Time Expression Recognition
Using a Constituent-based Tagging Scheme

Xiaoshi Zhong and Erik Cambria
Nanyang Technological University
{xszhong, cambria}@ntu.edu.sg
Outline

• Time expression analysis
  • Datasets: TimeBank, Gigaword, WikiWars, Tweets
  • Findings: loose structure, differentiable

• Time expression recognition
  • TOMN: a constituent-based tagging scheme
  • Baselines: HeidelTime, SUTime, SynTime, ClearTK, UWTime
  • Datasets: TE-3, WikiWars, Tweets
Time Expression - Examples

Today
Friday
September
Last week
2 years ago
September 2006
2006 September
January 30, 1998
1 September 2006
the third quarter of 1984
Time Expression - Constituents

- Time token
  - Explicitly express time information

- Modifier
  - Modify time tokens

- Numeral
  - Numbers and ordinals (except year)

Today
Friday
September
Last week
2 years ago
September 2006
2006 September
January 30, 1998
1 September 2006
the third quarter of 1984
Time Expression - Constituents

- **Time token**
  - Explicitly express time information
- **Modifier**
  - Modify time tokens
- **Numeral**
  - Numbers and ordinals (except year)

Today
Friday
September
Last week
2 years ago
September 2006
2006 September
January 30, 1998
1 September 2006
the third quarter of 1984
Time Expression - Constituents

- **Time token**
  - Explicitly express time information

- **Modifier**
  - Modify time tokens

- **Numeral**
  - Numbers and ordinals (except year)

Today
Friday
September
Last week
2 years ago
September 2006
2006 September
January 30, 1998
1 September 2006
the third quarter of 1984
Time Expression - Constituents

- **Time token**
  - Explicitly express time information

- **Modifier**
  - Modify time tokens

- **Numeral**
  - Numbers and ordinals (except year)

Today
Friday
September
Last week
2 years ago
September 2006
2006 September
January 30, 1998
1 September 2006
the third quarter of 1984
Time Expression Analysis

• Datasets
  • TimeBank
  • Gigaword
  • WikiWars
  • Tweets

• Findings
  • Loose structure
  • Differentiable
Time Expression Analysis - Datasets

• Datasets
  • TimeBank: a benchmark dataset used in TempEval evaluations
  • Gigaword: a large dataset with automatically generated labels
  • WikiWars: a war domain dataset collected from Wikipedia
  • Tweets: a tweet dataset collected from Twitter

• Dataset statistics

<table>
<thead>
<tr>
<th>Dataset</th>
<th>#Docs</th>
<th>#Words</th>
<th>#Timex</th>
</tr>
</thead>
<tbody>
<tr>
<td>TimeBank</td>
<td>183</td>
<td>61,418</td>
<td>1,243</td>
</tr>
<tr>
<td>Gigaword</td>
<td>2,452</td>
<td>666,309</td>
<td>12,739</td>
</tr>
<tr>
<td>WikiWars</td>
<td>22</td>
<td>119,468</td>
<td>2,671</td>
</tr>
<tr>
<td>Tweets</td>
<td>942</td>
<td>18,199</td>
<td>1,127</td>
</tr>
</tbody>
</table>

The datasets differ in size, source, domain, and text type, but their time expressions demonstrate similar characteristics.
Time Expression Analysis - Finding 1

- **Loose structure**: time expressions are formed by loose structure
Time Expression Analysis - Finding 1

- **Loose structure**: time expressions are formed by loose structure
  - Loose collocation
    - September
    - September 2006
    - 1 September 2006
  - Exchangeable order
    - September 2006
    - 2006 September
Time Expression Analysis - Finding 1

- **Loose structure**: time expressions are formed by loose structure
  - Loose collocation
    - September/B
    - September/B 2006/I
    - 1/B September/I 2006/I
  - Exchangeable order
    - September/B 2006/I
    - 2006/B September/I

In perspective of position within time expressions, under BIO scheme, ‘September’ may appear as (i) **Beginning** or (ii) **Inside** word of time expressions.
Time Expression Analysis - Finding 1

- **Loose structure**: time expressions are formed by loose structure
  - Loose collocation
    - September/U
    - September/B 2006/L
    - 1/B September/I 2006/L
  - Exchangeable order
    - September/B 2006/L
    - 2006/B September/L

Under BILOU scheme, ‘September’ may appear as
(1) Unit-word time expressions, (2) Beginning, (3) Inside, or (4) Last word of time expressions
Time Expression Analysis - Finding 1

- **Loose structure**: time expressions are formed by loose structure

Percentage of distinct time tokens and modifiers that appear in _different positions_ within time expressions

<table>
<thead>
<tr>
<th>Dataset</th>
<th>BIO Scheme</th>
<th></th>
<th></th>
<th>BILOU Scheme</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time Token</td>
<td>Modifier</td>
<td></td>
<td>Time Token</td>
<td>Modifier</td>
<td></td>
</tr>
<tr>
<td>TimeBank</td>
<td>58.18</td>
<td>33.33</td>
<td></td>
<td>63.64</td>
<td>33.33</td>
<td></td>
</tr>
<tr>
<td>Gigaword</td>
<td>61.29</td>
<td>45.83</td>
<td></td>
<td>77.05</td>
<td>46.00</td>
<td></td>
</tr>
<tr>
<td>WikiWars</td>
<td>53.57</td>
<td>26.19</td>
<td></td>
<td>61.40</td>
<td>29.55</td>
<td></td>
</tr>
<tr>
<td>Tweets</td>
<td>67.21</td>
<td>27.59</td>
<td></td>
<td>72.58</td>
<td>27.59</td>
<td></td>
</tr>
</tbody>
</table>
Time Expression Analysis - Finding 2

- **Differentiable**: time tokens can differentiate time expressions from common text
Time Expression Analysis - Finding 2

- **Differentiable**: time tokens can differentiate time expressions from common text

Percentage of time expression’s constituents that appear in time expressions ($P_{timex}$) and in common text ($P_{text}$)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Time Token</th>
<th></th>
<th>Modifier</th>
<th></th>
<th>Numeral</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P_{timex}$</td>
<td>$P_{text}$</td>
<td>$P_{timex}$</td>
<td>$P_{text}$</td>
<td>$P_{timex}$</td>
<td>$P_{text}$</td>
</tr>
<tr>
<td>TimeBank</td>
<td>94.61</td>
<td>0.34</td>
<td>47.39</td>
<td>22.56</td>
<td>22.61</td>
<td>3.16</td>
</tr>
<tr>
<td>Gigaword</td>
<td>96.44</td>
<td>0.65</td>
<td>28.05</td>
<td>22.82</td>
<td>20.24</td>
<td>2.03</td>
</tr>
<tr>
<td>WikiWars</td>
<td>91.81</td>
<td>0.14</td>
<td>31.64</td>
<td>26.14</td>
<td>38.01</td>
<td>9.82</td>
</tr>
<tr>
<td>Tweets</td>
<td>96.01</td>
<td>0.50</td>
<td>21.38</td>
<td>13.03</td>
<td>18.81</td>
<td>0.128</td>
</tr>
</tbody>
</table>

$$P_{timex}(T) = \frac{\# \text{timex that contain } T}{\# \text{total timex}}$$  $$P_{text}(T) = \frac{\# \text{tokens that are } T}{\# \text{total tokens}}$$
Fundamental Problem - Inconsistent Tag Assignment

• Position-based tagging scheme
  • BIO scheme: Beginning or Inside word of time expressions, Outside time expressions
  • BILOU scheme: Unit-word time expressions, Beginning, Inside, Last word of multi-word time expressions, Outside time expressions

• Inconsistent tag assignment
  • During training, a word is assigned with different tags simply because the word appears in different positions within labeled chunks
Inconsistent Tag Assignment

• Position-based tagging scheme
  • BILOU scheme: Unit-word time expressions, Beginning, Inside, Last word of multi-word time expressions, Outside time expressions
Inconsistent Tag Assignment

• Position-based tagging scheme
  • BILOU scheme: Unit-word time expressions, Beginning, Inside, Last word of multi-word time expressions, Outside time expressions

1) September/U  2) September/B 2006/L  
3) 2006/B September/L  4) 1/B September/I 2006/L
Inconsistent Tag Assignment

• Position-based tagging scheme
  • BILOU scheme: Unit-word time expressions, Beginning, Inside, Last word of multi-word time expressions, Outside time expressions

1) September/U  2) September/B 2006/L
3) 2006/B September/L  4) 1/B September/I 2006/L

1) (…, w=September, …, U)
2) (…, w=September, …, B)
3) (…, w=September, …, L)
4) (…, w=September, …, I)
Inconsistent Tag Assignment

- Position-based tagging scheme
  - BILOU scheme: Unit-word time expressions, Beginning, Inside, Last word of multi-word time expressions, Outside time expressions

1) September/U  
2) September/B 2006/L  
3) 2006/B September/L  
4) 1/B September/I 2006/L

1) (…, w=September, …, U)  
2) (…, w=September, …, B)  
3) (…, w=September, …, L)  
4) (…, w=September, …, I)

Inconsistent tag assignment reduces the predictive power of ‘September’

This contradicts finding 2 that time tokens can differentiate time expressions from common text.
Review

• Two findings
  • Finding 1: time expressions are formed by loose structure
  • Finding 2: time tokens can differentiate time expressions from common text

• Finding 1 leads BILOU scheme to inconsistent tag assignment
  • Reduce the predictive power of time tokens

1) September/U  2) September/B 2006/L
3) 2006/B September/L  4) 1/B September/I 2006/L

• Under BILOU scheme, Finding 1 contradicts Finding 2
Overcome Inconsistent Tag Assignment

• Constituent-based tagging scheme
  • TOMN scheme: Time token, Modifier, Numeral, Outside time expressions
  • TOMN scheme assigns a word with a tag according to its constituent role
Overcome Inconsistent Tag Assignment

• Constituent-based tagging scheme
  • TOMN scheme: Time token, Modifier, Numeral, Outside time expressions
  • TOMN scheme assigns a word with a tag according to its constituent role

1) September/T 2) September/T 2006/T
3) 2006/T September/T 4) 1/N September/T 2006/T
Overcome Inconsistent Tag Assignment

• Constituent-based tagging scheme
  • TOMN scheme: Time token, Modifier, Numeral, Outside time expressions
  • TOMN scheme assigns a word with a tag according to its constituent role

1) September/T   2) September/T 2006/T
3) 2006/T September/T   4) 1/N September/T 2006/T

1) (..., w=September, ..., T)
2) (..., w=September, ..., T)
3) (..., w=September, ..., T)
4) (..., w=September, ..., T)
Overcome Inconsistent Tag Assignment

• Constituent-based tagging scheme
  • TOMN scheme: Time token, Modifier, Numeral, Outside time expressions
  • TOMN scheme assigns a word with a tag according to its constituent role

1) September/T 2) September/T 2006/T
3) 2006/T September/T 4) 1/N September/T 2006/T

1) (…, w=September, …, T) 2) (…, w=September, …, T) Consistent tag assignment protects
3) (…, w=September, …, T) 4) (…, w=September, …, T) ‘September’s predictive power
Time Expression Recognition - TOMN

TOMN scheme

A set of token regular expressions

TOMN scheme

T (time token)
M (modifier)
N (numeral)
O (outside timex)

TOMN Scheme

Time Token
Modifier
Numeral

TmnRegex

Feature Extractor

Raw Text

TOMN Pre-tag Features
Lemma Features

CRFs-based Tagger

Annotated Text

CRFs framework
Time Expression Recognition - Examples

• Non-O words that appear together form a time expression

On/o September/T 1/N ,/M 1939/T ,/O … state/O in/o 1939/T ./O

... in/o a/M few/M days/T and/M weeks/T respectively/O ./O
Time Expression Recognition - Experiments

• Our method: TOMN

• Baselines
  • HeidelTime: rule-based
  • SUTime: rule-based
  • SynTime: type-based
  • ClearTK: learning-based
  • UWTime: learning-based

• Datasets
  • TE-3, WikiWars, Tweets
Performance of TOMN and baselines. **Best results** are in boldface and **second best** are underlined.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Method</th>
<th>Strict Match</th>
<th></th>
<th></th>
<th>Relaxed Match</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Pr.</td>
<td>Re.</td>
<td>F1</td>
<td>Pr.</td>
<td>Re.</td>
<td>F1</td>
</tr>
<tr>
<td>TE-3</td>
<td>HeidelTime (Strotgen et al., 2013)</td>
<td>83.85</td>
<td>78.99</td>
<td>81.34</td>
<td>93.08</td>
<td>87.68</td>
<td>90.30</td>
</tr>
<tr>
<td></td>
<td>SUTime (Chang and Manning, 2013)</td>
<td>78.72</td>
<td>80.43</td>
<td>79.57</td>
<td>89.36</td>
<td>91.30</td>
<td>90.32</td>
</tr>
<tr>
<td></td>
<td>SynTime (Zhong et al., 2017)</td>
<td>91.43</td>
<td><strong>92.75</strong></td>
<td>92.09</td>
<td>94.29</td>
<td><strong>95.65</strong></td>
<td><strong>94.96</strong></td>
</tr>
<tr>
<td></td>
<td>ClearTK (Bethard, 2013)</td>
<td>85.90</td>
<td>79.70</td>
<td>82.70</td>
<td>93.75</td>
<td>86.96</td>
<td>90.23</td>
</tr>
<tr>
<td></td>
<td>UWTime (Lee et al., 2014)</td>
<td>86.10</td>
<td>80.40</td>
<td>83.10</td>
<td>94.60</td>
<td>88.40</td>
<td>91.40</td>
</tr>
<tr>
<td></td>
<td><strong>TOMN</strong></td>
<td><strong>92.59</strong></td>
<td>90.58</td>
<td>91.58</td>
<td><strong>95.56</strong></td>
<td>93.48</td>
<td>94.51</td>
</tr>
<tr>
<td>WikiWars</td>
<td>HeidelTime (Strotgen et al., 2013)</td>
<td><strong>88.20</strong></td>
<td>78.50</td>
<td>83.10</td>
<td>95.80</td>
<td>85.40</td>
<td>90.30</td>
</tr>
<tr>
<td></td>
<td>SUTime</td>
<td>78.61</td>
<td>76.69</td>
<td>76.64</td>
<td>95.74</td>
<td>89.57</td>
<td>92.55</td>
</tr>
<tr>
<td></td>
<td>SynTime (Zhong et al., 2017)</td>
<td>80.00</td>
<td>80.22</td>
<td>80.11</td>
<td>92.16</td>
<td><strong>92.41</strong></td>
<td>92.29</td>
</tr>
<tr>
<td></td>
<td>ClearTK</td>
<td>87.69</td>
<td><strong>80.28</strong></td>
<td><strong>83.82</strong></td>
<td>96.80</td>
<td>90.54</td>
<td>93.56</td>
</tr>
<tr>
<td></td>
<td>UWTime (Lee et al., 2014)</td>
<td>87.70</td>
<td>78.80</td>
<td>83.00</td>
<td><strong>97.60</strong></td>
<td>87.60</td>
<td>92.30</td>
</tr>
<tr>
<td></td>
<td><strong>TOMN</strong></td>
<td>84.57</td>
<td><strong>80.48</strong></td>
<td>82.47</td>
<td>96.23</td>
<td><strong>92.35</strong></td>
<td><strong>94.25</strong></td>
</tr>
<tr>
<td>Tweets</td>
<td>HeidelTime</td>
<td><strong>91.67</strong></td>
<td>74.26</td>
<td>82.05</td>
<td>96.88</td>
<td>78.48</td>
<td>86.71</td>
</tr>
<tr>
<td></td>
<td>SUTime</td>
<td>77.69</td>
<td>79.32</td>
<td>78.50</td>
<td>88.84</td>
<td>90.72</td>
<td>89.77</td>
</tr>
<tr>
<td></td>
<td>SynTime (Zhong et al., 2017)</td>
<td>89.52</td>
<td><strong>94.07</strong></td>
<td>91.74</td>
<td>93.55</td>
<td><strong>98.31</strong></td>
<td><strong>95.87</strong></td>
</tr>
<tr>
<td></td>
<td>ClearTK</td>
<td>86.83</td>
<td>75.11</td>
<td>80.54</td>
<td>96.59</td>
<td>83.54</td>
<td>89.59</td>
</tr>
<tr>
<td></td>
<td>UWTime</td>
<td>88.36</td>
<td>70.76</td>
<td>78.59</td>
<td><strong>97.88</strong></td>
<td>78.39</td>
<td>87.06</td>
</tr>
<tr>
<td></td>
<td><strong>TOMN</strong></td>
<td>90.69</td>
<td><strong>94.51</strong></td>
<td><strong>92.56</strong></td>
<td>93.52</td>
<td>97.47</td>
<td>95.45</td>
</tr>
</tbody>
</table>
Performance of TOMN and baselines. **Best results** are in boldface and **second best** are underlined.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Method</th>
<th>Strict Match</th>
<th></th>
<th>Relexed Match</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Pr.</td>
<td>Re.</td>
<td>F1</td>
<td>Pr.</td>
</tr>
<tr>
<td>TE-3</td>
<td>HeidelTime (Strotgen et al., 2013)</td>
<td>83.85</td>
<td>78.99</td>
<td>81.34</td>
<td>93.08</td>
</tr>
<tr>
<td></td>
<td>SUTime (Chang and Manning, 2013)</td>
<td>78.72</td>
<td>80.43</td>
<td>79.57</td>
<td>89.36</td>
</tr>
<tr>
<td></td>
<td>SynTime (Zhong et al., 2017)</td>
<td>91.43</td>
<td><strong>92.75</strong></td>
<td><strong>92.09</strong></td>
<td>94.29</td>
</tr>
<tr>
<td></td>
<td>ClearTK (Bethard, 2013)</td>
<td>85.90</td>
<td>79.70</td>
<td>82.70</td>
<td>93.75</td>
</tr>
<tr>
<td></td>
<td>UWTime (Lee et al., 2014)</td>
<td>86.10</td>
<td>80.40</td>
<td>83.10</td>
<td><strong>94.60</strong></td>
</tr>
<tr>
<td></td>
<td>TOMN</td>
<td><strong>92.59</strong></td>
<td>90.58</td>
<td><strong>91.58</strong></td>
<td><strong>95.56</strong></td>
</tr>
<tr>
<td>WikiWars</td>
<td>HeidelTime (Strotgen et al., 2013)</td>
<td>88.20</td>
<td>78.50</td>
<td>83.10</td>
<td>95.80</td>
</tr>
<tr>
<td></td>
<td>SUTime</td>
<td>78.61</td>
<td>76.69</td>
<td>76.64</td>
<td>95.74</td>
</tr>
<tr>
<td></td>
<td>SynTime (Zhong et al., 2017)</td>
<td>80.00</td>
<td>80.22</td>
<td>80.11</td>
<td>92.16</td>
</tr>
<tr>
<td></td>
<td>ClearTK</td>
<td>87.69</td>
<td>80.28</td>
<td><strong>83.82</strong></td>
<td>96.80</td>
</tr>
<tr>
<td></td>
<td>UWTime (Lee et al., 2014)</td>
<td>87.70</td>
<td>78.80</td>
<td>83.00</td>
<td><strong>97.60</strong></td>
</tr>
<tr>
<td></td>
<td>TOMN</td>
<td>84.57</td>
<td><strong>80.48</strong></td>
<td>82.47</td>
<td>96.23</td>
</tr>
<tr>
<td>Tweets</td>
<td>HeidelTime</td>
<td>91.67</td>
<td>74.26</td>
<td>82.05</td>
<td>96.88</td>
</tr>
<tr>
<td></td>
<td>SUTime</td>
<td>77.69</td>
<td>79.32</td>
<td>78.50</td>
<td>88.84</td>
</tr>
<tr>
<td></td>
<td>SynTime (Zhong et al., 2017)</td>
<td>89.52</td>
<td>94.07</td>
<td><strong>91.74</strong></td>
<td>93.55</td>
</tr>
<tr>
<td></td>
<td>ClearTK</td>
<td>86.83</td>
<td>75.11</td>
<td>80.54</td>
<td>96.59</td>
</tr>
<tr>
<td></td>
<td>UWTime</td>
<td>88.36</td>
<td>70.76</td>
<td>78.59</td>
<td><strong>97.88</strong></td>
</tr>
<tr>
<td></td>
<td>TOMN</td>
<td>90.69</td>
<td><strong>94.51</strong></td>
<td><strong>92.56</strong></td>
<td>93.52</td>
</tr>
</tbody>
</table>
Time Expression Recognition - Experiments

• Cross-dataset experiments
  • Train on one dataset’s training set
  • Test on other datasets’ test sets

• Learning-based baselines
  • ClearTK
  • UWTime

• Datasets
  • TE-3
  • WikiWars
  • Tweets
Cross-dataset performance on test set of TE-3
(Color background indicates single-dataset results)

| Training Set | Method | Strict Match | | | Relexed Match | | |
|--------------|--------|--------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
|              |        | Pr.  | Re.  | F1  | Pr.  | Re.  | F1  | | | | |
| TE-3         | ClearTK | 85.90 | 79.70 | 82.70 | 93.75 | 86.96 | 90.23 | | | | |
|              | UWTIme | 86.10 | 80.40 | 83.10 | 94.60 | 88.40 | 91.40 | | | | |
|              | TOMN   | **92.59** | **90.58** | **91.58** | **95.56** | **93.48** | **94.51** | | | | |
| WikiWars     | ClearTK | 65.67 | 63.77 | 64.71 | 87.31 | 84.78 | 86.03 | | | | |
|              | UWTIme | 76.92 | 72.46 | 74.63 | 88.46 | 83.33 | 85.82 | | | | |
|              | TOMN   | **84.06** | **84.06** | **84.06** | **93.48** | **93.48** | **93.48** | | | | |
| Tweets       | ClearTK | 72.59 | 71.01 | 71.79 | **93.33** | 91.30 | 92.31 | | | | |
|              | UWTIme | 80.00 | 72.46 | 76.05 | 92.80 | 84.06 | 88.21 | | | | |
|              | TOMN   | **85.42** | **89.13** | **87.23** | 91.67 | **95.65** | **93.62** | | | | |
Cross-dataset performance on test set of WikiWars

<table>
<thead>
<tr>
<th>Training Set</th>
<th>Method</th>
<th>Strict Match</th>
<th></th>
<th></th>
<th>Relexed Match</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Pr.</td>
<td>Re.</td>
<td>F1</td>
<td>Pr.</td>
<td>Re.</td>
<td>F1</td>
</tr>
<tr>
<td>TE-3</td>
<td>ClearTK</td>
<td>74.38</td>
<td>60.76</td>
<td>66.89</td>
<td>97.54</td>
<td>79.68</td>
<td>87.71</td>
</tr>
<tr>
<td></td>
<td>UWTIme</td>
<td>87.01</td>
<td>79.34</td>
<td>83.00</td>
<td>96.07</td>
<td>87.60</td>
<td>91.64</td>
</tr>
<tr>
<td></td>
<td>TOMN</td>
<td>82.18</td>
<td>75.65</td>
<td>79.07</td>
<td>96.26</td>
<td>87.93</td>
<td>91.90</td>
</tr>
<tr>
<td>WikiWars</td>
<td>ClearTK</td>
<td>87.69</td>
<td>80.28</td>
<td>83.82</td>
<td>96.80</td>
<td>90.54</td>
<td>93.56</td>
</tr>
<tr>
<td></td>
<td>UWTIme</td>
<td>87.70</td>
<td>78.80</td>
<td>83.00</td>
<td>97.60</td>
<td>87.60</td>
<td>92.30</td>
</tr>
<tr>
<td></td>
<td>TOMN</td>
<td>84.57</td>
<td>80.48</td>
<td>82.47</td>
<td>96.23</td>
<td>92.35</td>
<td>94.25</td>
</tr>
<tr>
<td>Tweets</td>
<td>ClearTK</td>
<td>57.75</td>
<td>54.73</td>
<td>56.20</td>
<td>91.93</td>
<td>87.12</td>
<td>89.46</td>
</tr>
<tr>
<td></td>
<td>UWTIme</td>
<td>80.28</td>
<td>62.81</td>
<td>70.48</td>
<td>94.37</td>
<td>73.83</td>
<td>82.84</td>
</tr>
<tr>
<td></td>
<td>TOMN</td>
<td>60.29</td>
<td>66.00</td>
<td>63.02</td>
<td>84.74</td>
<td>92.76</td>
<td>88.57</td>
</tr>
</tbody>
</table>
Cross-dataset performance on test set of Tweets

<table>
<thead>
<tr>
<th>Training Set</th>
<th>Method</th>
<th>Strict Match</th>
<th></th>
<th>Relexed Match</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Pr.</td>
<td>Re.</td>
<td>F1</td>
</tr>
<tr>
<td>TE-3</td>
<td>ClearTK</td>
<td>81.16</td>
<td>47.26</td>
<td>59.73</td>
</tr>
<tr>
<td></td>
<td>UWTIme</td>
<td>89.66</td>
<td>65.82</td>
<td>75.91</td>
</tr>
<tr>
<td></td>
<td>TOMN</td>
<td><strong>92.92</strong></td>
<td><strong>88.61</strong></td>
<td><strong>90.71</strong></td>
</tr>
<tr>
<td>WikiWars</td>
<td>ClearTK</td>
<td>72.48</td>
<td>45.57</td>
<td>55.96</td>
</tr>
<tr>
<td></td>
<td>UWTIme</td>
<td><strong>87.43</strong></td>
<td>61.60</td>
<td>72.28</td>
</tr>
<tr>
<td></td>
<td>TOMN</td>
<td>85.00</td>
<td><strong>86.08</strong></td>
<td><strong>85.53</strong></td>
</tr>
<tr>
<td>Tweets</td>
<td>ClearTK</td>
<td>86.83</td>
<td>75.11</td>
<td>80.54</td>
</tr>
<tr>
<td></td>
<td>UWTIme</td>
<td>88.36</td>
<td>70.76</td>
<td>78.59</td>
</tr>
<tr>
<td></td>
<td>TOMN</td>
<td><strong>90.69</strong></td>
<td><strong>94.51</strong></td>
<td><strong>92.56</strong></td>
</tr>
</tbody>
</table>
Time Expression Recognition - Efficiency

• TOMN is more efficient

<table>
<thead>
<tr>
<th>Method</th>
<th>TE-3</th>
<th>WikiWars</th>
<th>Tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>ClearTK</td>
<td>152</td>
<td>223</td>
<td>86</td>
</tr>
<tr>
<td>UWTTime</td>
<td>864</td>
<td>1,050</td>
<td>160</td>
</tr>
<tr>
<td>TOMN</td>
<td>36</td>
<td>48</td>
<td>42</td>
</tr>
</tbody>
</table>

Runtime of going through a whole process (unit: seconds)
Summary

• Have two findings
  • Loose structure
  • Differentiable

• Reveal a fundamental problem in position-based tagging scheme
  • Inconsistent tag assignment

• Define a constituent-based tagging scheme
  • Good results
  • Less time