Technical Disclosure Commons

Defensive Publications Series

October 10, 2019

SONG SENTIMENT RECOGNITION BASED ON IMPLICIT USER FEEDBACK

Joseph Edwin Johnson Jr.

Shiblee Hasan

Chris Hluchan

SONG SENTIMENT RECOGNITION BASED ON IMPLICIT USER FEEDBACK

ABSTRACT

Computing devices utilize user feedback (e.g., explicit and/or implicit feedback) to determine a user's sentiment towards a song or music played by the computing device or another speaker within a given environment. The computing device may determine the user's sentiment, emotion, reaction, or opinion towards a song and utilize that information when selecting music, tunes, or songs to include in a playlist or otherwise output to the user. In one example, to determine the user's sentiment towards the song, the computing device executes a sentiment recognition algorithm that utilizes implicit feedback, such as a user singing along with the song, the user moving in time with the beat of the song, etc., which may be captured by one or more sensors of the computing device. In some instances, the computing device may also use explicit feedback provided by the user, such as selecting a "like" or a "dislike" button to determine the user's sentiment towards the song. If the user enjoys the song, the computing device may recommend, suggest, or output the song or similar songs (e.g., similar artists, genres, etc.) for playback, for example, when selecting songs to be included in an autogenerated playlist.

DESCRIPTION

Techniques are described that enable a computing device to determine a user's sentiment towards a song and to select songs for audio playback based on whether the user enjoys the song. In the example of FIG. 1, computing system 1 includes computing device 10 and computing device 20. In some examples, computing system 1 may not include computing device 20 and/or may include additional computing devices.

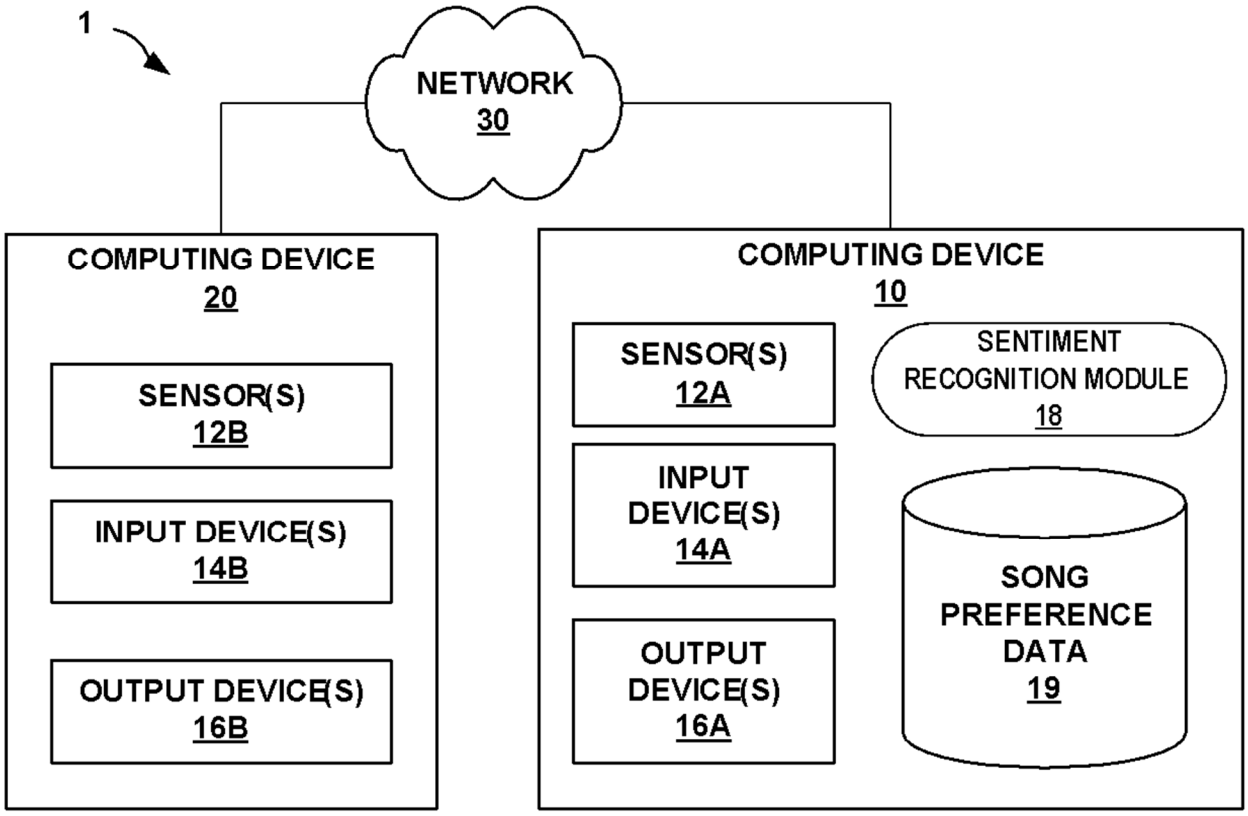

FIG. 1

In some examples, computing devices 10 and 20 are communicatively coupled to one another via network 30. Network 30 represents any public or private communications network, for instance, cellular, Wi-Fi, and/or other types of networks, for transmitting data between computing systems, servers, and computing devices. Network 30 may include one or more network hubs, network switches, network routers, or any other network equipment, that are operatively inter-coupled and thereby provide for the exchange of information between computing devices 10 and 20. Computing devices 10 and 20 may transmit and receive data across network 30 using any suitable communication techniques, such as via a via a cellular network (e.g., GSM, CDMA, LTE, etc.), via WIFI®, via BLUETOOTH® or any other wireless and/or wired communication techniques.

Computing devices 10 and 20 may include any type of computing device capable of communicating with another device over a network, such as a smart phone, a desktop computer, a laptop computer, a tablet computer, a smart watch, a smart speaker, a server, a vehicle infotainment system, or any other type of computing device or system. Computing devices 10 and 20 include one or more processors. Examples of processors include, but are not limited to, digital signal processors (DSPs), general purpose microprocessors, application specific integrated circuits (ASICs), field programmable logic arrays (FPGAs), or other equivalent integrated or discrete logic circuitry. In the example of FIG. 1, one or more processors execute the functionality of sentiment recognition module 18 of computing devices 10.

Computing device 10 may include one or more sensors 12A, one or more input devices 14A, one or more output devices 16A, and sentiment recognition module 18. Similarly, computing device 20 may include one or more sensors 12B, one or more input devices 14B, and one or more output devices 16B. Examples of sensors 12A and 12B (collectively, sensors 12) include motion sensors (e.g., accelerometers, gyroscopes, etc.), image sensors, radar sensors, among others. Examples of input devices 14A and 14B (collectively, input devices 14) include microphones; touchscreens; physical buttons, switches, or controls; etc. Examples of output devices 16A and 16B (collectively, output devices 16) include speakers, display screens, etc.

Computing device 10 may output audio output data, such as a song, for playback via an output device 16A (e.g., a speaker) of computing device 10. In another example, computing device 10 may transmit the song to another computing device 20, such as a vehicle infotainment device or a countertop speaker, to cause an output device 16B of computing device 20 to play the song. For purposes of illustration only, the audio output data is described as a song. However,

in some examples, the audio output data may include a song, podcast, audiobook, or other audio data.

Sentiment recognition algorithm 18 determines a user's sentiment towards a song based on data received from one or more sensors 12 and/or one or more input devices 14. Throughout the disclosure, examples are described where a computing device and/or a computing system analyzes data associated with a computing device and a user of a computing device, only if the computing device receives permission from the user of the computing device to analyze the information. For example, in situations discussed below, before a computing device or computing system can collect or may make use of information associated with a user, the user may be provided with an opportunity to provide input to control whether programs or features of the computing device and/or computing system can collect and make use of user information (e.g., information about a user's current location, current speed, etc.), or to dictate whether and/or how to the device and/or system may receive content that may be relevant to the user.

In addition, certain information may be treated in one or more ways before it is stored or used by the computing device and/or computing system, so that personally-identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined about the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user may have control over how information is collected about the user and used by the computing device and computing system.

One or more sensors 12 and/or one or more input devices 14 may generate user reaction data as a user of computing device 10 reacts to a song played by computing device 10 or

computing device 20. The user reaction data may include, as non-limiting examples, image data, motion data, audio input data, or touch input data captured by one or more of sensors 12 and/or input device 14. For example, when a user hears a song that he or she enjoys, the user may react to the song by singing along, tapping his or her feet or hands, dancing, or a combination thereof. In some examples, sensors 12 capture one or more images of the user singing, dancing, tapping his or her feet or hands, or otherwise moving while the song is being played. In such examples, sensors 12 generate image data indicative of the one or more images. In yet another example, sensors 12 may detect movement of computing device 10 (e.g., as the user dances, or taps his or her feet or hands) and may generate motion data indicative of the movement of computing device 10.

Additionally or alternatively, the user may react to a song her or she enjoys by adjusting (e.g., increasing) the volume of the song (e.g., via a physical or touch-sensitive button or via a voice command). In such instances, input device 14 may generate touch input data indicating a user input to change (e.g., increase or decrease) the volume of the song using touch input. Further, a user may sing along with the song being played or may provide a voice command (i.e., an audio command) to adjust the volume. In such instances, input device 14 may detect the singing or voice command and generate audio input data in response to detecting the audio.

Sensors 12 and/or input devices 14 may transmit the user reaction data to sentiment recognition module 18. Sentiment recognition module 18 may determine whether the user enjoys the song based on the user reaction data. For example, sentiment recognition module 18 may determine whether the user enjoys the song by determining whether the user sings along with the song. In some examples, sentiment recognition module 18 determines whether the user sings along with the song based on characteristics of the audio input data. Examples of

characteristics include frequency spectrum, quefrency cepstrum (MFCC's), pitch, volume, etc. In one example, sentiment recognition module 18 compares the characteristics of the audio input data to a database or library of known singing examples using a dynamic time warping (DTW) algorithm to determine whether the user is signing. In another example, sentiment recognition module 18 applies raw audio input data and/or the characteristics of the audio input data to a machine trained model (e.g., a convolutional neural network) to determine whether the user is singing. For example, sentiment recognition module 18 may concatenate the characteristics of the audio input data for various samples of the audio input data into a matrix and input the matrix into a machine trained model. In one example, the output of the machine trained model indicates whether the user is singing or not singing.

Additionally or alternatively, sentiment recognition module 18 may analyze the audio input data to determine whether the audio input data includes the user voicing words that are included in the song. For example, sentiment recognition module 18 may perform voice recognition on the audio input data and/or compare the audio input data to a fingerprint of the song to differentiate between words said or sung by the user and words played by output device 16.

Sentiment recognition module 18 may not only determine if the user is saying words that are included in the song, but may also determine whether the user is saying such words at approximately the same time as the words occur in the song (e.g., within half of a second, one second, etc.). According to some examples, sentiment recognition module 18 may determine whether the user has said at least a threshold number of words of the song and/or whether the words sung by the user are part of the refrain or other portion of the song. In some instances, sentiment recognition module 18 determines that the user enjoys the song in response to

determining that the user is singing along with the song (and optionally, that the user says the words at approximately the same time as the words occur in the song, the user speaks at least a threshold number of words, and/or that the words said by the user are included in a particular portion of the song).

In instances where the user reaction data includes image data, to determine if the user is singing, sentiment recognition module 18 may analyze the image data (e.g., by performing visual speech recognition to detect movement of the user's lips) in addition to or instead of the audio data. For example, sentiment recognition module 18 may determine whether the user enjoys the song by determining whether the user is tapping his or her hands or feet, bobbing his or her head, dancing, or otherwise moving while the song is playing (e.g., moving to the beat or the rhythm of the song. Sentiment recognition module 18 may analyze a plurality of images captured over time to identity a portion of the user's body and determine whether that portion of the user's body moves with approximately the same rhythm as the song.

In examples where the user reaction data includes motion data, sentiment recognition module 18 may determine whether the computing device 10 (and thereby the user of computing device 10) moves with substantially the same rhythm as the song. In some scenarios, sentiment recognition module 18 may determine the user's body moves with approximately the same rhythm as the song when the user taps his or her foot or hand and the time between taps is approximately equal to the time between beats or pulses of the song (e.g., within one quarter of a second, one half of a second, etc.). Sentiment recognition module 18 may determine that the user enjoys the song in response to determining that the user is moving to the beat or rhythm of the song while the song is playing.

Sentiment recognition module 18 may also determine whether the user enjoys the song based on user reaction data that includes touch input data or audio input data indicating a command to change (e.g., increase or decrease) the volume of the song. In some examples, sentiment recognition module 18 determines the user enjoys the song if the user increases the volume of the song at least a threshold number of times (e.g., at least one out of every three times the song is played, every ten times the song is played, etc.). Sentiment recognition module 18 may determine that the user does not enjoy the song based on receiving a user input to reduce to volume of the song or skip the song. In one example, a user may decrease the volume of a song if he or she is engaged in a conversation, such that decreasing the volume of the song once may not indicate that the user does not enjoy the song. However, in some examples, if the user repeatedly (e.g., at least a threshold number of times) decreases the volume of a song or skips the song, sentiment recognition module 18A may determine that the user does not enjoy the song. In this way, sentiment recognition module 18 may determine whether the user enjoys the song by determining whether the user sings along with the song, moves along with the song, and/or increases a volume of the song.

According to some examples, sentiment recognition module 18 determines whether different individuals within the same environment enjoy the song. For example, computing device 10 may be located in a vehicle, karaoke bar, or other environment that includes several people, some of whom may be singing along with the song and/or moving along with the song. In some examples, sentiment recognition module 18 may identify the individuals within the environment based on the user reaction data. For example, sentiment recognition module 18 may identify the users or individuals in the environment by performing voice recognition and/or

facial recognition. In such examples, sentiment recognition module 18 may determine whether any individuals enjoy the song, as described above, and if so, determine who enjoys the song.

Responsive to determining that a user enjoys the song, sentiment recognition module 18 stores data indicating the user enjoyed the song to song preference data 19. Song preference data 19 may include data indicating whether the user enjoys each song of a plurality of different songs. Song preference data may include data indicating the artist, title, genre, and/or musical characteristics of each song.

Sentiment recognition module 18 may determine or select songs that the user may enjoy based on song preference data 19. For example, sentiment recognition module 18 may select songs that sentiment recognition module 18 has determined the user enjoys, songs by the same artist as a song the user enjoys, songs within the same genre as a song the user enjoys, songs with similar musical qualities as a song the user enjoys, or a combination thereof. Sentiment recognition module 18 may output the selected songs via output device 16A or may cause computing device 20 to output the song via output device 16B. In some examples, sentiment recognition module 18 causes computing device 20 to output the song by transmitting the songs to computing device 20, transmitting a title of the song to computing device 20, etc.

It is noted that the techniques of this disclosure may be combined with any other suitable technique or combination of techniques. As one example, the techniques of this disclosure may be combined with the techniques described in U.S. Patent Publication No. 2019/042,647 entitled “Systems and methods for creation of a listening log and music library” by Hyun Oh Oh. As another example, the techniques of this disclosure may be combined with the techniques described by U.S. Patent Publication No. 2014/232,516 entitled “Contextual and presence sensing to operate electronic device” by Stivoric, et al.