Using User Location and User Interest To Determine Content In Voice Activated Computing Systems

John D. Lanza
Foley & Lardner LLP
Voice activated computing systems provide a user with content or services in response to voice commands spoken by the user. Such systems can capture voice commands from a user, process the voice commands to determine requests and keywords in the voice commands, and provide the user with content or services related to the determined requests and keywords.

As discussed herein, a voice activated computing system processes the voice commands and generate responses that take into consideration a user location as well as user interests. The system processes the voice commands to identify requests and trigger keywords. For example, the voice command received from the user can include the statement: “I would like to go out to a restaurant.” The system can identify, for example, a request for a “restaurant” and trigger keywords “to go” or “to go out to.” Based on the request and the trigger keywords, the system can communicate with one or more service providers to provide the user with a list of options for restaurants. However, the system not only takes into consideration the request and the trigger keyword identified in the voice command, but also considers the user’s current location. For example, if the user provided a voice command using a mobile device, which includes a global positioning system (GPS), the system can request the mobile device to provide the current location of the user. Furthermore, the system can take into consideration the user’s interest. For example, the system can analyze historical user data to determine user interests such as, favorite cuisine, dietary restrictions, etc. Based on the location and the user interest, in addition to the request and the trigger keywords, the system can determine which restaurant options to present to the user.
The system also determines a bid value that a merchant would be willing to pay to have its product or service presented to the user. For example, the system can determine a bid value for each restaurant service provider would be willing to pay to present their restaurant to the user. The bid value can be based on the distance of the restaurant from the current location of the user, as well as based on the user’s interest.

Figure 1 shows an example voice activated computing system. The system includes a voice assistant device, a service provider, a content provider, and a data processing system communicating over a network. The voice assistant device can be a device that accepts voice commands, and provides audio or visual output. The voice assistant can include one or more mics and cameras, such that voice commands received by the user are converted into corresponding audio signals. The voice assistant can send the audio signals to the data
processing system and the service provider. The voice assistant device also can receive data such as audio signals or video signals from the data processing system or the service provider. The voice assistant device also can include audio speakers that can convert the audio signals received from the data processing system or the service provider into sound.

The data processing system can process voice commands received from the voice assistant device. The data processing system includes a natural language processor, an audio signal generator, a task predictor, and a content selector. The natural language processor is capable of processing voice commands included in the audio signals received from the voice assistant device. The natural language processor can convert the audio signals into recognized text by comparing the audio signals against a stored, representative set of audio waveforms, and choosing the closest matches. The representative waveforms are generated across a large set of users, and can be augmented with speech samples. After the audio signals are converted into recognized text, the natural language processor can match the text to words that are associated, for example via training across users or through manual specification, with actions that the data processing system can serve. Basically, the natural language processor identifies requests and trigger words in the converted text, based on which the natural language processor can determine the content and actions to be carried out. The task predictor can predict tasks or actions based on the converted text, and in particular by identifying requests and trigger keywords in the converted text. The task predictor also can predict the most likely sequence in which the tasks would be executed. The content selector can select content, such as services to be offered to the user based on the actions identified by the task predictor. In addition, the content selector also can alter the sequence or the order in which the actions related to the services offered to the user are executed. The audio signal generator can generate audio signals based on the services
selected by the content selector. The audio signals can be representative of voice responses or voice instructions provided to the user in response to the voice commands.

The service provider can provide one or more services to the user. For example, the service provider can be a taxi or car sharing service provider, dining or reservation service provider, and the like. The service provider can communicate with the voice assistant device independently of the data processing system and provide the user the ability to request a ride, do a dinner reservation, or avail of other services provided by the service provider. The service provider can also include a natural language processor, similar to the one discussed above in relation to the data processing system, to convert user voice commands into text, and identify requests and keywords to determine the services requested by the user. The content provider can provide sponsored content items related to the requests and trigger keywords identified in the voice command. The content items provided by the content provider can be provided to the user in addition to the response to the requests made by the user in the voice command.

Referring again to the voice command example mentioned above, the user can speak the voice command “I would like to go out to a restaurant,” to the voice assistant device. The mics at the voice assistant device can convert the voice commands into corresponding audio signals, which are be transmitted by the voice assistant device to the data processing system over the network.

At the data processing system, the natural language processor processes the audio signal received form the voice assistant device and identify a request for a “restaurant.” The natural language processor also can identify a trigger keyword “go” or “to go out to,” which can indicate a need for transportation. Even though the user’s voice command does not directly express an intent for transportation, the trigger keyword indicates that transportation may be needed.
The task predictor, based on the requests for “restaurant” and on the trigger keywords, can determine a most likely sequence of actions related to the voice command. For example, the task predictor can determine a task sequence that includes restaurant reservations, booking a ride to the restaurant, and booking a ride from the restaurant.

The content selector can determine the names of the restaurants to be presented to the user. As mentioned above, the content selector can determine which of the many names of restaurants provided by the service providers to be presented to the user based on a current location of the user and the user’s interests.

The current location of the user can be a location of a user device that was used to transmit the voice command to the system. For example, if the user used a mobile device to send a voice command to the system, the location of the mobile device can be used as a representative location of the user. In some implementations, the current location of the user can be a location of a second user device that is associated with the user device that transmitted the voice command to the data processing system. For example, multiple devices can be associated with one another through a unique login, a family account, or other grouping. The system can use the location of the user device from which the voice command originated or the location of any other user device associated with the originating user device. In some implementations, the current location of the user can be determined based on location information communicated from the user device to the system for the purpose of locating the device (such as geographic coordinates communicated by the device when the user has granted location-based information to be communicated). In some implementations, the location can be determined based on network information associated with the user device. This network information can include, for example, an IP address of the user device, an IP address of a carrier associated with the user device, or a
network identification name (e.g., an SSID of a nearby wireless network). The location of the user device may also be manually set by the user. For example, for a home voice assistant device, the user can provide a location to the system to associate with the home voice assistant device. In some implementations, the location of the user device can dynamically update. The location of the user device may update when polled, every minute, every 30 minutes, every hour, or every several hours. The location can be determined using a GPS module within the client device or cellular tower location information.

The user interest can be determined, for example, based on prior requests or voice commands associated with the user. For example, the interest can be based on past search histories provided by the user to the system. The interests can also be based on prior web browsing histories, selection of products or services, selection of sponsored content, or actions associated with the request. For example, the request can trigger a thread that includes multiple actions. The actions of the thread can occur in a set, sequential order. For a thread including three actions, for the second action, the prior request can be a request associated with the first action. The method can also include providing an indication of the third action prior to an occurrence of the second action. For example, the content selector component can receive the indication of the third action prior to occurrence of the first action or prior to occurrence of the second action. The indication of the third action can be generated by the task predictor. The task predictor can generate the indication based on the request or the prior request. For example, for the request, "Ok, what time does the movie start," the prediction component can determine there is a high probability the user will provide the request of "Ok, what restaurants are near the movie theater." The indication of the third action can be an indication that the user will request restaurant locations near the movie theater.
The content selector can identify a set of candidate content items based on the current location of the user and the user interest, where the content items can include, with reference to the example voice command discussed above, a list of restaurants. The content selector can further determine a prediction value for each of the candidate content items. Each prediction value can be based on the distance between the current location of the user and the geographical location of the respective one of the candidate digital components. The prediction value can also be based on the prior request. The prediction value can also be based on an interest of the user associated with the user device. Each prediction value can indicate a probability that a corresponding restaurant will be selected by a user. For example, using both the location of the user and the interest of the user, the system may identify a likelihood that a restaurant will be selected by the user. In the foregoing example, the system may determine that the user is interested in Italian cuisine. Therefore, the system may determine that a restaurant serving Italian cuisine may be more likely to be selected by the user than a restaurant serving other types of cuisines. Also in the foregoing example, a restaurant serving Italian cuisine that is near the user may be more likely to be selected by the user than a restaurant serving Italian cuisine that is farther from the user.

A candidate item or a list of candidate items to be presented to the user can be selected based on the prediction values calculated for each of the candidate content items. The selection can be made by ranking each of the candidate content items based on the prediction values and selecting the content item with the highest prediction value, or selecting a predetermined number of top ranking content items.

In some implementations, the selection process can also include an auction process. As mentioned above, the system can determine a bid value that a merchant would be willing to pay
to have its product or service presented to the user. The bid may include a value or other indication that indicates a dollar or other currency amount that an originator, such as an content provider, of a content item is willing to pay to have the content item presented to the user. For example, the system may receive or otherwise obtain the bid form the service providers. In these implementations, the system may determine an expected revenue return value based on the bid value and the prediction corresponding to the content item and may identify a content item based on the expected revenue return value. In other words, the system may identify content items in a manner that is expected to maximize fees from entities such as content providers that wish to promote their goods and/or services via content items.

As an example, the expected revenue return value can be given by the equation: \( E = p \times b \) (Equation 1), where \( E \) is the expected revenue return value (obtained by the product of \( p \) and \( b \)); \( p \) is the probability that the content item will be selected by the user; and \( b \) is the bid. Thus, the expected revenue return value may be based on a combination of the prediction/probability of being selected (which itself may be based on the location and the interest of the user) and the bid. The bid may be sufficiently high to overcome a prediction with a low probability and vice versa. \( p \) can be a function of various input predictive factors, \( x \), as given by the equation: \( p = f(x) \) (Equation 2), where \( x \) may include, for example, distance from physical location, interest-based information, historical performance of content items (such as whether a content item was selected by the user), and/or other predictive factors described below. Based on one or more of the predictive factors, a probability that a content item will be selected by the user may be generated. The bid can be changed as a function of the distance, \( d \), as given by the equation: \( b = g(d) \) (Equation 3). In other words, the system may associate different bids of a content item for different distances. For example, an entity such as an content provider may be willing to pay
more for a content item delivered to the user via mobile communication when the user is at a
closer distance from a physical location than a farther distance. The entity may submit—and the
system may receive—different bids for different distances.

Based on the bidding process, and the user location and user interests discussed above,
the content selector can select one or more restaurants, from the candidate restaurants, to be
presented to the user. The system can send an audio file that includes the selected restaurant, or
the selected list of restaurants, to the voice assistant device, which can play the audio to the user.
The system may also send the selected restaurants, or the selected list of restaurants to a mobile
device of the user, where the restaurant names can be either delivered aurally or displayed on a
display screen of the mobile device.

The system can select a type of display content of the digital component based on the
present location associated with the client device and the geographical location of the digital
component. For example, if the client device is within a predetermined threshold distance of the
geographical location associated with the selected digital component, the type of display content
can be walking directions to the geographical location. If the client device is outside of the
predetermined threshold distance and the digital component is associated with a restaurant, the
type of display content can be an action to set a reservation restaurant.

The system may also alter the task sequence determined by the task predictor based on
the user location or the user interests. For example, if the user requests go to a restaurant and to
shop for shoes, the task predictor may initially determine a task sequence as first going to a
restaurant and then shopping for shoes. However, if it is determined that the current location of
the user is near a merchant that offers user’s favorite shoes, the system may provide the user with
the merchant’s information prior to providing the user with restaurant information.
Abstract

This document describes a technique for processing voice commands in a voice activated computing system. In particular, the system processes the voice commands and generates responses that take into consideration a user location as well as user interests. The system processes the voice commands to identify requests and trigger keywords. Based on the request and the trigger keywords, the system can communicate with one or more service providers to determine a list of candidate content items. The system then selects one or a list of content items from the list of candidate content items based on the user location and the user interest. The system can also use the user location and the user interests as factors in a bidding process in which content providers compete to provide their content items in response to user requests. The system may also modify a sequence of tasks based on the user’s current location or the user interests.