EXAMTOPICS

- Expert Verified, Online, **Free.**

☰ MENU 🔍

← Google Discussions

Exam Professional Machine Learning Engineer All Questions

View all questions & answers for the Professional Machine Learning Engineer exam

Go to Exam

📄 **EXAM PROFESSIONAL MACHINE LEARNING ENGINEER TOPIC 1 QUESTION 104 DISCUSSI...**

Actual exam question from Google's Professional Machine Learning Engineer

Question #: 104

Topic #: 1

[All Professional Machine Learning Engineer Questions]

You work for a gaming company that has millions of customers around the world. All games offer a chat feature that allows players to communicate with each other in real time. Messages can be typed in more than 20 languages and are translated in real time using the Cloud Translation API. You have been asked to build an ML system to moderate the chat in real time while assuring that the performance is uniform across the various languages and without changing the serving infrastructure.

You trained your first model using an in-house word2vec model for embedding the chat messages translated by the Cloud Translation API. However, the model has significant differences in performance across the different languages. How should you improve it?

- A. Add a regularization term such as the Min-Diff algorithm to the loss function.

- B. Train a classifier using the chat messages in their original language.

- C. Replace the in-house word2vec with GPT-3 or T5.

- D. Remove moderation for languages for which the false positive rate is too high.

Show Suggested Answer

---

by 👤 **kunal_18** at *Dec. 24, 2022, 5:23 p.m.*

Comments

☐ 👤 **TNT87** `Highly Voted 👍` 1 year, 8 months ago

`Selected Answer: B`

Answer B
Since the performance of the model varies significantly across different languages, it suggests that the translation process might have introduced some noise in the chat messages, making it difficult for the model to generalize across languages. One way to address this issue is to train a classifier using the chat messages in their original language.

👍 ↩ 🚩 upvoted 8 times

☐ 👤 **Zwi3b3l** `Most Recent ⊘` 9 months, 2 weeks ago

`Selected Answer: A`

uniform performance

👍 ↩ 🚩 upvoted 1 times

⊟ 👤 **pinimichele01** 6 months, 3 weeks ago

Adding a regularization term to the loss function can help prevent overfitting of the model, but it may not necessarily address the language-specific differences in performance. The Min-Diff algorithm is a type of regularization technique that aims to minimize the difference between the model predictions and the ground truth while ensuring that the model remains simple. While this can improve the generalization performance of the model, it may not be sufficient to address the language-specific differences in performance. Therefore, training a classifier using the chat messages in their original language can be a better solution to improve the performance of the moderation system across different languages.

👍 ↩ 🚩 upvoted 1 times

☐ 👤 **ciro_li** 1 year, 3 months ago

`Selected Answer: B`

Min-diff may reduce model unfairness, but here the concern is about improving performance. Training models avoiding Cloud Natural API should be more suitable.

👍 ↩ 🚩 upvoted 2 times

⊟ 👤 **tavva_prudhvi** 1 year, 3 months ago

Adding a regularization term to the loss function can help prevent overfitting of the model, but it may not necessarily address the language-specific differences in performance. The Min-Diff algorithm is a type of regularization technique that aims to minimize the difference between the model predictions and the ground truth while ensuring that the model remains simple. While this can improve the generalization performance of the model, it may not be sufficient to address the language-specific differences in performance. Therefore, training a classifier using the chat messages in their original language can be a better solution to improve the performance of the moderation system across different languages.

👍 ↩ 🚩 upvoted 1 times

☐ 👤 **[Removed]** 1 year, 3 months ago

`Selected Answer: A`

A is correct since it encourages the model to have similar performance across languages.

B would entail training 20 word2vec embeddings + maintaining 20 models at the same time. On top of that, there would be no guarantee that those models will have comparable performance across languages. This is certainly not something you would do after training your first model.

👍 ↩ 🚩 upvoted 2 times

☐ 👤 **friedi** 1 year, 4 months ago

`Selected Answer: A`

A is correct, the key part of the question is „[…] assuring the performance is uniform […]" which is baked into the Min-Diff regularisation: https://ai.googleblog.com/2020/11/mitigating-unfair-bias-in-ml-models.html

👍 ↩ 🚩 upvoted 1 times

☐ 👤 **M25** 1 year, 6 months ago

`Selected Answer: B`

Went with B

👍 ↩ 🚩 upvoted 1 times

☐ 👤 **tavva_prudhvi** 1 year, 7 months ago

`Selected Answer: B`

Since the current model has significant differences in performance across the different languages, it is likely that the translations produced by the Cloud Translation API are not of uniform quality across all languages. Therefore, it would be best to train a classifier using the chat messages in their original language instead of relying on translations.

This approach has several advantages. First, the model can directly learn the nuances of each language, leading to better performance across all languages. Second, it eliminates the need for translation, reducing the possibility of errors and improving the overall speed of the system. Finally, it is a relatively simple approach that can be implemented without changing the serving infrastructure.

👍 ↩ 🚩 upvoted 4 times

☐ 👤 **hakook** 1 year, 8 months ago

Selected Answer: A

should be A
https://ai.googleblog.com/2020/11/mitigating-unfair-bias-in-ml-models.html

👍 ↩ 🚩 upvoted 2 times

☐ 👤 **Ml06** 1 year, 8 months ago

B i think is the correct answer
C is an overkill , you have just developed your first model you don't jump into solution like C , in addition the problem is that there is a significant difference between language note the model is enormously underperforming .
Finally you are serving millions of users , running chat GPT or T5 for a task like chat moderation (and in real time) is extremely wasteful .

👍 ↩ 🚩 upvoted 3 times

☐ 👤 **John_Pongthorn** 1 year, 9 months ago

Given that GPT-3 is rival of google , C is not possible certainly .

👍 ↩ 🚩 upvoted 3 times

☐ 👤 **John_Pongthorn** 1 year, 9 months ago

we are taking into account 20 muti classification, it is relevant about FP or FN.

👍 ↩ 🚩 upvoted 1 times

☐ 👤 **egdiaa** 1 year, 10 months ago

Selected Answer: C

GPT-3 is best for generating human-like Text

👍 ↩ 🚩 upvoted 3 times

☐ 👤 **lightnessofbein** 1 year, 8 months ago

Does "moderate" means we need to generate text?

👍 ↩ 🚩 upvoted 2 times

☐ 👤 **kunal_18** 1 year, 10 months ago

Ans : C
https://towardsdatascience.com/poor-mans-gpt-3-few-shot-text-generation-with-t5-transformer-51f1b01f843e

👍 ↩ 🚩 upvoted 1 times

Start Learning for free

👥

Social Media

Facebook   Twitter

EXAMTOPICS

We are the biggest and most updated IT certification exam material website.

Using our own resources, we strive to strengthen the IT professionals community for free.