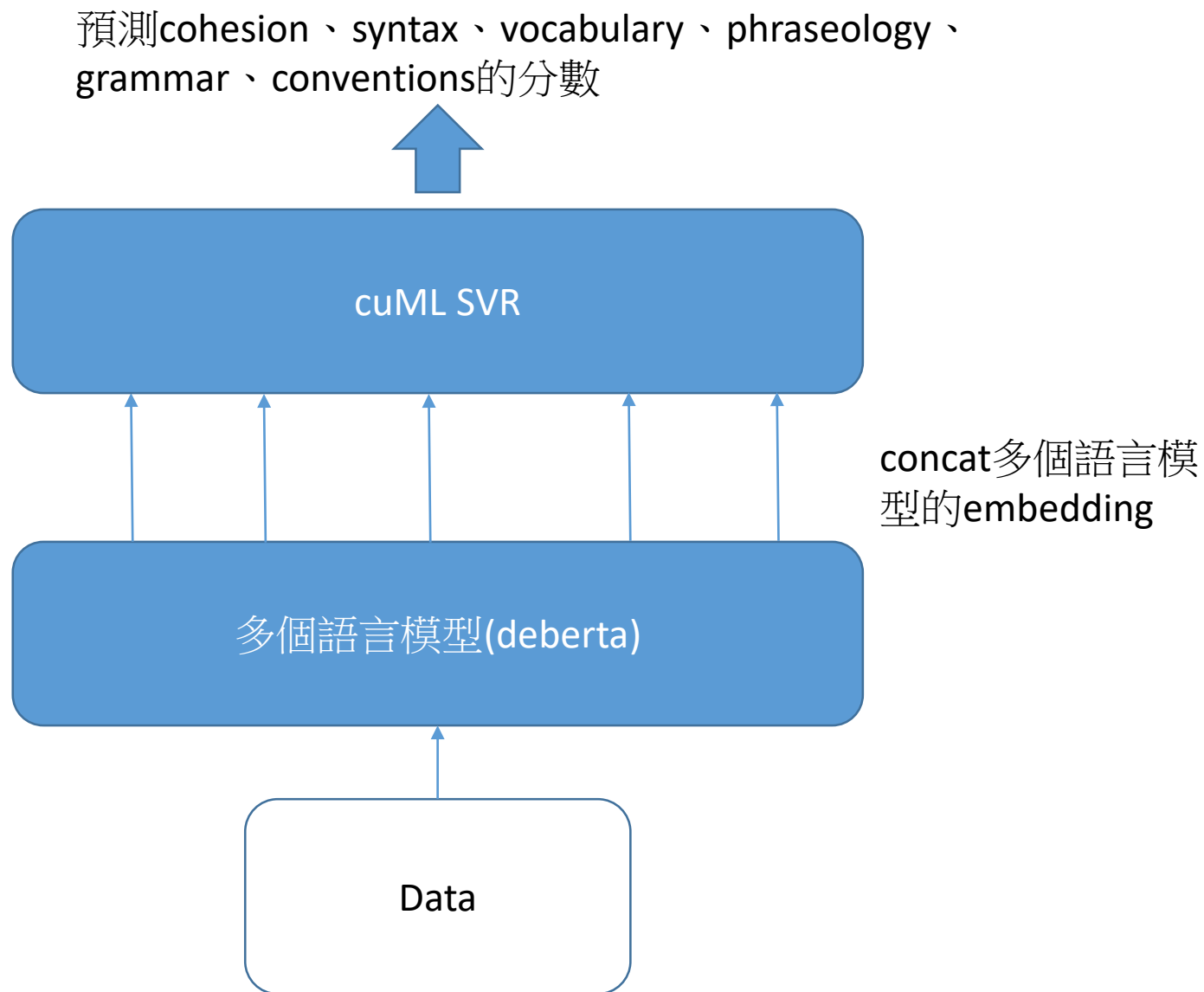


Kaggle  
Feedback Prize - English  
Language Learning

# Overview

- 輸入學生的作文，輸出這篇作文的cohesion、syntax、vocabulary、phraseology、grammar、conventions
- 是一個多label的任務

# 方法



# trick

- 1.將一個句子不同token mean pooling 。
- 2.依照不同的label的rmse來選定不同的語言模型 。
- 3.依照不同的label的rmse來選定token\_max\_length要多少 。



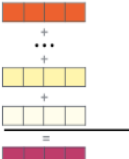



# cuML SVR

- 在GPU上執行所以速度快可以跑很多實驗

# Future work

1. Train with big max\_len, infer low max\_len
2. Fine tune

What is the best contextualized embedding for “Help” in that context?  
For named-entity recognition task CoNLL-2003 NER

		Dev F1 Score
First Layer	Embedding 	91.0
Last Hidden Layer	12 	94.9
Sum All 12 Layers		95.5
Second-to-Last Hidden Layer	11 	95.6
Sum Last Four Hidden		95.9
Concat Last Four Hidden		96.1

在產生last hidden embeddings時，可以選擇使用concat後四層embedding的方式