Bahubhashi: A Federated Translator For The Tone of The Language.

By: Venkateshprasad (Udacity verified mail id: vensadk@gmail.com)

Aim:

Project Bahubhashi ('polyglot' in Kannada) aims to translate the tone or the context that passage has to say(possibly in multiple ways based on the audience while not losing the originality). Although now it is English to Kannada, the aim is to extend it to all Indian languages.

Objectives:

- Making translation dynamic by improving it continuously.
- Democratizing the translation.
- Translating the tone rather than words.
- Bringing a new standard in translating scientific articles to make it more readable in Indic languages.
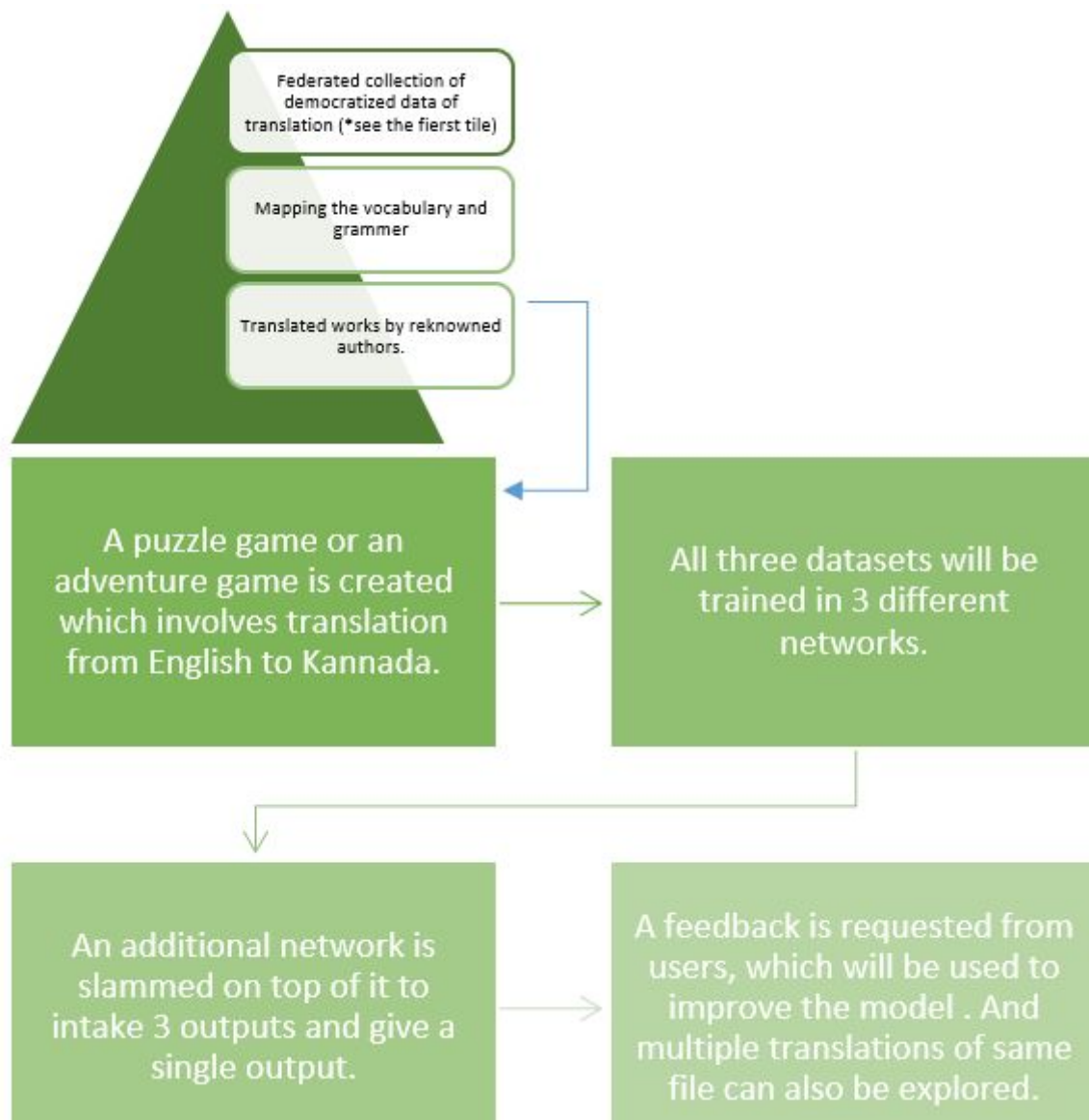
Existing problems:

- Language translators are quite inaccurate.
- Instead of common words, dictionary equivalents are used while translating.
- Translators aren't dynamic. (due to the 2nd reason, a lot of translated statements, even when are correct, don't collocate with the current usage)
- As a language is heavily influenced by the culture of the land, translating a European language like English to Indian languages like Kannada needs to translate the meaning not just statement.
- And while doing so, we need to see that originality is maintained too.
- And language has been by the people and of the people, while current translators aren't sufficiently democratized.
- The translation between English to Indic languages is becoming increasingly important as India is digitizing at a rapid pace.

What I propose(The structure):

- Part 1:
  I am building network-1 which will be fed with English and Kannada versions of books translated by profound authors.
- Part 2:
  Network-2 will be fed with grammatical information and vocabulary for mapping both the languages.

- Part 3:
    1. Network-3 will use a federated learning.
    2. A simple puzzle game(or simple adventure game) which involves people translating English to Kannada or vice versa. And additionally translating custom statements.
    3. The model deployed on local machine will send back only the weights.
    4. Securing and privatizing the whole network is necessary as from every person's translation a lot of personal data can be known.

Flowchart:

Federated collection of democratized data of translation (*see the fierst tile)

Mapping the vocabulary and grammer

Translated works by reknowned authors.

A puzzle game or an adventure game is created which involves translation from English to Kannada.

All three datasets will be trained in 3 different networks.

An additional network is slammed on top of it to intake 3 outputs and give a single output.

A feedback is requested from users, which will be used to improve the model . And multiple translations of same file can also be explored.

How has the course helped:

- I started learning deep learning with the 'Secure and private AI' course. (Network 1 & 2)
- Concepts used in Network-3 are the main aspects the course aimed at teaching.

Udacity - "Talk is cheap, show the code"

Me - "It turns out the idea is very novel and there is very less research done in how to translate European to Kannada. And as a beginner I am putting up a network and also research about the techniques to translate. (I'm working on a paper on that). I hope I'll make it to the 2nd phase which will speed up its deployment"