

Lab 3 Report

Implementation

In this program, three models have been implemented for training data then used to predict the tag for each word in given sentence. These three models require different feature set. In the model 1, there is only one type of feature that is extracted from the data set according to the rule of current word and current label. In the model 2, based on model 1, one more feature set would be extracted according to the rule of current label and previous label, then combined with the model 1's feature set. In the model 3 (which is a bonus model), based on model 2, two more feature sets would be considered. These two feature sets are extracted, based on the rule of suffix-3 of word and current word-previous word, then combined with the feature set in model 2. To be specific, the model 1 is ϕ_1 ; the model 2 is $\phi_1 + \phi_2$; the model 3 is $\phi_1 + \phi_2 + \phi_3 + \phi_4$.

The sparse matrix has been used in the program and it is used to construct the Φ which is sparse matrix in the formulation. Using sparse is not only for saving memory but also for saving time.

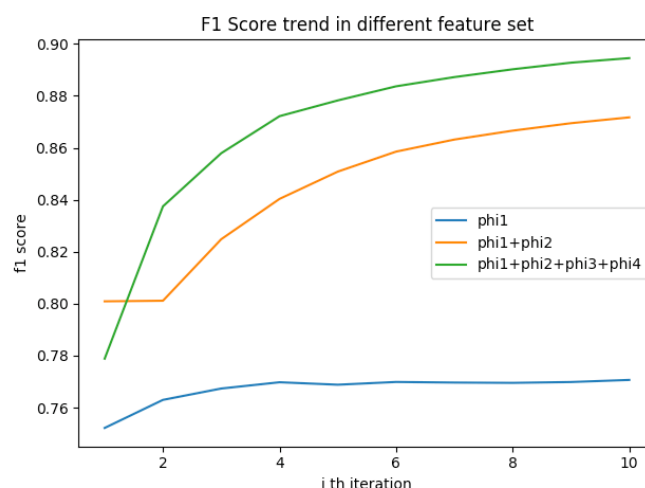
Evaluation

Firstly, for each iteration, the f1 score for three models would be calculated separately. The specific result would be shown in the table below.

F1 Score in Different Model			
i.th iteration	Model 1	Model 2	Model 3
1	0.752184	0.800920	0.778851
2	0.762989	0.801149	0.837471
3	0.767356	0.824828	0.857931
4	0.769770	0.840345	0.872184
5	0.768828	0.850851	0.878253
6	0.769885	0.858544	0.883678
7	0.769655	0.863186	0.887225
8	0.769540	0.866609	0.890230
9	0.769860	0.869476	0.892771
10	0.770667	0.871724	0.894529

From the table above, it is easy to find that the f1 score increase with increased iteration. The model 3 has the highest score at final and the model 1 has lowest score. That is a reasonable and expected result, because more feature sets mean more information provided in prediction. Also, considering the choice of the feature is critical. In model 4, it has 4 useful feature sets so that make it could be more accurate and precision in prediction than other 2 models.

The trend of the increase for three models would be shown as followed image.



This image shows the different trend f1 score in different score. From the image, it could be concluded that more feature set result in better f1 scores or better prediction. For the model 2 and model 3, it seems that if they are given more iterations the result would be improved continuously until reaching the highest. However, for model 1, after 5th iteration, there is no obvious improvement.

After training and validation, the top ten of the most positively-weighted feature for each tag in three models would be selected and displayed in followed table.

Top 10 for each tag in Model 1				
LOC	MISC	O	ORG	PER
ABABA	C\$	10,650,407	&	A.
ABDERDEEN	AMERICAN	14,775,000	ABC	A.de
ABIDJAN	American	15-10	AD-DIYAR	Aamir
ADDIS	Argentine	unq	AD	Adrian
AIRES	Australian	0.056	AHOLD	Affleck
AJACCIO	Austrian	0.69	AHRONOTH	Ahmed
AKRON	Baseball	1.09	AL-ANWAR	Akam
AL-MUNTAR	Bedi	11.38	AL-WATAN	Akram
ALKHAN-YURT	Belgian	17-16	AN-NAHAR	Alan
ALLENTOWN	Brazilian	25.00	ANGOLA	Alastair

In model 1 table, the result is correct and appropriate. The reason why the most words are beginning with character 'A' is that in model 1, less feature extracted result in the same weight for lots of word and when sorting it, if the word has the same weight, the word would be sorted according to the first character.

Top 10 for each tag in Model 2				
LOC	MISC	O	ORG	PER
England	Dutch	,	Newsroom	Mark
LONDON	French	(LG	Younis
AMSTERDAM	C\$)	TEXAS	Armstrong
ATHENS	English	--	Atletico	Fogarty
BONN	GMT	-	CINCINNATI	Jim
BRUSSELS	CENTRAL	0	Grimsby	Kocinski
Finland	DIVISION	1	Hanwha	M.
HONG	GTR	2	Huddersfield	Michael

JOHANNESBURG	German	3	Milwaukee	Peter
MOSCOW	Korean	:	Montreal	Slight

In the model 2, the result in the table is more obviously appropriate. The top ten are all obtained by only one type of feature, even though two types of feature have been used. The reason might be that the tags' order is not important in this case.

Top 10 for each tag in Model 3				
LOC	MISC	O	ORG	PER
bia	C\$,	Newsroom	Jim
TON	GMT)	HOUSTON	Koerts
KANSAS&AT	ish	-15	Oakland	Peter
SAN&AT	DIVISION	2	eld	R.&<s>
Australia	Dutch	:	ham	R.
BALTIMORE&AT	English	DAY	Cleveland	ker
COLORADO&AT	French	on	HOUSTON&<s>	nis
DAM	GTR	vs	St	ott
DETROIT&AT	German	000	ire	ski
ENS	League	0	nre	yne

In the model 4, the result consists of three types of feature. All the results are reasonable and appropriate. The texts in the table, in the yellow highlight, are the suffix-3 of the word. The texts, in the light gray highlight, are the current word and previous word. The texts without the highlight are only text.