Adaptive Technologies for the Developmentally Disabled

Computational Linguistics, Machine Learning, and Computer Vision in Healthcare

Justin Le[1], Pushkin Kachroo[1], Kate Martin[2]

A. Background

Recent efforts to examine social determinants of health within the general public have revealed that people with disabilities experience greater obstacles, specifically in attaining their full health potential. Within the disability population, a subpopulation of individuals with intellectual or development disabilities (ID/DD) appear to experience this to a greater degree. They receive lower rates of preventative care as well as under-diagnosis of chronic conditions such as diabetes and hypertension, when compared to other Americans. There is limited information available about the care received by people with ID/DD, but qualitative data from focus group studies suggest that they face four main barriers in health care: access, knowledge, communication and quality.

In order to better characterize the needs of the developmentally/intellectually disabled in Southern Nevada, families of these individuals were surveyed regarding their quality of life and services currently received. Based on the initial survey data, Dr. Kate Martin, Associate Professor of Family Medicine at UNSOM, has been developing a health assessment and promotion program for Opportunity Village, a not-for-profit organization that serves people with intellectual and developmental disabilities. One of the areas identified as potentially beneficial to the individuals is introduction of novel assistive devices to improve the ability of people with ID/DD to interact with others.

The ideal assistive device would recognize, or learn to recognize, the non-verbal expressions used by an individual and convey the meanings of these expressions in a verbal form to the caretaker of the individual. (We henceforth refer to individuals with ID/DD as "individuals".) We envision that such a device has the following properties: (1) it learns to recognize not only the spatial and temporal characteristics of these expressions but also the semantics that underlie them, and (2) it learns to perform this recognition with comparable accuracy across various individuals whose expressions may vary vastly in characteristic and meaning. In practice, we expect that the first property requires the device to detect and track the areas of the body that an individual employs when communicating, a task that relies greatly on the quality of hardware, environmental conditions, and experimental factors as outlined in the following paragraphs. Additionally, we expect that the first property requires the device to automatically infer the semantics of expressions wherever ambiguity occurs, and such inference requires mathematical treatment. We further expect that the second property can only be accomplished with the guidance of a human who can "train" the device manually until it "learns" to automate the task of recognition. Our mathematical approach will encompass this training process.

The performance of computerized recognition relies largely on the quality of the information given to the system. A sophisticated mechanism for recognition will perform

[1]Dept. of Electrical and Computer Engineering, University of Nevada, Las Vegas
[2]School of Medicine, University of Nevada, Las Vegas

poorly if given data that represents features of the individual that do not closely correlate with the meanings of non-verbal expressions.

For instance, data can be collected regarding the motion of the individual's elbows, but if the individual tends to move elbows in precisely the same manner across all expressions, recording data on this motion may be irrelevant to the task of recognizing the meanings of the expressions. However, if the individual moves elbows in a different manner between expressions, the motion may be useful in the recognition task and should consequently be recorded.

The record of this motion may include its speed, range, continuity, repetitiveness, and other properties. In this study, we call this record an *observation* and each of these properties a *feature*. Performing good recognition requires the selection of good features.

Recording such an observation requires that the elbow be distinguished from other areas of the individual, and likewise, recording observations on other areas, such as the face, hands, or feet, require a mechanism for distinguishing one area from another. In this study, we call this task as *localization*. Recording the features of an observation that involve motion, such as speed or continuity, requires that we follow the area across its range of motion. We call this task *tracking*.

We define an adaptive system to be one that interprets languages across multiple individuals. Ideally, its performance varies negligibly with respect to the changes in the alphabet and grammar across non-verbal languages used by different individuals. This degree of adaptation can be accomplished with the guidance of a caretaker, who provides the system with recordings of an individual's expressions, along with the meanings of those expressions. The system can then learn to associate those expressions and their respective meanings with a specific individual and thereby learn to recognize the language specific to the individual. We call these recordings *examples*, and the learning process we call *supervised training* of the system.

B. Tasks

Aim I: Train the system to recognize ideas and emotions implied by the non-verbal expressions of a person with ID/DD.

Rationale: Communication between a person with ID/DD and a caretaker is limited by the caretaker's understanding of that person's non-verbal expressions. A system that recognizes these expressions can translate them into text that reflects the messages they represent.

Approach:

- Using footage taken by digital camera of a person with ID/DD at distance of one to three feet (depending on the expressions' range of motion), we will apply methods from a software library, OpenCV, to localize the hands and face of an individual. Other areas, such as feet, legs, or torso, will be considered as experiments progress.

- We will program the system to identify gestures performed by each area of interest. Each gesture's features include speed, range, continuity, frequency, orientation, and magnitude of its associated movements, among others to be determined by experiment.
- The gestures typical of the individual form an alphabet of gestures that are unique to that individual. By developing a formal theory of the individual's grammar, as per Aim III, we will model the structure of phrases as combinations of gestures. Each structure will then imply a meaning, and the meanings across phrases form the semantics of the individual's language. The system will be trained to either record the possible structures and the implied meaning of each or to infer a meaning when the structure of a phrase is unfamiliar.
- The performance of the system will be measured by the ratio of correct to incorrect inferences of the meanings of phrases. We refine our selection of features in Step 1 by studying how the performance changes with respect to feature selection. This may involve the removal or adding of features, adjusting the method of recording features, and other changes to be considered as the work progresses.

Aim II: Combine various instances of the system to form a system that adapts to different individuals with ID/DD.

Rationale: Each individual uses a unique alphabet and grammar in constructing non-verbal phrases, and thus, the system must be trained to adapt to multiple non-verbal languages, invariant to these changes in alphabet and grammar.

Approach:

- We will work personally with an individual and a caretaker to provide examples for supervised training of the system.
- The process will be repeated for a number of different individuals, with each individual's system undergoing a separate process of supervised training using examples that are unique to the individual.
- By repeating the process multiple times, we will find which learning algorithms are best suited to the task of learning each individual's unique language. This stage will therefore involve testing of multiple learning algorithms, varying the parameters for each algorithm as we proceed, until the most suitable algorithm is found for each individual.

Aim III: Develop a formal theory of non-verbal languages used by individuals with ID/DD.

Rationale: The task of recognizing non-verbal expressions requires the decomposition of each phrase into well-defined elements. In order to describe the relationships between them, they must have features in common with each other. As an example, verbal communication relies on elements called words, each of which has a part-of-speech. Additionally, words that have the same part-of-speech, such as verbs, each have a conjugation or a tense to indicate a temporal context. Describing the relationship between these elements can accomplished systematically if each element is interpreted as a mathematical object and their relationships as the products of processes that can be formally analyzed.

Approach:

- We will observe the entire range of expressions typical of an individual's daily communication behavior and form an alphabet from the elements that are common across these expressions.
- By identifying the elements in an expression, we will identify the element's role, or the part-of-speech. These non-verbal parts-of-speech may not correspond to the parts-of-speech that are familiar to verbal communication, such as nouns and verbs, but will be chosen to represent the elementary ideas, or perhaps emotions, that they represent for the individual.
- We will design a parser, a computing system that recognizes the relationship between the elements according to their interpreted parts-of-speech and thereby produces the potential semantics associated with the expressions formed by those elements.
- Ambiguous elements will inevitably cause the parser to produce multiple semantic interpretations of the same expression, at which point, we will extend the system to form predictions regarding the correct interpretation by employing a probabilistic algorithm that leverages prior examples of expressions that were recorded and labelled during supervised training of the system.