CS 224N: Assignment 3

PETER888@STANFORD.EDU SUNDAY 18TH FEBRUARY, 2018

Problem 1. A window into NER (30 points)

1.1 (a) Understanding NER (5 points, written)

1.1.1 i) Ambiguous Examples (2 points)

Answer:

1. "Walt Disney has just bought Fox for 52bn", here "Walt Disney" and "Fox" are ambiguous type. "Walt Disney" can stands for person, the founder of Walt Disney Company, or can stands for "Walt Disney" organization, the company.

2. "What does the Fox say", here "Fox" is ambiguous type. "Fox" can stands for the animal, or can stands for "Fox" organization, the company.

1.1.2 ii) Why use features (1 point)

Answer:

Words can have multiple means depending on it's context. We need predict named entity labels from their context as well.

1.1.3 iii) Feature examples (2 points)

Answer:

1. Adjacent Verb. For example "Rose said", here Rose should be a person, not the flower rose, because it's adjacent verb "said"

2. Location. If "Chase" happening on highway, that's normally the verb chase, instead of the bank "Chase"

1.2 (b) Computational complexity (5 points, written)

1.2.1 i) Dimensions (2 points)

Answer:

$\boldsymbol{e}^{(t)} \in \mathbb{R}^{1\times((2\omega+1)\times D)}$, $\boldsymbol{W} \in \mathbb{R}^{(2\omega+1)\times H}$, $\boldsymbol{U} \in \mathbb{R}^{H\times C}$

1.2.2 ii) Complexity (3 point)

Answer:

For each step $\boldsymbol{e}^{(t)}$ requires $\boldsymbol{O}((2\omega+1)\times D)$ computation, $\boldsymbol{h}^{(t)}$ requires $\boldsymbol{O}((2\omega+1)\times D\times H)$ computation, and $\hat{\boldsymbol{y}}^{(t)}$ requires $\boldsymbol{O}(H\times C)$ computation.

In total, requires $\boldsymbol{O}(T\times((2\omega+1)\times D+(2\omega+1)\times D\times H+H\times C))$ computation. Ignore the constants and $D, H >> C$, the computation complexity of whole sentence is $\boldsymbol{O}(T\times\omega\times D\times H)$

1.3 (c) Implement model(15 points, code)

Problem 2. Recurrent neural nets for NER (40 points)

2.1 (a) Computational complexity (4 points, written)

2.1.1 i) How many more (1 point)

Answer:

2.1.2 ii) Complexity (3 point)

Answer:

2.2 (b) F_1 score (2 points, written)

2.2.1 i) When CE cost and F_1 decreasing at same time (1 point)

Answer:

2.2.2 ii) Why not F_1 (1 point)

Answer:

2.3 (c) RNN cell (5 points, code)

2.4 (d) RNN model (8 points, code/written)

2.4.1 i) Loss and Gradient Update (3 points, written)

Answer:

2.4.2 ii) (5 points, code)

2.5 (e) More RNN model (12 points, code)

2.6 (e) Train RNN model (3 points, code)

Problem 3. Grooving with GRUs (30 points)

3.1 (a) Modeling latching behavior (4 points, written)

3.1.1 i) RNN cell values (1 point)

Answer:

3.1.2 ii) GRU cell values (3 points)

Answer:

3.2 (b) Modeling toggling behavior (6 points, written)

3.2.1 i) 1D RNN (3 points)

Answer:

3.2.2 ii) GRU cell values (3 points)

Answer:

3.3 (c) GRU cell (6 points, code)

3.4 (d) Learn dynamics (6 points, code)

3.5 (e) Analyze graphs (5 points, written)

Answer:

3.6 (f) Train GRU (3 points, code)