Google Discussions

EXAM PROFESSIONAL MACHINE LEARNING ENGINEER TOPIC 1 QUESTION 282 DISCUSSI...

Actual exam question from Google's Professional Machine Learning Engineer

Question #: 282

Topic #: 1

[All Professional Machine Learning Engineer Questions]

You work at an organization that maintains a cloud-based communication platform that integrates conventional chat, voice, and video conferencing into one platform. The audio recordings are stored in Cloud Storage. All recordings have an 8 kHz sample rate and are more than one minute long. You need to implement a new feature in the platform that will automatically transcribe voice call recordings into a text for future applications, such as call summarization and sentiment analysis. How should you implement the voice call transcription feature following Google-recommended best practices?

A. Use the original audio sampling rate, and transcribe the audio by using the Speech-to-Text API with synchronous recognition.

B. Use the original audio sampling rate, and transcribe the audio by using the Speech-to-Text API with asynchronous recognition.

C. Upsample the audio recordings to 16 kHz, and transcribe the audio by using the Speech-to-Text API with synchronous recognition.

D. Upsample the audio recordings to 16 kHz, and transcribe the audio by using the Speech-to-Text API with asynchronous recognition.

Show Suggested Answer

by Yan_X at *Feb. 12, 2024, 9:24 a.m.*

Comments

Type your comment...

Submit

CHARLIE2108 Highly Voted 7 months ago

Selected Answer: D

I went with D.
"following Google-recommended best practices"
https://cloud.google.com/speech-to-text/docs/optimizing-audio-files-for-speech-to-text#:~:text=We%20recommend%20a%20sample%20rate%20of%20at%20least%2016%20kHz%20in%20the%20audio%20files%20that%20you%20use%20for%20transcription%20with%20Speech%2Dto%2DText

upvoted 7 times

wences Most Recent 1 month, 1 week ago

Selected Answer: B

Agree on B. If you read carefuly the documentation pointed will come to the conclusion that there is no need to upsample voice

upvoted 2 times

asmgi 3 months, 1 week ago

Selected Answer: B

We have longer than minute, 8KHz recordings.

https://cloud.google.com/speech-to-text/docs/best-practices-provide-speech-data
"avoid re-sampling. For example, in telephony the native rate is commonly 8000 Hz, which is the rate that should be sent to the service."
-> 8KHz
https://cloud.google.com/speech-to-text/docs/sync-recognize
"Synchronous speech recognition returns the recognized text for short audio (less than 60 seconds). To process a speech recognition request for audio longer than 60 seconds, use Asynchronous Speech Recognition."
-> asynchronous

So, the correct answer is B.

upvoted 4 times

PhilipKoku 4 months, 1 week ago

Selected Answer: B

B) Use original sampling rate and use asynchronous recognition...
"If possible, set the sampling rate of the audio source to 16000 Hz. Otherwise, set the sample_rate_hertz to match the native sample rate of the audio source (instead of re-sampling)."
https://cloud.google.com/speech-to-text/docs/best-practices-provide-speech-data#sampling_rate

upvoted 3 times

livewalk 5 months ago

Selected Answer: B

According to google recommandation on Sampling rate: "If possible, set the sampling rate of the audio source to 16000 Hz. Otherwise, set the sample_rate_hertz to match the native sample rate of the audio source (instead of re-sampling)."
So we should match the native sample (8kHz) in the question.

upvoted 2 times

pinimichele01 6 months ago

Selected Answer: B

https://cloud.google.com/speech-to-text/docs/best-practices-provide-speech-data: Capture audio with a sampling rate of 16,000 Hz or higher. Lower sampling rates may reduce accuracy. However, avoid re-sampling. For example, in telephony the native rate is commonly 8000 Hz, which is the rate that should be sent to the service.

https://cloud.google.com/speech-to-text/docs/optimizing-audio-files-for-speech-to-text#sample_rate_frequency_range: It's possible to convert from one sample rate to another. However, there's no benefit to up-sampling the audio, because the frequency range information is limited by the lower sample rate and can't be recovered by converting to a higher sample rate.

-----> B, not D

upvoted 1 times

SahandJ 6 months ago

Selected Answer: B

According to the documentation, it's best to have 16 KHz sample rate, however one should avoid up-sampling and rather use the native sample rate

upvoted 2 times

ludovikush 6 months, 1 week ago

Selected Answer: B

Following best practices, the easiest choice is B

upvoted 2 times

omermahgoub 6 months, 1 week ago

Selected Answer: D

Upsample to 16 kHz and Use Asynchronous Speech-to-Text Recognition

upvoted 1 times

tavva_prudhvi 6 months, 3 weeks ago

Selected Answer: D

Upsampling to 16 kHz:
The Speech-to-Text API recommends an audio sample rate of 16 kHz for optimal transcription accuracy. Upsampling the 8 kHz recordings to 16 kHz will improve the quality of the transcription.

Asynchronous Recognition:
Asynchronous recognition is suitable for longer audio recordings (more than one minute). It allows you to submit the audio file and receive the transcription results later, which is more efficient for batch processing.

https://cloud.google.com/speech-to-text/docs/best-practices-provide-speech-data

upvoted 4 times

guilhermebutzke 8 months ago

Selected Answer: B

My Answer: B

- Not necessary upsampling (exclude C and D)
- Asynchronous means executing different tasks with no sequential order. Therefore, is preferred over synchronous recognition for longer audio recordings as it allows for more efficient processing, especially when dealing with larger volumes of data.

upvoted 2 times

guilhermebutzke 8 months ago

My Answer: B

- Not necessary upsampling (exclude C and D)
- Asynchronous means executing different tasks with no sequential order. Therefore, is preferred over synchronous recognition for longer audio recordings as it allows for more efficient processing, especially when dealing with larger volumes of data.

upvoted 1 times

Yan_X 8 months, 1 week ago

Selected Answer: B

B

https://cloud.google.com/speech-to-text/docs/speech-to-text-requests#:~:text=Synchronous%20recognition%20requests%20are%20limited,periodically%20poll%20for%20recognition%20results.

upvoted 3 times