第十組

M11115020　吳宇祥　　M11015084　魏向晨

M11115079　張家維　　M11115099　郭建鴻

M11115073　張晉嘉　　M11115103　施昱民

實作心得:

```
先安裝 tensorflow==2.5.1
```

1. Training 訓練資料集 -d 後面可以填 wiki_tfidf 或 wiki_sent 此步驟為用維基的資料集去訓練標籤

```
python train_tf.py -m 'bigru_bahdanau_attention'  -d 'wiki_tfidf'
```

```
**calculation on GPU
 tf.Tensor(
[[22. 28.]
 [49. 64.]], shape=(2, 2), dtype=float32)
###############
save flag: True
number of training  226282
number of validating  12424
number of vocab in x 181794
number of vocab in y 13948
x_tr shape (226282, 32)
```

訓練資訊會印在 terminal

我們除了論文資料及外還有嘗試自己找的資料集

從政府的計畫內容系統 GRB 下載資料 並將資料改成論文的 input

指標分數是還蠻高的但是實際看預測標籤好像都沒有預測正確的，因此我們推斷這個 model 可能還是有 domain knowledge 的限制，我們跑了另外一個從維基抓下來的資料集，指標分數仍然不錯，在觀察預測的 label 就有比較明顯的抓出一些關鍵字了。因此我們推測是因為訓練資料是用 wiki 來作訓練，導致一些專門領域的資訊還是沒有學習到，但是對於大部分常見的議題這個模型的預測標籤就有呈現出來~

政府計畫資料

Deep Learning,Convolution Neural Networks,Object Detection Networks,MLP,Digital Ima	The automatic inspection of Printed Circuit Board Assembly (PCBA) is one of th
Deep Learning, Machine Learning, Reinforcement Learning, Robotic Manipulation.	This three-year project will work on the design, development, and implementatio
Deep recommendation model, matrix factorization, latent factors, deep learning, autoencod	This project plans to leverage on the deep learning technology to build a multi-objective recommendation model, which relaxes the constraints of the matrix factorization-based recom- mendation models. The matrix factorization-based recommendation models at least have the following problems. 1. MF-based models assume that the weight of each latent factor is the same, which may not always be a reasonable assumption. 2. MF-based models assume that each latent factor is independent of other factors, which may not always be a reasonable assumption. 3. MF-based models require the number of latent factors of a user equaling the number of latent factors of an item. However, this may not be a good encoding mechanism. 4. We probably over-simplify the interaction between a user and an item if only applying the inner-product operation. Particularly, we should consider higher-order interactions, such as the kernel method or the deep learning related operations. This research attempts to design the recommendation model based on the deep learning architecture to relax these restrictions. For the first issue, we have already developed a prototype such that each latent factor has distinct weight. Initial experimental results show that the new method performs better than the other MF-based approaches on the test dataset in terms of the root-mean-

實際預測結果:

```
the
st st st st st st st st st st
list of the summer olympics
the
antaeotricha
list of
national cup
railway station
list of
list of the league
john
the
john
minuscule
the
john
list of the
the
th brigade
st st st st st st st st st st
high school
list of the summer olympics
the
st st st st st st st st st st
list of the summer olympics
st st st st st st st st st st
```

可以發現其實都猜得不太好都是用沒有意義的字來當答案，我們推測可能跟訓練資料集沒有考慮到 domain knowledge 的關係。

維基新資料:這是額外從維基百科的官方載下來的資料

election,electoral college,by-election,general election,primary election	vote house election poll bill republican party voter candidate senate
hospital,patient,health,medicine,neurology,oncology,cancer	health patient medical doctor hospital disease cancer care drug study
cooking,pasta,barbecue sauce	food eat cook chicken recipe cup cheese add taste tomato
son of god,catholicism,holy spirit (christianity),baptism,faith,christian theology,sacram	god church jesus christian faith lord christ catholic give prayer
olympic gold,summer olympic games,winter olympic games,2008 summer olympics,	china chinese olympics gold olympic team win beijing medal sport
conservative democrat,sarah palin	palin sarah mccain alaska governor president republican woman john
blue,black	look white color black red hair blue light green eye
investment banking,subprime mortgage crisis,great recession,capital market,financial	financial bank market government mortgage bailout billion street wal
weight loss,appetite,weight management,dieting,diet (nutrition),obesity,eating,saturate	weight body fat diet food health eat exercise loss healthy
front-wheel drive,engine,automatic transmission	car power vehicle seat rear wheel engine drive air front
internet,web portal,google search,mobile app,website,web browser,user-generated con	site web google search website online internet social content user
biofuel,petroleum,diesel fuel,energy development,fossil fuel,natural gas,biomass,altern	oil energy gas water power fuel global price plant natural
gene expression	cell gene protein function study elegan expression mutant suggest bra
mobile device,mobile web,android (operating system),smartphone,mobile phone	phone mobile iphone device internet service google app read source
rock music,the band,album	music song band album play sound rock track guitar listen
playoffs	game team run hit season win baseball sox pitch start
father,brother	family brother die sister father mother dad tell son wife
state prosecutor,grand jury,indictment,jury trial,criminal law,acquittal,state court (unite	court law case police charge legal judge crime report federal
footwear,shoe	size shoe bag wear pair fit foot pocket leather weight
primary education,school,secondary education,matriculation,university,high school,stu	school student university college teacher class education learn high p
player-coach,football team,championship,professional football (gridiron)	game team season play football player win week fan coach
grocery store,coupon,sales,buyer,costco,supermarket,purchasing	store sell sale item buy coffee shop price box gift
marketing,business development,industry,supply chain management,project managem	business company technology product customer service provide man
united states army,military police,artillery,aerial warfare,united states marine corps,arm	military force war army soldier gun fire air guard u.s.
house,the living room,dining room,room,kitchen	room bedroom floor apartment kitchen bath call large include center
tax revenue,consumer spending,market share,percentage point,tax rate,revenue,annua	million price percent tax market increase rate number report average
costume	wear baby dress wedding look love clothes girl kid need
marriage,husband	child woman family kid parent baby mother home husband young

模型猜測結果:

```
la hacienda
richard
berts
bridge
santa rosa local fauna
jihad
bacterial transporter
street
swedish albanian association
keith robinson
brian
fernando
vijaya fm
richard
on the road to freedom
france
louis michel
bedford
composite rules hurling
avenue oakland road historic district
she died
ursula rani
aubrey lewis
adam shand
john
terence
homer center junior senior high school
tome
st mary royal
service
robin
microcotyle
induced protein
jamaica air shuttle
globe village fire house
all ireland senior camogie championship
```

可以觀察到出現的字比較有意義了，這樣好像驗證了我們的推測，但是作者沒有多加說明，之後可以再深入探討。