Technical Disclosure Commons

Defensive Publications Series

March 20, 2017

Multiple Input Modes Without Specific Mode Selection

Thomas Deselaers

Emmanuel Mogenet

Victor Cărbune

MULTIPLE INPUT MODES WITHOUT SPECIFIC MODE SELECTION

ABSTRACT

A system and method are disclosed that enables multiple text input modes for a device without explicitly specifying the desired mode. A machine learning (ML) model is used to handle and interpret the inputs (text, voice, handwritten, etc.). The ML model analyzes sequences of data, and trains itself so that the correct final output sequence is given to the application requiring the text input. A decoder is used to combine the output of the sequence interpretation model with other knowledge sources such as character or word recognition models. Then, a language-model is used in a decoder to obtain the most likely sequence of words or characters given all user inputs. The recognition of the most likely inputs improves and enables automatic mode selection, eliminating explicit segmentation between the modalities.

BACKGROUND

At present, there are a variety of methods for text input, such as typing on a touch screen keyboard, gesture typing on a touch screen keyboard, handwriting on touch screen, voice input and typing on a real hardware keyboard. Switching between these different modes typically requires a conscious effort to select another input method or mode of input, e.g. hitting the "voice" button in a keyboard, or switching to a handwriting input method.

Input methods usually have a way to toggle between one or more of the input modalities explicitly through button toggles (e.g. the keyboard has a voice button on top of it). There are translation apps available to support multiple input modalities by explicitly selecting the mode of input. Similarly, mode toggling is available in handwriting recognition (e.g. switching from drawing to handwriting characters). The latest version of the keyboard may be used to enter text in several languages. The keyboard allows switching automatically between input languages. If multiple languages are enabled, the user may enter text in any of the languages without switching explicitly.

DESCRIPTION

A system and method are disclosed that enable multiple input modes without specifying the desired mode. The system includes a user interacting device which receives inputs from the user in the form of typing on a touch screen keyboard, gesture typing on a touch screen keyboard, handwriting on a touch screen, voice input, and typing on a real hardware keyboard. The method includes multiple input modes without specifying the required mode using a machine learning model as shown in FIG. 1. The machine learning model is selected in a way that it copes with analyzing sequences of data and trains it in a way that the correct final output sequence is given to the application, without any explicit segmentation between the modalities. For example, recurrent neural networks, hidden Markov models, or finite state transducer (FST) graphs may be used. A recurrent neural network allows end-to-end training, given appropriate aforementioned input (sequence of points, [x, y, t, p], sequence of voice cues, sequence of points) and gives the final output sequence that is expected. The training data may be generated based on existing usage from users generated by the machine learning algorithm. The explicit button toggling is removed and a continuous stream of points is considered to be the input.

A decoder is used to combine the output of the sequence interpretation model with other knowledge sources such as word- or character-based language models. A language-model perceives the sequence of words or characters which is likely in a given language. This also improves and enables automatic mode selection by improving the chances for the recognition of more likely inputs. The user may seamlessly switch between input modes.

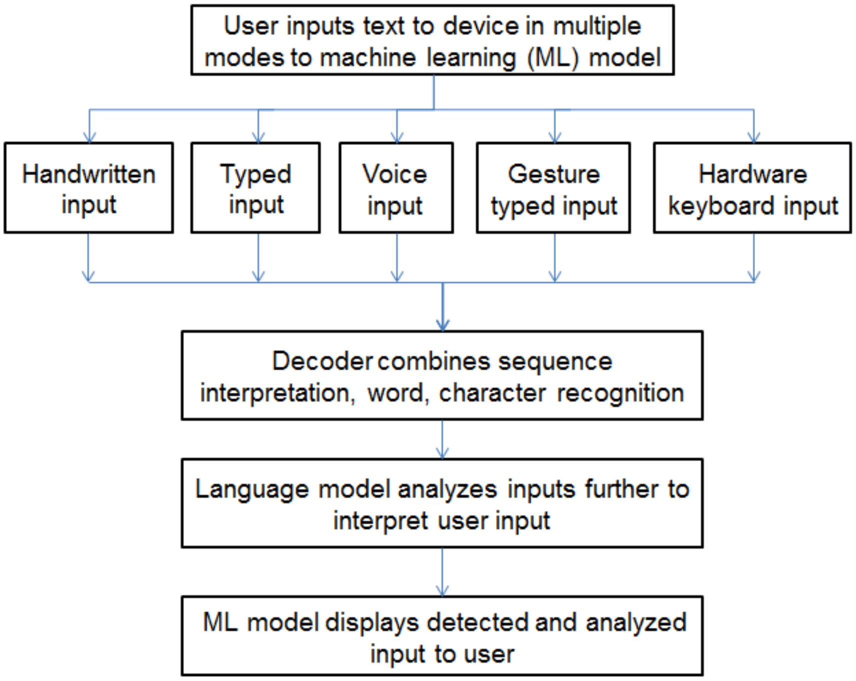
User inputs text to device in multiple modes to machine learning (ML) model
Handwritten input
Typed input
Voice input
Gesture typed input
Hardware keyboard input
Decoder combines sequence interpretation, word, character recognition
Language model analyzes inputs further to interpret user input
ML model displays detected and analyzed input to user

FIG. 1: The method to enable multiple input modes

For example, if a user is typing something while someone else is talking in the background, it's important that the model is able to determine that these two sources of text are not supposed to be combined. The use of speaker identification makes it less likely to combine text spoken by others into the text input.

Examples of the properties taken into consideration are:

i. If the user has touch points on the screen in the keyboard area, it's quite likely that the user wants to type.

ii. Voice can be overheard from the environment (speaker identification may be a useful feature here).

iii. Typing using a physical keyboard is considered to be a very strong signal - in some cases, e.g. when the keyboard is folded entirely away, the environment might be noisy; Hence taking into account the working state of the keyboard is necessary.

iv. Distinguishing handwriting and typing on a touch screen may be done using a similar framework.

Another outcome may be that the user may seamlessly switch between typing and handwriting by just handwriting on top of the keyboard: In the image below, a keyboard is displayed with both ways to enter the word “Hello”, in blue, the gesture typing input is shown, and in red the handwritten input.

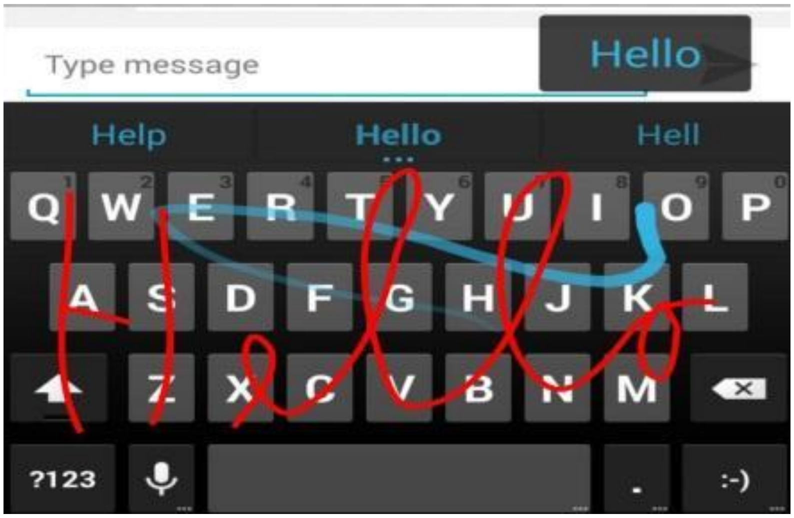

FIG. 2: Illustration of simultaneous gesture typing and handwritten inputs

If the user enters a text by voice input and then wants to enter a punctuation mark the user may just type the punctuation mark when he is giving a voice input:

spoken: "I am leaving home now"

type:","

spoken: "will be at your place in 20 minutes"

type: "."

Additionally, feature extraction or normalization may be integrated to convert the input to a common higher-level representation.

Alternatively, the system and method may exclude the use of machine learning models and instead include a set of heuristics based on predetermined signals that would allow some of the interaction described. Also, the various machine learning models may be run in parallel and their outputs finally combined and implemented.

The system and method disclosed would allow better, seamless interactions and more complex interaction models such as using different areas of the screen for different input types, than currently possible without an explicit mode switch.