Exploiting Bi-LSTM for Named Entity Recognition in Indian Culinary Science

- Ravi Kiran Selvam
- Makesh Narsimhan

Outline:

- Objective
- Related Work
- Data Preprocessing
- Data Annotation
- Proposed Architecture
- Results
- Conclusion & Future Work

Objective:

- There is no dearth for Indian Recipes on the Internet.
- However, finding such information in an organized manner is a problem.
- Thus, the goal of our project is create a dataset in an untapped domain so that it can stimulate further research.
- We used deep learning model to tag the named entities in the Recipes.

Related Work:

- Timothy used a method that involved Conditional Random Fields which learnt the structures of sentences using an undirected model. The dataset used was a set of Tweets and they achieved an F1 score of 49.88.
- An approach that used Bi- Directional LSTMs with Orthographic Sequence Generators yielded an F1 score of 52.41 as proposed by Collier
- Chiu used a Bi-LSTM with character level CNN on the Brown Corpus for NER and achieved an F1 score of 85.53.

Data Preprocessing:

- Removal of extraneous, unnecessary characters that were inadvertently included as part of the dataset during scraping.
- Punctuations, especially commas, were highly improper and proved to be erroneous while handling the dataset as a CSV file.
- Spelling variations led to tagging of different ingredients as separate entities which is undesired.

Data Annotation:

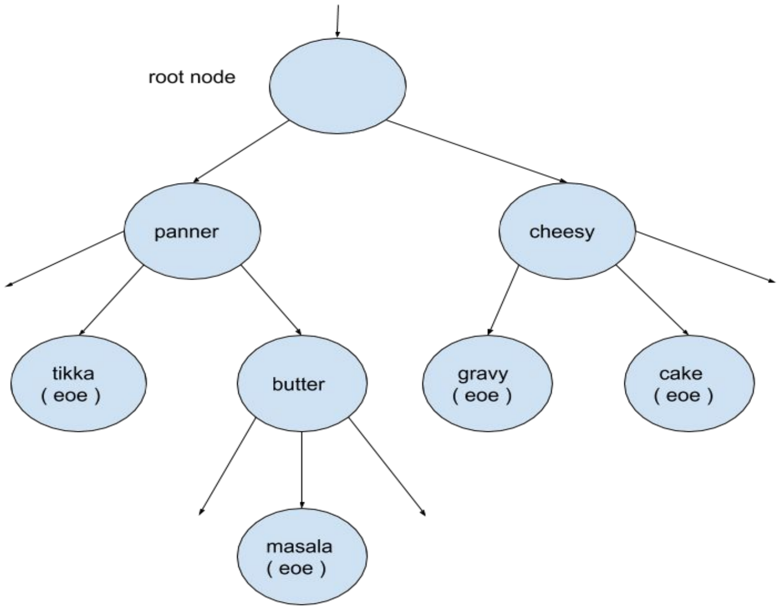


Fig 1: Trie of Words of Named Entities

Proposed Architecture:

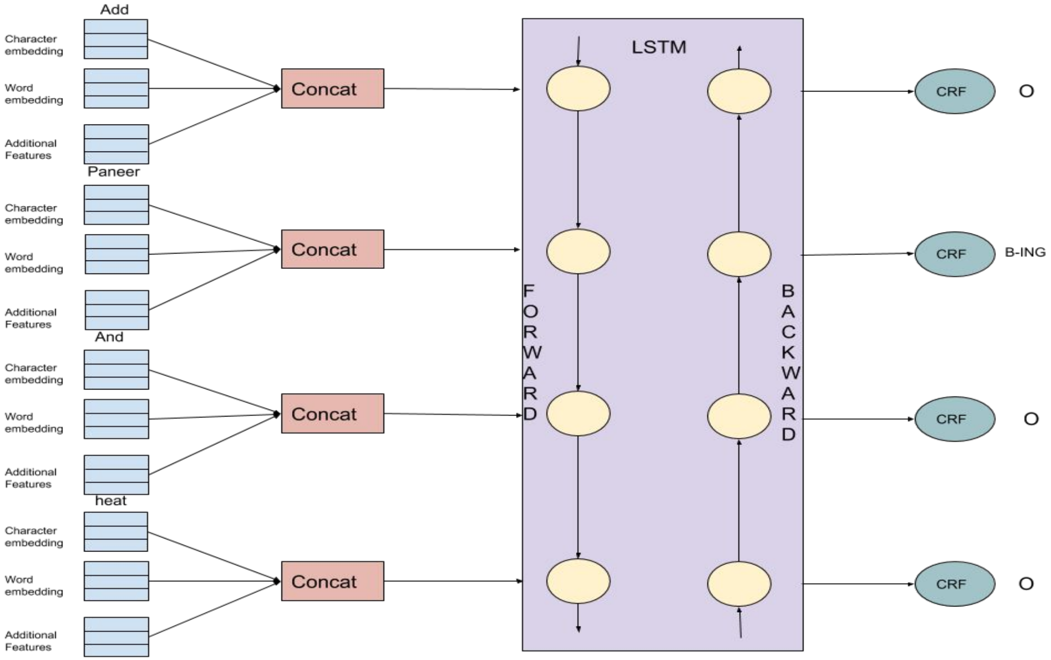


Fig 2: Overall Architecture of the model

Evaluation Metrics:

Precision:

$$precision = \frac{true\ positives}{true\ positives + false\ positives}$$

Mean Average Precision:

$$mean\ average\ precision = \frac{\sum_{i=1}^{i=n} precision_i}{n}$$

Recall:

$$recall = \frac{true\ positives}{true\ positives + false\ negatives}$$

Mean Average Recall:

$$mean\ average\ recall = \frac{\sum_{i=1}^{i=n} recall_i}{n}$$

F1 Score:

$$F1\ score = \frac{2 \cdot precision \cdot recall}{precision + recall}$$

Evaluation Metrics:

Mean F1 Score:

$$mean\ average\ F1\ Score = \frac{\sum_{i=1}^{i=n} F1\ Score_i}{n}$$

where, n = no of folds of validation used.

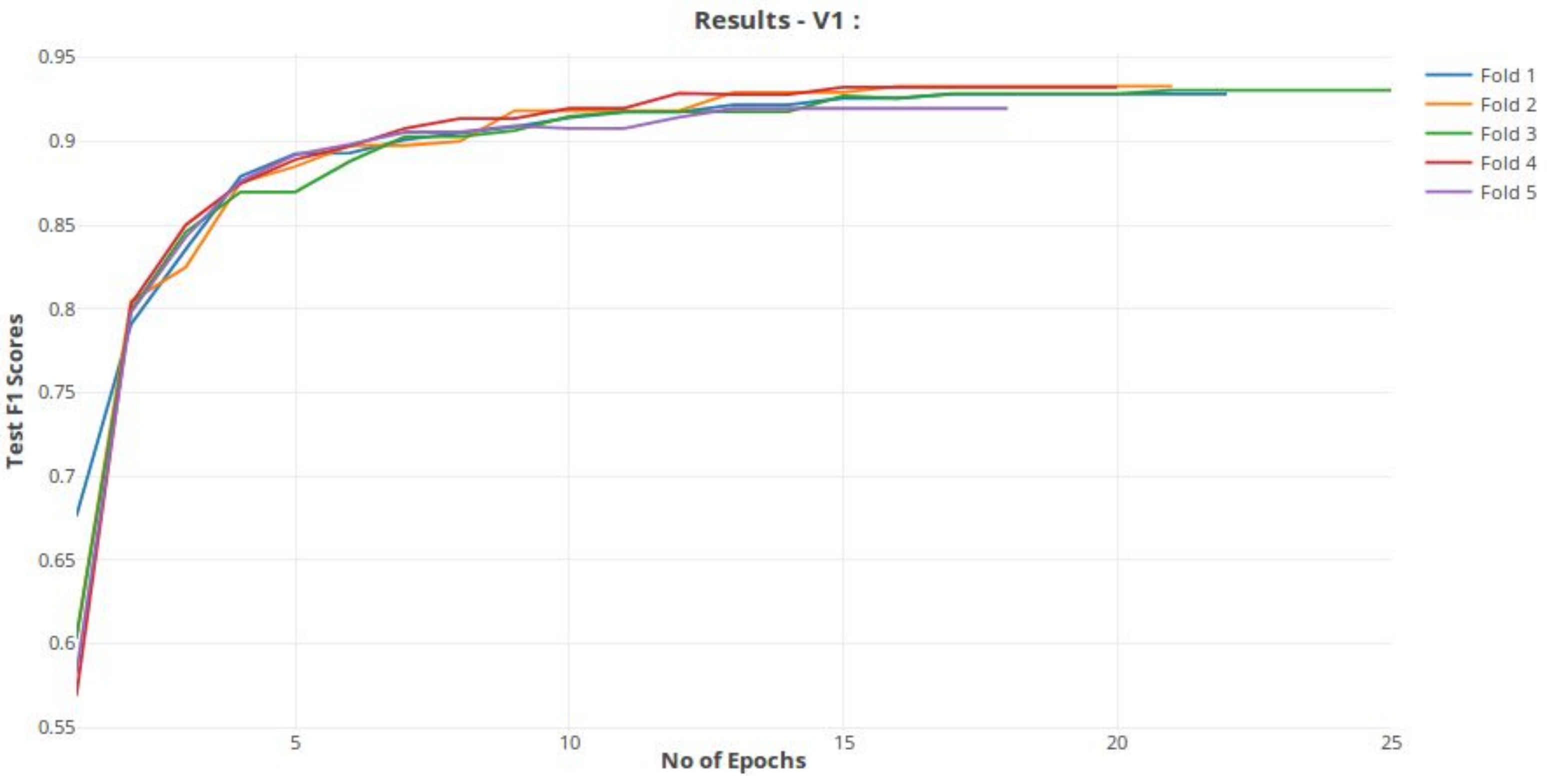
Results:

Version 1:

Train Sentences: 13,682
Dev Sentences: 5,145
Test Sentences: 5,071
Embeddings: Glove [7] 50 dims word embeddings.
Dataset: Not cleaned.
No of Folds of Validation: 5

Mean Average Precision (Test data) = 0.9205
Mean Average Recall (Test data) = 0.9371
Mean F1 Score (Test data) = 0.9287

Results:


Results - V1 :
Test F1 Scores
No of Epochs
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
5
10
15
20
25
Fold 1
Fold 2
Fold 3
Fold 4
Fold 5

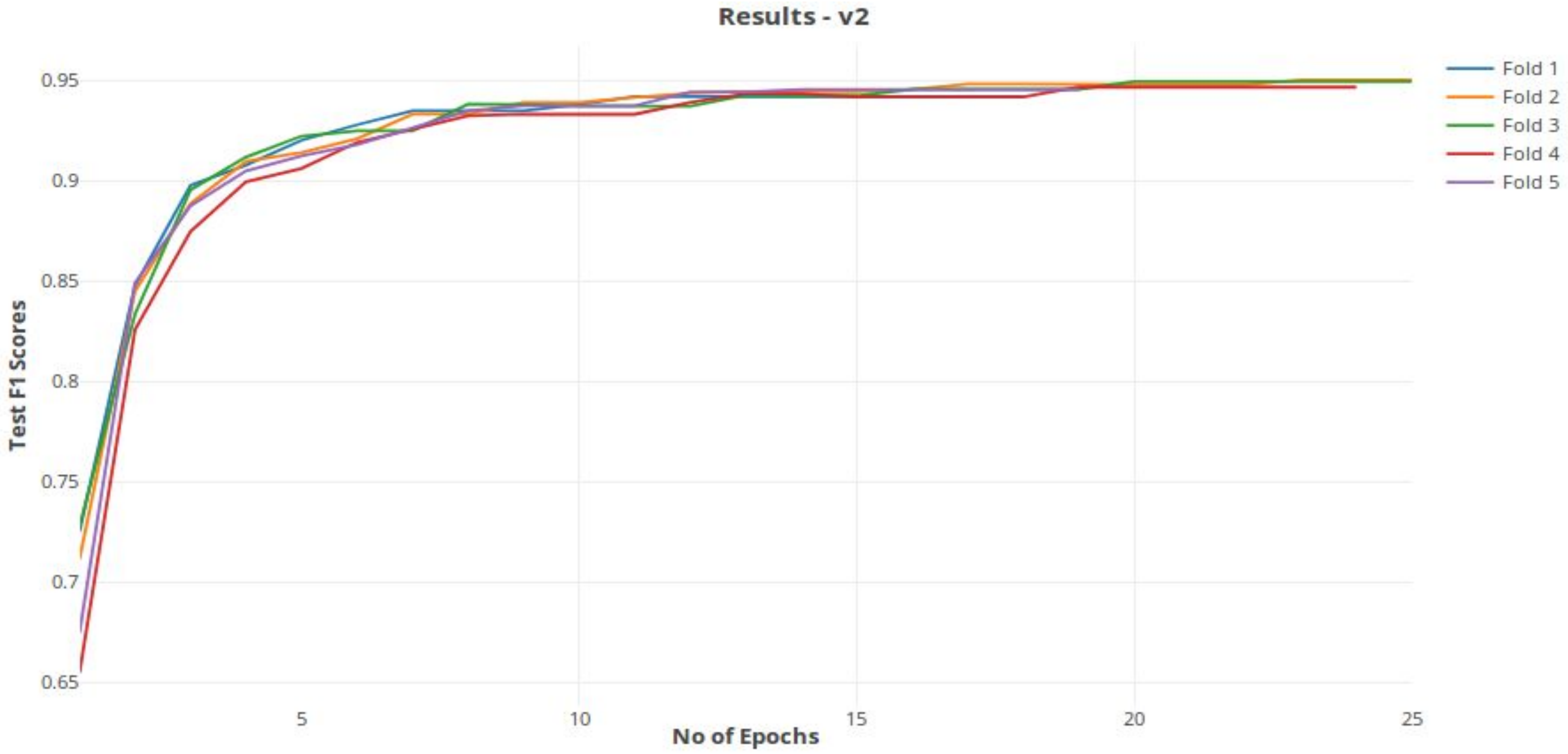
Results:

Version 2:

Train Sentences: 14,323
Dev Sentences: 4869
Test Sentences: 4733
Embeddings: Glove[7] 200 dims word embeddings.
Dataset: Cleaned.
No of Folds of Validation: 5

Mean Average Precision (Test data) = 0.9403
Mean Average Recall (Test data) = 0.9530
Mean F1 Score (Test data) = 0.9466

Results:


Results - v2
Test F1 Scores
0.95
0.9
0.85
0.8
0.75
0.7
0.65
5
10
15
20
25
No of Epochs
Fold 1
Fold 2
Fold 3
Fold 4
Fold 5

Sample Output:

1. Add O
2. two O
3. tbsp O
4. **onion B-ING**
5. and O
6. two O
7. tbsp O
8. **tomato B-ING**
9. to O
10. the O
11. **ginger B-ING**
12. **paste I-ING**
13. along O
14. with O
15. few O
16. amount O
17. of O
18. **salt B-ING**
19. . O

Conclusion & Future Work:

- The data set can be utilized for further research in Culinary domain.
- Other NLP tasks such as breaking of a complex sentence, performing the coreference resolution in culinary science dataset can be performed.


Thank you!