Multi-language translator

HOW TO PRIORITIZE LANGUAGE OPTIONS

There are a few factors to take into consideration when choosing the languages to focus on. We want to make our website as user friendly as possible to our clients. We should first build a profile for our ideal customer and optimize based on that. Here are a few considerations:

- customers from some countries are more likely than others to create an investment portfolio
- customers from well developed countries might produce higher revenues because they have larger capital
- customers from English-native countries or other areas that are well known for speaking English very well (like Scandinavian countries) are less likely to require translation
- some languages are spoken by a very large population, so it therefore cancels out a lot of the potential economical drawbacks for the prospects

Based on the aforementioned points, I would recommend focusing on the following translation features:

1. Chinese (for being widely spoken)
2. French (for also being widely spoken, with a large population of its speakers having low proficiency in English)
3. Spanish (same as above)
4. German (because the DACH space has a high level of financial education and its speakers are generally wealthy)
5. Arabic (a language that's spoken in a lot of countries that are undergoing an economic boom)

MODEL EVALUATION

The model translation performance has been evaluated using these 3 metrics:

1. **Bleu score**: stands for bilingual evaluation understudy, and is a pretty standard metric that works by matching n-grams between the 2 sentences. It typically works decently for short sentences like the ones from the test set, however if translations have very few matching words the score can be rather low even if the semantics were close or even identical. In conclusion, this would be too simplistic to use.

2. **Chrf score**: also based on n-grams, but way less penalizing than bleu (see for example the translation of row 17: "[Broker name] [param name] fees are about half of the industry average.", which was one of the best performing samples. It received a staggering 23,7% rating from bleu, but 93,7% from chrf).
3. **Bert f1-score**: a more advanced metric that computes the cosine similarity of the sentences in the embedded space. This has several advantages over n-gram based metrics like capturing equivalent expressions that use completely different words and accommodating longer range dependencies of tokens. This is the most appropriate method to do evaluation out of the 3.

After running the comparison between our HuggingFace based translator and the Google Translate API we came to the following results:

Model	Bleu score (range [0..1])	Chrf score (range [0..100])	Bert f1-score (range [0..1])
HuggingFace	0.065	**29.434**	**0.859**
Google Translate	**0.072**	25.519	0.831

FUTURE MODEL IMPROVEMENTS

Suggestions to improve the production model:

- In case we have more domain specific data, we can fine tune the model to adapt to the task at hand. For that we also need a solid sweeping strategy for selecting the best hyper-parameters, potentially using some bayesian optimisation
- We can also perform some in-depth analysis over the examples where the model tends to fail more frequently, and go from there to see how we can fix this issue. That could imply, for example, gathering more labelled data of a specific type and fine tuning the model with it.
- We haven't still considered more engineering related factors, like storage and computation costs. If we analyze multiple options we could decide to use a slightly less performant model as far as the scientific metrics are concerned, but which needs less processing power and storage space and could significantly reduce both the cloud computing costs and the latency the client experiences when they use the website.

FUTURE APP POTENTIAL INNOVATIONS

One business proposal would be to also offer the reverse service to the broker partners. In other words, provide information about the potential market opportunities that large brokers could be targeting, as a consulting service.

For this initiative, the **risks** are rather low because it wouldn't require much additional costs since BrokerChooser already has (I assume) an internal pool of information regarding what choices the retail investors prefer.

The **potential** is quite high, because large retail investors could be interested in spending a lot of money on externalizing this consultation to an entity that has collected a lot of relevant data.