# Will Repeated Reading Benefit Natural Language Understanding?

Lei Sha, Feng Qian, Zhifang Sui

Peking University EECS

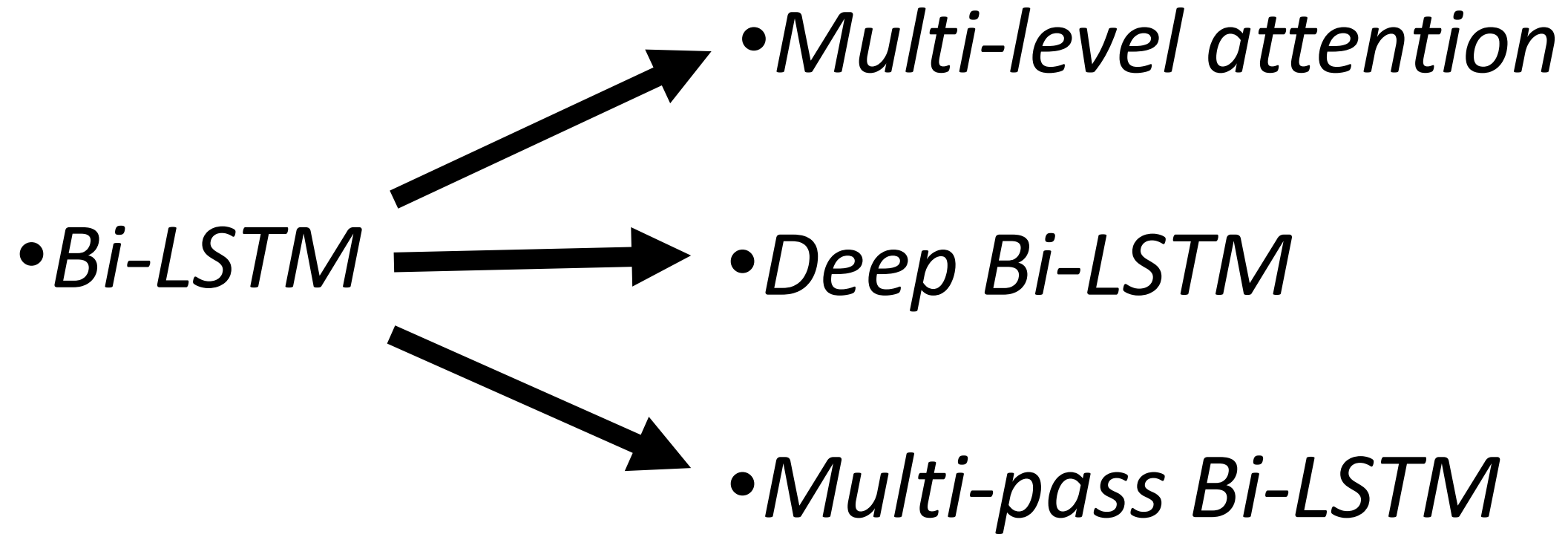
{shalei, nickqian, szf} @ pku.edu.cn

# Question

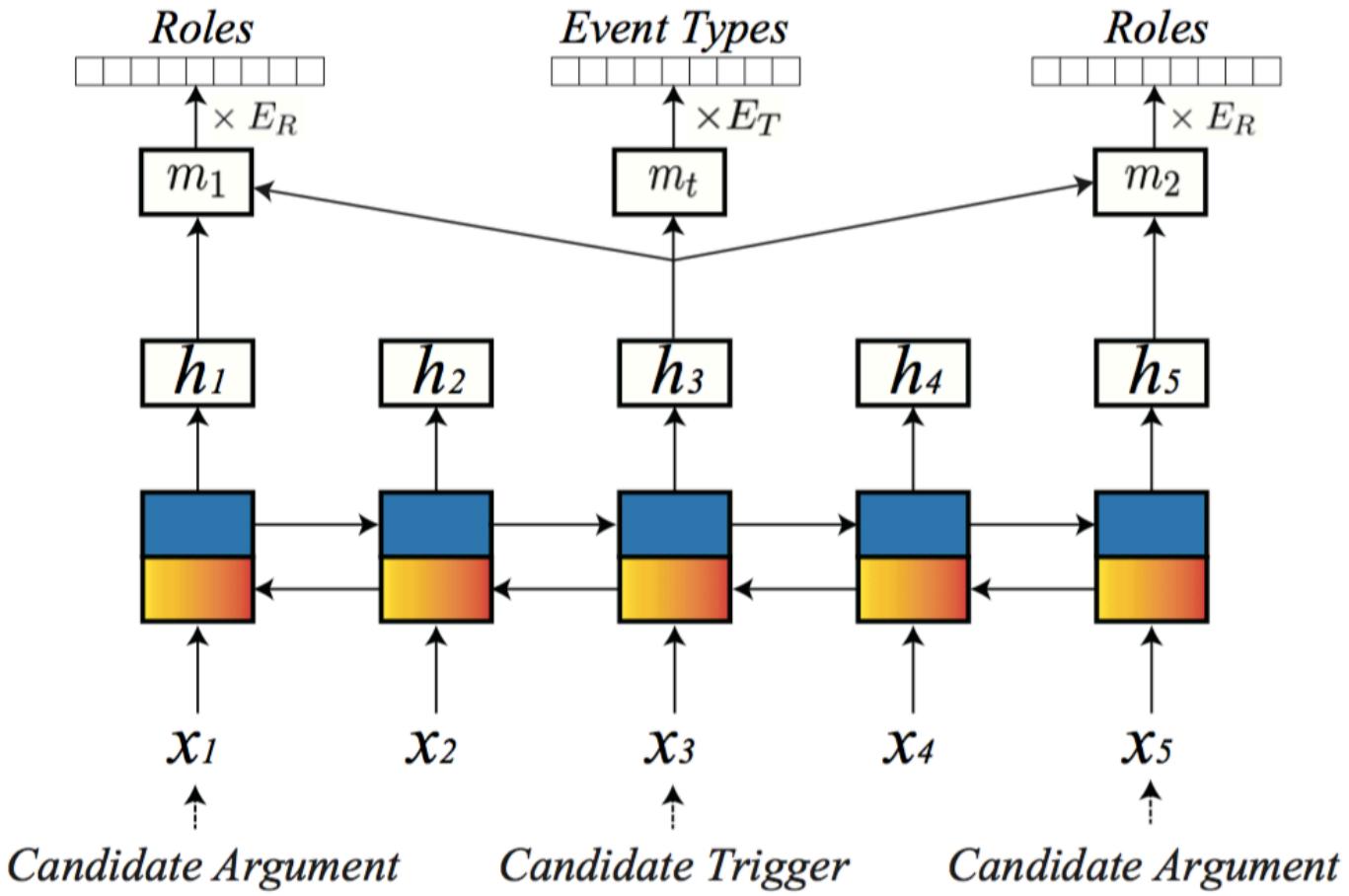
- *Human beings can gain a better understanding by reading a sentence repeatedly.*

- *Will Repeated Reading (re-read) also benefit natural language understanding?*

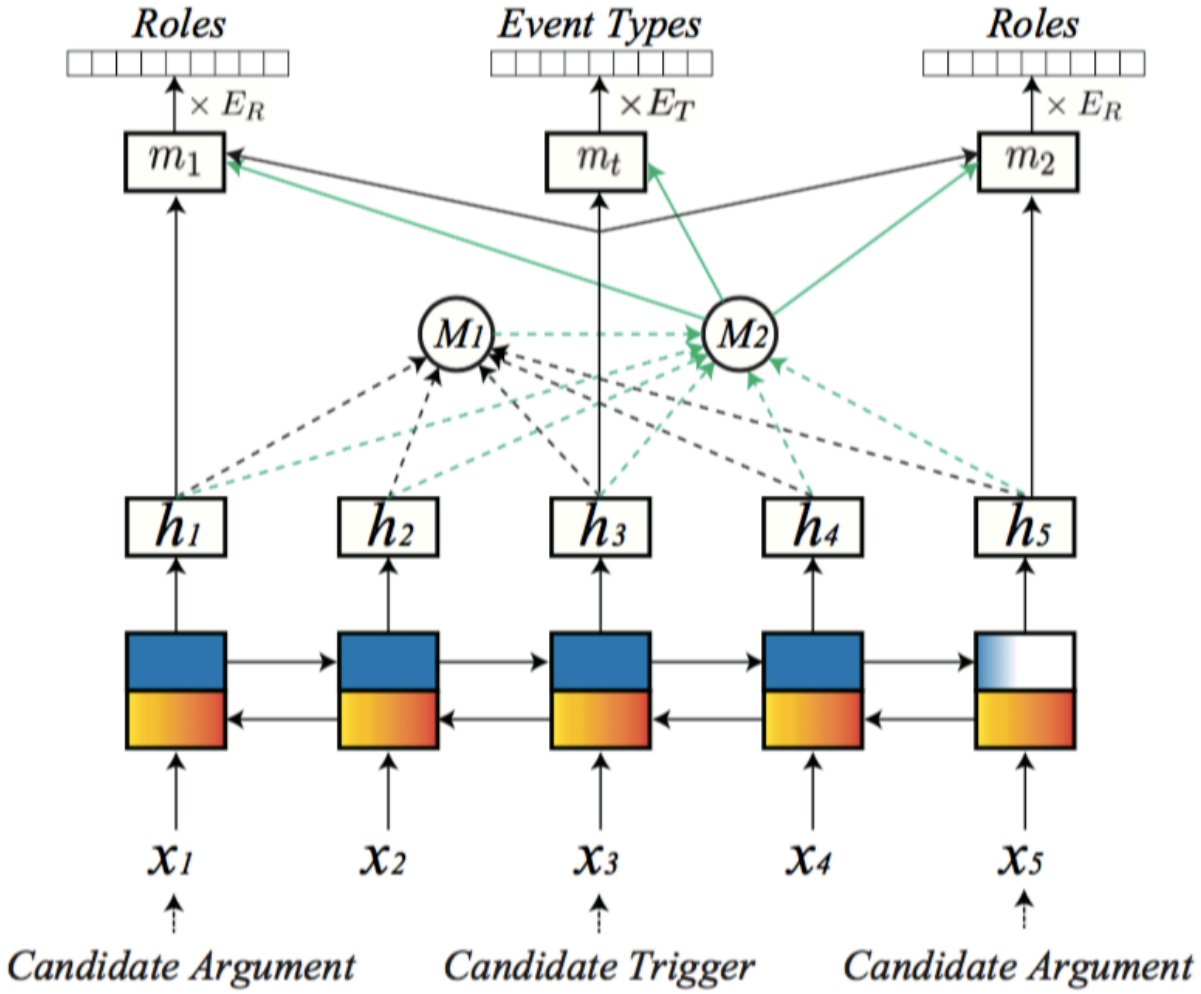
# What structure has the ability to re-read?



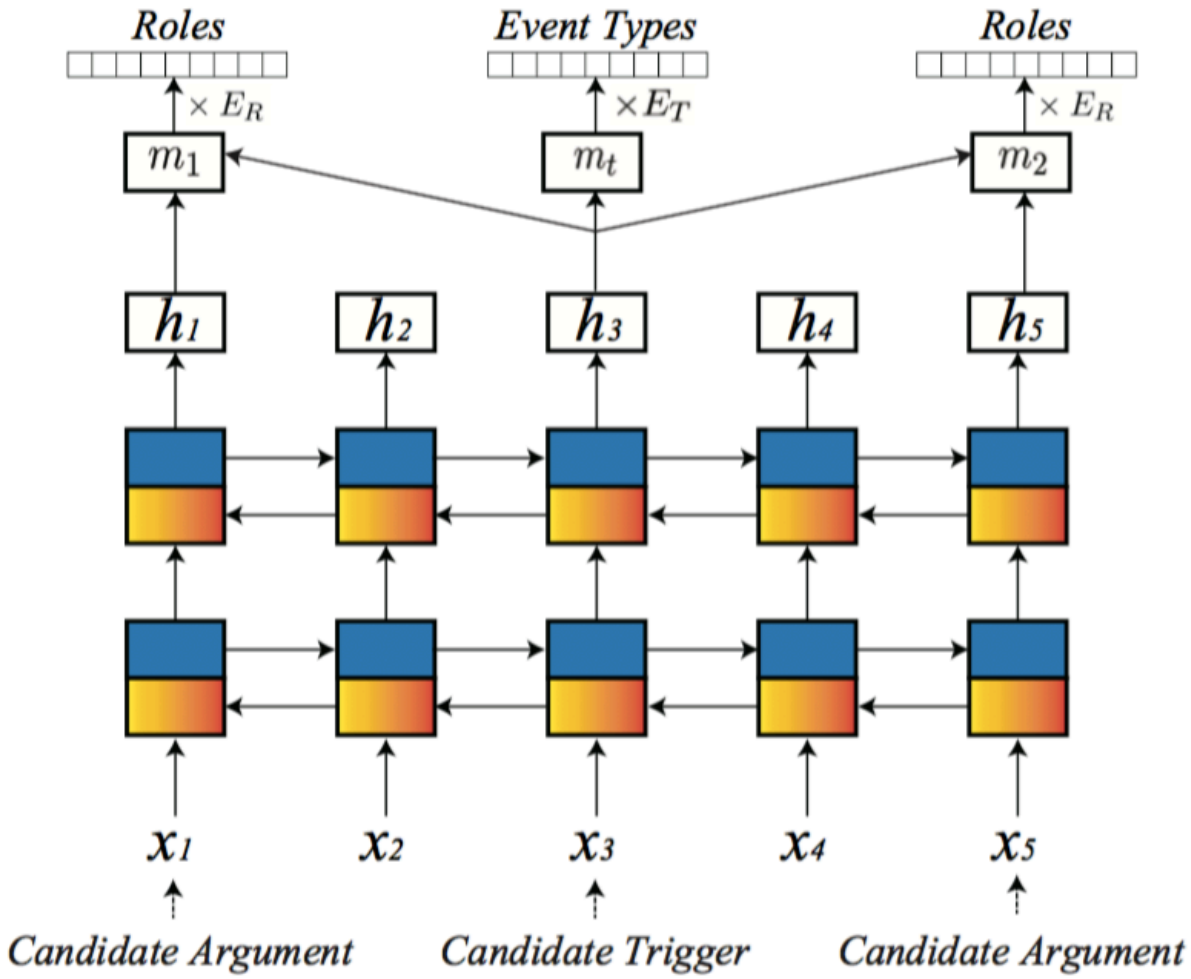
# Standard Bi-LSTM



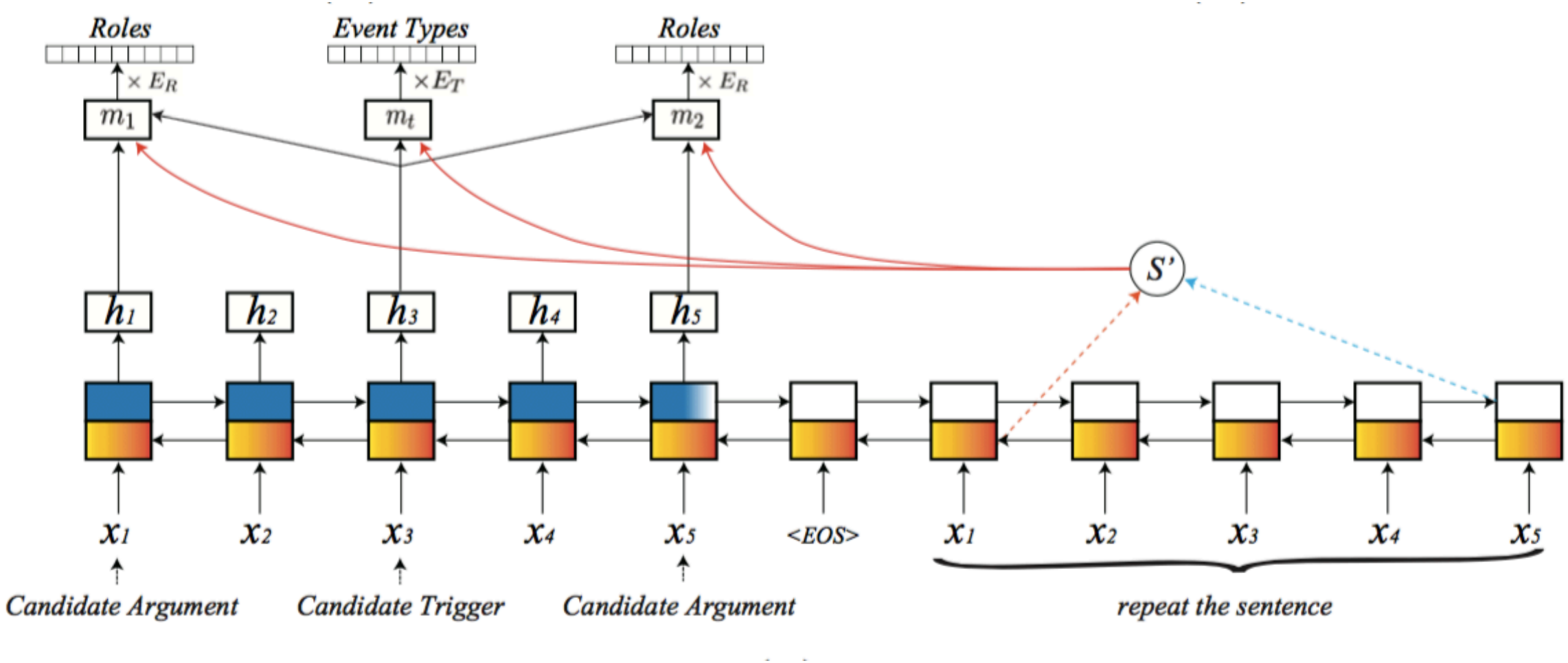
# Multi-level attention mechanism



# Deep Bi-LSTM


Roles
Event Types
Roles
$\times E_R$
$\times E_T$
$\times E_R$
$m_1$
$m_t$
$m_2$
$h_1$
$h_2$
$h_3$
$h_4$
$h_5$
$x_1$
$x_2$
$x_3$
$x_4$
$x_5$
Candidate Argument
Candidate Trigger
Candidate Argument

# Multi-pass Bi-LSTM



# What tasks can we evaluate on?

- *Part-of-speech Tagging*
- *Sentiment Analysis*
- *Semantic Relationship Classification*
- *Event Extraction*

# Part-of-speech tagging

- *Data:* Wall Street Journal (WSJ) data

- Sections 0-18 of the Wall Street Journal (WSJ) data for training
- sections 19-21 for validation
- sections 22-24 for testing.

# Part-of-speech tagging result

| | Accuracy | $p$ value |
|---|---|---|
| Standard | **91.30** | - |
| MLA(1-level) | 91.09 | 0.10960 |
| MLA(2-level) | 90.92 | 0.39532 |
| DB | 81.31 | 0.00096* |
| MPB | 90.74 | 0.07346 |

# Sentiment analysis

- Data: Stanford Sentiment TreeBank
- We reconstruct 11,855 sentences into 215,154 phrases, so that the reconstructed dataset contains 215,154 examples.
- Fine grained and coarse grained: Fine-grained classification result (very negative, negative, neutral, positive, very positive). Coarse-grained classification result (negative, positive).

# Sentiment analysis result (fine grained)

| Fine-grained | Phrase-level | Root-level | Total |
|---|---|---|---|
| Standard | 80.72 | **42.25** | 79.91 |
| MLA(1-level) | 81.25(+0.53) | 40.68(-1.57) | 80.06(+0.15) |
| $p$ value | 0.0002* | 0.00578* | 0.0008* |
| MLA(2-level) | **81.61**(+0.89) | 39.58(-2.67) | **80.15**(+0.24) |
| $p$ value | 0.008* | 0.006* | 0.0007* |
| DB | 79.61(-1.11) | 41.63(-0.62) | 78.26(-1.63) |
| $p$ value | 0.0003* | 0.03156* | 0.03156* |
| MPB | 81.11(+0.39) | 42.08(-0.17) | 79.88(-0.08) |
| $p$ value | 0.0003* | 0.10524 | 0.87288 |

# Sentiment analysis result (coarse grained)

| Coarse-grained | Phrase-level | Root-level | Total |
|---|---|---|---|
| Standard | 80.79 | 72.57 | 79.89 |
| MLA(1-level) | 81.47(+0.68) | 73.04(+0.47) | 80.95(+1.06) |
| $p$ value | 0.0022* | 0.0028* | 0.0129* |
| MLA(2-level) | **81.65**(+0.86) | **73.64**(+1.07) | **81.31**(+1.42) |
| $p$ value | 0.0008* | 0.0006* | 0.0033* |
| DB | 75.99(-4.80) | 69.10(-3.47) | 75.51(-4.38) |
| $p$ value | 0.0004* | 0.0001* | 0.0002* |
| MPB | 80.71(-0.08) | 72.60(+0.03) | 79.78(-0.11) |
| $p$ value | 0.0600 | 0.158 | 0.0238 |

# Semantic relationship classification

- Data: SemEval-2010 Task 8 dataset
- The dataset includes 8,000 training instances and 2,717 test instances.

# Semantic relationship classification result

| | Accuracy | $p$ value |
|---|---|---|
| Standard | 75.54 | - |
| MLA(1-level) | 75.83 (+0.29) | 0.06010 |
| MLA(2-level) | **76.24** (+0.70) | 0.01552* |
| DB | 66.23 (−9.31) | 0* |
| MPB | 75.43 (−0.11) | 0.81034 |

# Event Extraction

- Data: Event Extraction on the ACE 2005 dataset
- The newswire texts in ACE2005 dataset are divided into training (529 documents) / develop (10 documents) / testing (40 documents).

# Event extraction result

| | Trigger Cl | Argument Id | Argument Cl |
|---|---|---|---|
| | $F_1(\%)$ | $F_1(\%)$ | $F_1(\%)$ |
| Standard | 51.68 | 57.44 | 42.09 |
| MLA(1-level) | 53.77(+2.09) | 59.34(+1.90) | 41.43(-0.66) |
| $p$ value | 0.0022* | 0.0003* | 0.8891 |
| MLA(2-level) | 54.68(+3.00) | 60.64(+3.20) | 42.87(+0.78) |
| $p$ value | 0.0043* | 0.0001* | 0.0124* |
| DB | **57.22**(+5.54) | **60.75**(+3.31) | **43.65**(+1.56) |
| $p$ value | 0.0003* | 0.0001* | 0.0002* |
| MPB | 55.21(+3.53) | 59.03(+1.59) | 41.32(-0.77) |
| $p$ value | 0.0015* | 0.0102* | 0.0230* |

# Suggestions for NLPers

- When to use?
- Which to use?

# When to use?

- If the task requires to understand the meaning of the whole sentence instead of single words, we suggest to use repeated reading mechanism.

- Thanks!

# Will Repeated Reading Benefit Natural Language Understanding?

Lei Sha, Feng Qian, Zhifang Sui

Peking University EECS

{shalei, nickqian, szf} @ pku.edu.cn