VOCI

Enhancing Customer Experience for Health Insurance *Case Study*

One of the country's largest health insurance providers with nearly 15 million members wanted to improve the outcomes of its automatic speech recognition (ASR) system. This included overall accuracy as well as an improved ability to identify industry-specific terms spoken by its customers and call center agents.

Voci's V-Blaze ASR solution was selected as the provider's speech-to-text (STT) transcription engine. After Voci integrated custom language models into the ASR system, the provider experienced significantly enhanced transcription accuracy and a better understanding of how best to serve their customers.

Challenge

Key terms are important for categorizing, identifying discussion topics, and assessing sentiment. Improving key term recognition can often have a greater impact on the delivery of actionable insights than generalized transcription accuracy improvement of a language model. The provider's ASR system had been automatically generating transcripts that did not include many key healthcare terms spoken by customers and call center agents. Without accurate transcriptions of industry-specific terms, the provider would not be able to improve its Net Promoter Score® (NPS®), which measures customer experience, predicts business growth, and reduces customer turnover.

Key Term Examples

Spoken by Call Center Agents
Ambulance, Anesthesia, Copays, Enroll, HMO, Visits

Spoken by Customers
Canceled, Colonoscopy, HMO, Rehab, Resubmitted

What Is a Corpus?

Corpus is the Latin word for "body." In linguistics, it refers to a collection of related writings. In relation to ASR, the writings are typically transcripts of speech that include timing information and tags indicating non-word vocalizations. It can also refer to the speech audio itself, indicating a complete body of material suitable for training an ASR system using machine-learning techniques.

Solution

The key to improving speech recognition for a specific enterprise is to tune the ASR system with audio captured from the enterprise's telephone system. This ensures the tuning will result in better alignment between the ASR system and the specifics of what agents and customers are discussing, their accents and speech patterns, and even noises and distortions peculiar to the telephone system. The components of the ASR system subject to this kind of tuning are called *language* and *acoustic* models. Language models capture the statistical relationships of word groupings while acoustic models capture the relationships between digitized audio sounds and basic components of speech known as phonemes.

Approximately 10,000 hours of call audio were used to produce the custom language and acoustic models. Roughly 10% of these calls were manually transcribed to provide primary model training material, while the remainder were processed through a Voci proprietary system to provide supplemental training material. As is typical, approximately half of the audio contained speech useful for training. This speech audio and transcripts generated from it comprise the *corpus* used for creating the custom models.

Over all, nearly 4 million words were manually transcribed. This yielded about 8,500 unique vocabulary words spoken by agents and 21,000 unique vocabulary words spoken by customers. More than 1,000 new words were added to the pronunciation dictionary used by these models. Due to significant differences in agent and customer speech, separate models were produced, specifically tuned for each group. This enabled greater improvements in accuracy than could be achieved by modeling agent and customer speech combined.

Results

By incorporating agent and customer language models that were tuned using the provider's own call audio into Voci's V-Blaze ASR system, significant transcription accuracy improvements were achieved. Custom language models enabled high-precision text analytics, with relative word error rate improvements of over 35% for agent speech and over 20% for customer speech. *Word error rate* is the strictest measure of transcription error, weighting all error types equally.

In addition to overall accuracy improvements, substantial improvements to high-value terms were also observed. An overall 20% or greater improvement in *recall* was observed with 650 key terms recognized by the delivered custom agent models. The tailored customer models delivered the same degree of improvement for more than 2,000 key terms. Many of these terms had recall improvements of several hundred percent with some exceeding 1,000% improvement.

Greater transcription accuracy enabled the health insurance provider to better identify and categorize issues and trends that were important to their customers, as well as to assess their customer's overall happiness with the provider's services. What's more, the provider was able to more effectively gauge its agents' performance and productivity and improve their training.

Conclusion

By using Voci's custom language models with its V-Blaze ASR system, the health insurance provider is now able to quickly extract high-quality, actionable intelligence from vast quantities of call audio data. The provider better understands what its customers and agents are saying during calls, as well as the sentiments and emotions they are expressing. This in turn has led to improved customer experience and reduced churn.

Having experienced the benefits of greater transcription accuracy, the provider has requested further enhancements. These include recognizing more healthcare terminology and adding support for more languages such as Spanish. With this additional investment, the quality and quantity of insights extracted from their voice assets will improve further and drive even greater business success.

What Is Language Modeling?

Language modeling is the process of creating a machine-learning component that a software system can utilize to convert a stream of phoneme symbols into a stream of written words. The process involves creating high-quality transcripts of spoken words, then feeding those words into training software that analyzes the relationships between the words and produces a statistical model of the likelihood of each possible word grouping. The resulting model is then composed into a data structure that can be traversed at high speed during the ASR transcription process.

General-purpose language models are trained against a broad range of speech obtained from many sources. This enables a model to perform well against speech audio from any source. However, like a "one size fits all" garment, this model does not provide a perfect fit for any particular source, especially one that was not available at the time the general model was created.

A custom language model is created using audio from a single source. In this sense, it is like a tailored garment, specifically designed to fit its subject perfectly. Such a model naturally delivers better accuracy for the subject than a general model can.

What Is Recall?

Recall is a statistical measure indicating the percentage of items that were found out of the total of number of such items existing within a population. For example, if the word "Voci" is said 10 times in a call and the ASR system recognizes and transcribes the word "Voci" 8 times, the recall is 80%. Recall can be thought of as the "hit rate."

About Voci Technologies

Voci Technologies combines artificial intelligence (AI) and deep learning algorithms to deliver the best-in-class enterprise speech analytics platform. Voci's innovative technology and strategic partnerships enable contact centers of all sizes to extract actionable intelligence from voice data to improve customer experience, operational efficiency and compliance requirements. For information, visit www.vocitec.com.