidence can be set to control the number of errors.

Other ML Contributions: Other ML methods based on online learning, active learning, learning from multiple sources and transfer learning form a significant component of such an assistive orthotic, where the system needs to learn and evolve with the user over time and adapt to changing environments. In the SIA, to ensure appropriate recognition when the blind user encounters a new subject, an online active learning algorithm based on the user-conformal confidence measures, called Generalized Query by Transduction, has been implemented [4]. Also, often in interaction settings, an entire video is recorded and it may be necessary to choose the salient face images in the entire video consisting of thousands of frames. To address this issue, a batch mode active learning algorithm based on an optimization framework has been implemented [5]. In ongoing work, we are consolidating evidence and learning from multiple sources by integrating information from face and speech modalities in a manner such that one modality can supplement/complement the predictions from the other modality. Also, to build person-specific adaptive systems, we are employing transfer learning techniques to be able to adapt models to each user uniquely. Most importantly, our research has shown us how the study and understanding of disabilities leads to a better understanding of human requirements in any human machine interaction [6], thus providing a pathway to building better technologies for the general population.

REFERENCES

[1] S Krishna, D Colbry, J Black, V N Balasubramanian, S Panchanathan, A Systematic Requirements Analysis and Development of an Assistive Device to Enhance the Social Interaction of People Who are Blind or Visually Impaired, ECCV Workshop on Computer Vision Applications for the Visually Impaired, France, Oct 2008.

[2] T McDaniel, S Krishna, V Balasubramanian, D Colbry, S Panchanathan, Using a Haptic Belt to Convey Non-Verbal Communication Cues during Social Interactions to Individuals who are Blind, IEEE International Workshop on Haptic Audio Visual Environments and Games, Ottawa, Canada, Oct 2008.

[3] G Shafer, V Vovk, A Tutorial on Conformal Prediction. J. Mach. Learn. Res. 9, Jun 2008, 371-421.

[4] V Balasubramanian, S Chakraborty, S Panchanathan, Generalized Query by Transduction for Online Active Learning, IEEE ICCV International Workshop on Online Learning for Computer Vision, Japan, Oct 2009.

[5] S. Chakraborty, V. Balasubramanian, S. Panchanathan, Learning from Summaries of Videos: Applying Batch Mode Active Learning to Face-based Biometrics, IEEE CVPR Workshop on Biometrics, June 2010.

[6] S Panchanathan, N C Krishnan, S Krishna, T McDaniel, V N Balasubramanian, Enriched Human-Centered Multimedia Computing through Inspirations from Disabilities and Deficit-Centered Computing Solutions, 3rd International Workshop on Human-Centered Computing, ACM Multimedia, Vancouver, Oct 2008.