

METHOD AND SYSTEM FOR REAL TIME RECOMMENDATION

TECHNICAL FIELD

[0001] The present invention relates generally to customer service technologies, and more particularly to a system and method for providing real-time recommendations to customer service agents based on voice data.

BACKGROUND

[0002] In today's fast-paced customer service environment, agents are often required to assist clients in a quick and efficient manner. To ensure satisfactory customer experiences, agents must understand customer queries and provide accurate, timely responses. Traditionally, customer service agents have relied on personal experience, written manuals, and internal knowledge databases to answer queries. However, the process of listening to a customer, comprehending their issue, and manually searching for relevant information can be time-consuming and prone to errors.

[0003] Moreover, as the volume of customer interactions has increased with the growth of various communication channels like voice calls, chats, and social media, the pressure on agents to provide instant and accurate information has escalated. Consequently, there is a growing need for automated systems that can assist agents in real-time, ensuring a consistent quality of service regardless of the agent's experience or familiarity with a particular issue.

[0004] Speech recognition technologies, like Automatic Speech Recognition (ASR), have been developed to transcribe voice data. While they have advanced considerably, these technologies primarily focus on transcription accuracy and often do not cater to real-time requirements or specific applications like assisting customer service agents.

[0005] Furthermore, the prior art does not sufficiently address the dynamic nature of customer interactions. There is a lack of systems that can predict, with accuracy, when a customer might stop speaking and present information to the agent in an easily digestible format, allowing them to respond promptly.

[0006] Thus, there exists a need in the art for an integrated system that leverages real-time transcription of voice data, predicts optimal end times based on speech rates, and

retrieves relevant content for agents to ensure a more efficient and enhanced customer service experience.

[0007] The present invention seeks to address these challenges by introducing a real-time recommendation method and system for customer service agents based on processed voice data.

SUMMARY

[0008] Embodiments of the present disclosure may include a real-time recommendation method for customer service agent voice data, including receiving a customer's voice input. Embodiments may also include processing the voice input to determine an optimal end time based on speech rate. Embodiments may also include transcribing the voice input using an ASR model. Embodiments may also include retrieving relevant content based on the transcribed voice input. Embodiments may also include displaying the content to a customer service agent in a manner designed to enable rapid reading and response.

[0009] Embodiments may also include processing the voice input includes using a sliding window time series model to estimate current speech rate and predict voice end time. In some embodiments, the ASR model transcription duration may be optimized by combining cache mechanisms and NLP models to enhance the transcription's real-time retrieval and accuracy.

[0010] In some embodiments, the method may include the steps of preprocessing the ASR transcribed text to obtain cleaned text. Embodiments may also include matching the preprocessed text with known vocabulary and phrases in a cache to calculate a matching score. Embodiments may also include further processing the ASR text with an NLP model to obtain advanced semantic representations for enhanced content retrieval. In some embodiments, the matching score may be computed as a ratio of matching vocabulary words from the ASR text to the total vocabulary in the cache. In some embodiments, the displayed content to the customer service agent uses a highlighting mechanism and a keyword mode for faster comprehension and improved response times.

[0011] Embodiments of the present disclosure may also include a real-time recommendation system for customer service agent voice data, including an input module configured to receive and process voice input from a customer. Embodiments may also include an ASR module for transcribing the voice input. Embodiments may also include a

content retrieval module interfaced with a knowledge base, designed to fetch relevant content based on the transcribed input. Embodiments may also include and a display module to present the content to a customer service agent in a format designed for rapid comprehension.

[0012] In some embodiments, the input module uses a sliding window time series model to optimize the end time of voice input based on an estimated speech rate. In some embodiments, the ASR module employs a combined cache and NLP model system to optimize the transcription duration, enhancing both the real-time nature and the accuracy of transcription content retrieval. In some embodiments, the display module employs a user interface that highlights relevant content and keywords to aid the customer service agent in quickly comprehending and responding to the customer's query.

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] FIG. 1 is a flowchart illustrating a real-time recommendation method, according to some embodiments of the present disclosure.

[0014] FIG. 2 is a flowchart further illustrating the real-time recommendation method from FIG. 1, according to some embodiments of the present disclosure.

[0015] FIG. 3 is a block diagram illustrating a real-time recommendation system, according to some embodiments of the present disclosure.

DETAILED DESCRIPTION

[0016] FIG. 1 is a flowchart that describes a real-time recommendation method, according to some embodiments of the present disclosure. In some embodiments, at 110, the real-time recommendation method may include receiving a customer's voice input. At 120, the real-time recommendation method may include processing the voice input to determine an optimal end time based on speech rate. At 130, the real-time recommendation method may include transcribing the voice input using an ASR model. At 140, the real-time recommendation method may include retrieving relevant content based on the transcribed voice input. At 150, the real-time recommendation method may include displaying the content to a customer service agent in a manner designed to enable rapid reading and response. In some embodiments, processing the voice input may include using a sliding window time series model to estimate current speech rate and predict voice end time. In some

embodiments, the displayed content to the customer service agent may use a highlighting mechanism and a keyword mode for faster comprehension and improved response times.

[0017] In one embodiment of the invention, the system utilizes a sliding window time series model to efficiently process voice input and predict voice end times. The underlying concept revolves around processing segmented portions of the continuous voice input stream to establish an ongoing estimation of the speech rate, which is then employed to predict when the current speaker might pause or finish speaking.

[0018] To begin, as the voice data is streamed into the system, it is segmented into small, overlapping windows of fixed duration. For instance, if the duration chosen is five hundred milliseconds, the model might process voice data from 0-500 milliseconds, then 200-700 milliseconds, then 400-900 milliseconds, and so on, always overlapping the previous window. Each of these windows undergoes a transcription process, converting the voice data into text. The transcribed text is then evaluated for word count or syllable count, which gives an estimate of the speech rate for that particular window. By aggregating the speech rates from multiple overlapping windows, the system can establish a more stable and accurate ongoing estimation of the current speech rate.

[0019] With an accurate estimate of the current speech rate in hand, the system is equipped to make predictions about the speaker's behavior. By analyzing historical data of similar speech patterns and their corresponding end times, combined with the current speech rate, the system can generate a probability distribution over potential voice end times. For instance, if a speaker's current speech rate indicates they are speaking quickly and summarizing, the system might predict an imminent pause or conclusion. On the other hand, a slower, more deliberate speech rate might indicate the speaker is beginning a detailed explanation, and the system would adjust its prediction accordingly. By continuously updating its predictions as new voice data windows are processed, the system can provide real-time estimations for when the speaker might finish their current statement, allowing customer service agents to prepare their responses accordingly.

[0020] In a one embodiment of the invention, to further optimize the interface for customer service agents, the system integrates a highlighting mechanism and a keyword mode, both designed to enable agents to understand the gist of the voice input rapidly and respond more efficiently. The primary goal is to minimize the cognitive load on the agent,

allowing them to comprehend the core concerns of the caller quickly and formulate an appropriate and timely response.

[0021] Upon transcription of the voice input, the highlighting mechanism springs into action. Using NLP techniques, the system identifies and emphasizes key phrases or sentences that are likely to represent the main concerns or queries of the caller. These emphasized segments are visually highlighted in the displayed transcription, making them stand out to the agent at a glance. For instance, a sentence like, "I can't access my account," might be bolded or colored differently to signal its importance. This visual emphasis guides the agent's attention to the most critical parts of the transcription, enabling faster comprehension of the caller's intent.

[0022] In tandem with the highlighting mechanism, the keyword mode further aids the agent's quick understanding. Rather than presenting the full transcription, the system, in keyword mode, extracts and displays a list of critical keywords or phrases from the voice input. These keywords are chosen based on their relevance to common customer queries, concerns, or potential issues. For example, words or phrases like "password reset," "billing error," or "refund" would be displayed prominently. The agent can then quickly scan these keywords to get an immediate sense of the conversation's direction without having to parse through the entire transcription. By integrating these two features – the highlighting mechanism and the keyword mode – the system ensures that agents can grasp the essence of the conversation rapidly, leading to more efficient responses and improved customer interactions.

[0023] FIG. 2 is a flowchart that further describes the real-time recommendation method from FIG. 1, according to some embodiments of the present disclosure. In some embodiments, the ASR model transcription duration may be optimized by combining cache mechanisms and NLP models to enhance the transcription's real-time retrieval and accuracy. In some embodiments, at 210, the method may include preprocessing the ASR transcribed text to obtain cleaned text. At 220, the method may include matching the preprocessed text with known vocabulary and phrases in a cache to calculate a matching score. At 230, the method may include further processing the ASR text with an NLP model to obtain advanced semantic representations for enhanced content retrieval. The steps of, the real-time recommendation method may include 210 to 230. In some embodiments, the matching score

may be computed as a ratio of matching vocabulary words from the ASR text to the total vocabulary in the cache.

[0024] In a particular embodiment of the invention, the duration taken for transcription by the ASR model is significantly optimized by incorporating a unique combination of cache mechanisms and NLP models. This approach aims to maximize the efficiency and accuracy of real-time transcription retrieval, thereby reducing any lag that might disrupt the fluidity of the conversation between the customer and the service agent.

[0025] The cache mechanism operates by storing frequently used words, phrases, or even whole sentence structures that are commonly encountered in customer interactions. Over time, the system learns and anticipates which segments of speech are likely to reoccur. When a customer's voice input is detected, the ASR first checks the cache for matching segments before reverting to full transcription. This immediate retrieval of common phrases from the cache accelerates the transcription process and reduces the system's workload. Additionally, as these cached segments have been pre-validated, they offer a higher accuracy rate than real-time transcription, ensuring that the displayed text is more precise.

[0026] Parallely, the system employs NLP models to further enhance transcription accuracy. These models analyze the context in which words or phrases are spoken, predicting the most likely following words or sentences. For instance, after the phrase "I forgot my," the NLP model might anticipate words like "password," "username," or "account number." By integrating predictions based on context, the system can pre-fetch potential cached segments or even correct potential errors made by the ASR in real-time. This combination of caching common phrases and using NLP models for context-aware prediction creates a synergistic effect, ensuring that transcriptions are not only delivered faster but are also more contextually accurate.

[0027] FIG. 3 is a block diagram that describes a real-time recommendation system 300, according to some embodiments of the present disclosure. In some embodiments, the real-time recommendation system 300 may include an input module 310 configured to receive and process voice input from a customer and an ASR module 320 for transcribing the voice input. The real-time recommendation system 300 may also include a content retrieval module 330 interfaced with a knowledge base, designed to fetch relevant content based on the

transcribed input. And a display module to present the content to a customer service agent in a format designed for rapid comprehension.

[0028] In some embodiments, the input module 310 may use a sliding window time series model to optimize the end time of voice input based on an estimated speech rate. In some embodiments, the ASR module 320 may employ a combined cache and NLP model system to optimize the transcription duration, enhancing both the real-time nature and the accuracy of transcription content retrieval. In some embodiments, the display module may employ a user interface that highlights relevant content and keywords to aid the customer service agent in quickly comprehending and responding to the customer's query.

What is Claimed is:

1. A real-time recommendation method for customer service agent voice data, comprising:

receiving a customer's voice input;

processing said voice input to determine an optimal end time based on speech rate;

transcribing the voice input using an ASR model;

retrieving relevant content based on the transcribed voice input; and

displaying the content to a customer service agent in a manner designed to enable rapid reading and response.

2. The method of Claim 1, wherein processing said voice input includes using a sliding window time series model to estimate current speech rate and predict voice end time.

3. The method of Claim 1, wherein the ASR model transcription duration is optimized by combining cache mechanisms and NLP models to enhance the transcription's real-time retrieval and accuracy.

4. The method of Claim 3, further comprising the steps of:

preprocessing the ASR transcribed text to obtain cleaned text;

matching the preprocessed text with known vocabulary and phrases in a cache to calculate a matching score; and

further processing the ASR text with an NLP model to obtain advanced semantic representations for enhanced content retrieval.

5. The method of Claim 4, wherein the matching score is computed as a ratio of matching vocabulary words from the ASR text to the total vocabulary in the cache.

6. The method of Claim 1, wherein the displayed content to the customer service agent uses a highlighting mechanism and a keyword mode for faster comprehension and improved response times.

7. A real-time recommendation system for customer service agent voice data, comprising:
 - an input module configured to receive and process voice input from a customer;
 - an ASR module for transcribing said voice input;
 - a content retrieval module interfaced with a knowledge base, designed to fetch relevant content based on the transcribed input;
 - and a display module to present said content to a customer service agent in a format designed for rapid comprehension.
8. The system of Claim 7, wherein the input module uses a sliding window time series model to optimize the end time of voice input based on an estimated speech rate.
9. The system of Claim 7, wherein the ASR module employs a combined cache and NLP model system to optimize the transcription duration, enhancing both the real-time nature and the accuracy of transcription content retrieval.
10. The system of Claim 7, wherein the display module employs a user interface that highlights relevant content and keywords to aid the customer service agent in quickly comprehending and responding to the customer's query.