ATIS slot filling - RNN

Chrysomallis Iason – 2014030078

Papadopoulos Argiris - 2014030158

Contents

Introduction

- Natural Language Problem (NLP)
- Automated customer service via phone
- Speech recognition
- Extraction of semantic information (relevancy of words)

- Solution: use of Dense NN (specifically sequential models)

ATIS data set

- Airline Travel Information System (ATIS) – plane tickets
- Collected by DARPA 90s

Words	Show	flights	from	Boston	to	New	York	today
Labels	O	O	O	B-dept	O	B-arr	I-arr	B-date

- 4876 total sentences (labeled) – this is called slot-filling task
- Vocabulary size: 572 words, 127 labels (classes)
- Small data set, K-fold cross validation approach

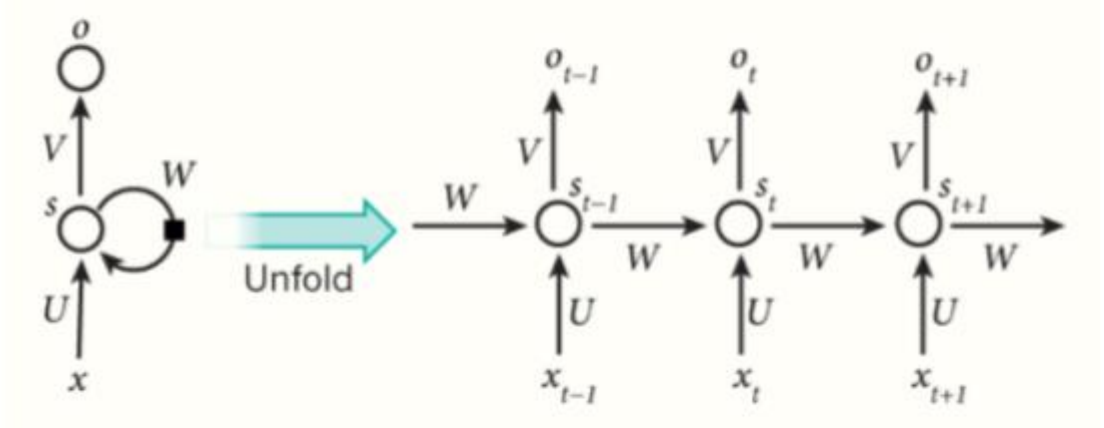
Word Embeddings

- Word Embeddings map words to high-dimensional vectors.
- Capable of learning semantic and syntactic information.
- Nearest neighbors:

sunday	boston	august	time	car
wednesday	nashville	september	schedule	rental
saturday	toronto	july	times	limousine
friday	chicago	june	schedules	rentals
monday	phoenix	december	dinnertime	cars
saturdays	columbus	january	departure	ap

RNN Architecture

- Perfect for sequential problems
- Word prediction based on the previous one



- $x_1, x_2, \ldots, x_{t-1}, x_t, x_{t+1}$: inputs, o_t output at step t, $o_t = f(Vs_t)$
- s_t hidden state at t, $s_t = f(Ux_t + Ws_{t-1})$, f activation function
- U, V, W learnable parameters

Training

- Small data set, 5-fold cross validation
- Word Embeddings as inputs
- Multiple models, distinguish performance (hidden layers):
  - Simple RNN layer
  - LSTM layer (capable of forgetting)
  - Convolutional1D layer between Embeddings and RNN layer (implementation of lookahead)
- Output layer: TimeDistributed
- 30 epochs

Performance (evaluation)

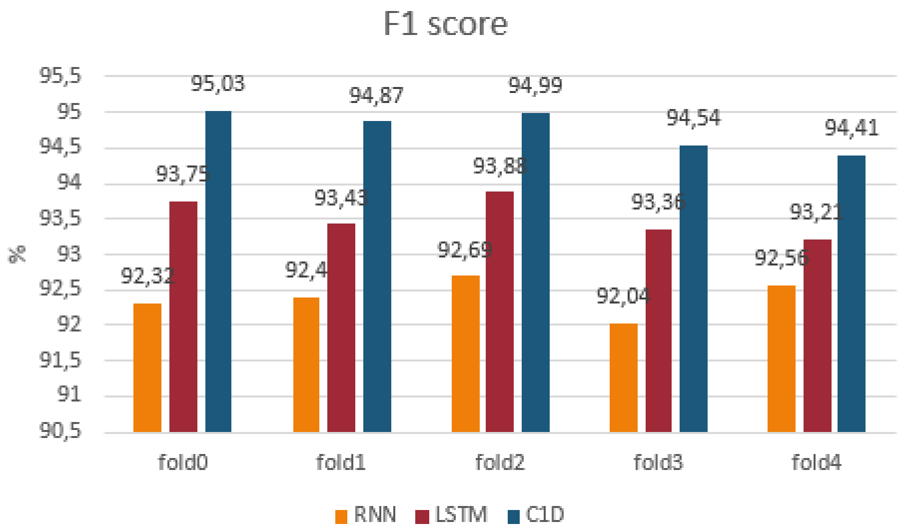
- Criterion: f1 score
- Accuracy is not dependable (good predictions can come from bad samples) due to small size of data set

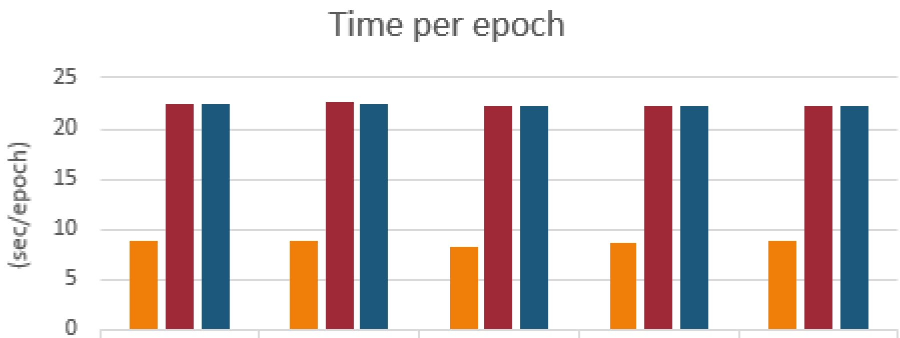
$$F_1 = 2\,\frac{precision \cdot recall}{precision + recall}$$

- Precision: when the model predicts positive, how often is it correct?
- Recall: helps when the cost of false negatives is high

Performance

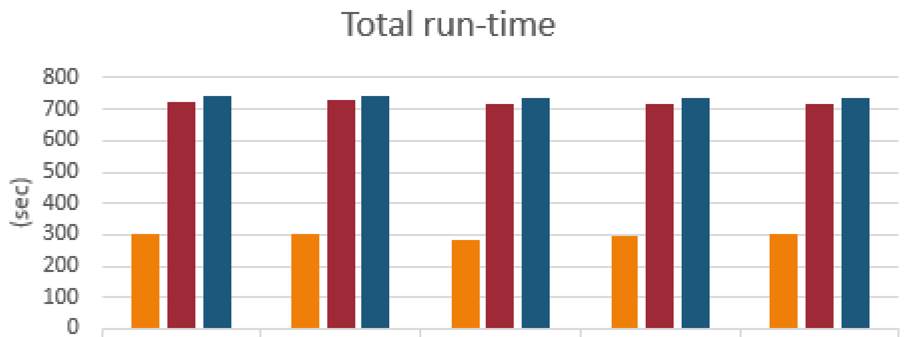
- Specs: i5 6600 3.3GHz
- CPU version of Python 3.7





	fold0	fold1	fold2	fold3	fold4
RNN	8,92	8,82	8,28	8,67	8,81
LSTM	22,46	22,63	22,13	22,16	22,21
C1D	22,41	22,44	22,23	22,27	22,22

RNN LSTM C1D



	fold0	fold1	fold2	fold3	fold4
RNN	304,89	301,89	283,88	297,04	301,51
LSTM	724,96	730,7	714,86	715,73	717,46
C1D	739,49	740,3	733,03	733,97	733,54

RNN LSTM C1D

Conclusion

- NLP problems can be tackled with RNNs
- Simple implementation
- Time and resource efficient
- Complexity reduction and automated systems for customer service

Thank you for your time

Any questions?

- Grégoire Mesnil, Xiaodong He, Li Deng and Yoshua Bengio. Investigation of Recurrent-Neural-Network Architectures and Learning Methods for Spoken Language Understanding. Interspeech, 2013.
- Gokhan Tur, Dilek Hakkani-Tur and Larry Heck. What is left to be understood in ATIS?
- Christian Raymond and Giuseppe Riccardi. Generative and discriminative algorithms for spoken language understanding. Interspeech, 2007.
- Bastien, Frédéric, Lamblin, Pascal, Pascanu, Razvan, Bergstra, James, Goodfellow, Ian, Bergeron, Arnaud, Bouchard, Nicolas, and Bengio, Yoshua. Theano: new features and speed improvements. NIPS Workshop on Deep Learning and Unsupervised Feature Learning, 2012.
- Bergstra, James, Breuleux, Olivier, Bastien, Frédéric, Lamblin, Pascal, Pascanu, Razvan, Desjardins, Guillaume, Turian, Joseph, Warde-Farley, David, and Bengio, Yoshua. Theano: a CPU and GPU math expression compiler. In Proceedings of the Python for Scientific Computing Conference (SciPy), June 2010.