TEMPORAL EXPRESSION AND EVENT EXTRACTION USING GENERAL CONDITIONAL RANDOM FIELDS

An Experimental Evaluation

Presented by Ankush Israney & Jaidev Ramakrishna for CS 613

Original Objectives

- Our initial proposal considered the usage of a Deep Neural Network to extract events and temporal expressions.
- We found recent research with an identical line of inquiry.

  V.R Chikka claims in "Extraction of Temporal Information from Clinical documents using Machine Learning techniques" :- *"The results show that both approaches (give) relatively same performance on the provided train and test datasets of the challenge."*

Table 2: Phase 1 Evaluation on events, times and temporal relations

	Approach 1 (CRF and SVM)			Approach 2 (DNN)		
Task	**Precision**	**Recall**	**F-Score**	**Precision**	**Recall**	**F-Score**
EVENT:<span> (ES)	0.835	0.797	0.815	0.838	0.786	0.811
EVENT:Modality	0.764	0.729	0.746	0.779	0.731	0.754
EVENT:Degree	0.830	0.793	0.811	0.834	0.783	0.807
EVENT:Polarity	0.750	0.716	0.733	0.813	0.764	0.788
EVENT:Type	0.806	0.769	0.787	0.814	0.765	0.789
TIMEX3:<span> (TS)	0.752	0.515	0.612	0.614	0.560	0.586
TIMEX3:Class	0.644	0.439	0.522	0.468	0.426	0.446
EVENT:DocTimeRel (DR)	0.481	0.460	0.470	0.643	0.604	0.623
TLINK:Type (CR)	0.431	0.167	0.241	0.285	0.225	0.252

Project Objectives

- Many researchers have produced statistical results in favor of employing Machine Learning techniques in the following tasks of the TempEval Competition Series.
  - *Time span (TS) identification*
  - *Event span (ES) identification*

- Therefore, we use CRF classifiers to extract events and temporal expressions in the News domain. (TempEval3 – 2013 publicly available dataset)
- Additionally, we perform a detailed quantitative and qualitative analysis of these tasks.
- We have built an extensible system that integrates existing NLP frameworks, which can be used on a variety of domains (annotated in TimeML).

Conditional Random Fields

- CRFs are a type of <u>discriminative undirected probabilistic graphical model</u>. It is used to encode known relationships between observations and construct consistent interpretations. (Sounds like HMM!, but different – conditioned on input and features)
- Popular for pattern recognition tasks, especially Named Entity Recognition in NLP.
- CRF's exploit the concept of Feature Functions:

  1 feature for each state transition and a 1 feature for each state-observation pair.
- Important Distinction from generative models like HMM (Discriminative: More like logistic regression here, Is that why it performs like DNN?)
- In it's most intuitive explanation, In the NER task, the label for a word may not depend on its previous word but could be relevant to any word in the sentence.

Conditional Random Fields

- Feature function $$f_k(y, y_{y-1}, x_t)$$

- Linear CRF $$p(y,x) = \frac{1}{Z} \prod_{t=1}^{T} exp(\sum_{k=1}^{K} \theta_k f_k(y_t, y_{t-1}, x_t))$$

- General CRF $$p(y,x) = \frac{1}{Z(x)} \prod_{t=1}^{T} exp(\sum_{k=1}^{K} \theta_{ak} f_{ak}(y_a, x_a))$$

Input Features

- useClassFeature - Include a feature for the class (as a class marginal). Puts a prior on the classes which is equivalent to how often the feature appeared in the training data.
- useWord= Gives Feature for w
- **useNGrams - Make features from letter n-grams, i.e., substrings of the word**
- **noMidNGrams - Do not include character n-gram features for n-grams that contain neither the beginning or end of the word**
- **maxNGramLeng - If this number is positive, n-grams above this size will not be used in the model**
- usePrev - Gives you feature for (pw,c), and together with other options enables other previous features, such as (pt,c) [with useTags)
- useNext - Gives you feature for (nw,c), and together with other options enables other next features, such as (nt,c) [with useTags)

Input Features

- useDisjunctive - Include in features giving disjunctions of words anywhere in the left or right disjunctionWidth words (preserving direction but not position)
- useSequences - Does not use any class combination features if this is false
- usePrevSequences - Does not use any class combination features using previous classes if this is false
- useTypeSeqs - Use basic zeroeth order word shape features.
- useTypeSeqs2 - Add additional first and second order word shape features
- useTypeySequences - Some first order word shape patterns.
- wordShape - Either "none" for no wordShape use, or the name of a word shape function.

Data Sets and Preprocessing

- Transcripts of news pieces from the BBC, Associated Press, New York Times etc.
- Training set
  - 2452 documents annotated in TimeML format

```
<TITLE>Malaysian share market falls 1.1 percent at
midday</TITLE>


<TEXT>

Malaysian share prices <EVENT class="OCCURRENCE" eid="e1">
dipped</EVENT> 1.1 percent by midday <TIMEX3 type="DATE"
value="1997-04-01" tid="t1">Tuesday</TIMEX3>, with the market
barometer <EVENT class="OCCURRENCE" eid="e2">falling</EVENT>
below the 1,200 support level for the first time in <TIMEX3
type="DURATION" value="P3M" tid="t2">three months</TIMEX3>
following credit curbs and the Dow's fall.
```

- We convert this to .col format using the open source TimeMLtoColumns converter developed by Paramita Mirza, a PhD student at the Max Planck Institute.
  - https://github.com/paramitamirza/TimeML-CAT-Converter

Data Sets and Preprocessing

- Testing Set – 20 documents manually annotated.

```
<TITLE>105 U.S. Kids Died From Flu, CDC Says</TITLE>


<TEXT>

The flu season is winding down, and it has killed 105 children
so far - about the average toll.

The season started about a month earlier than usual, sparking
concerns it might turn into the worst in a decade. It ended up
being very hard on the elderly, but was moderately severe
overall, according to the Centers for Disease Control and
Prevention.

Six of the pediatric deaths were reported in the last week, and
it's possible there will be more, said the CDC's Dr. Michael
Jhung said Friday.
```

- Converted to tokens using our own preprocessor and supplied to CRF Classifier.

Evaluation Method

- Token-wise evaluation (not “strict” like the TempEval3 validator).
- We provide true annotated token files to the Stanford NER CRF Classifier to test it against our trained model.
- Stanford CRF Classifier compares its predictions to the true annotations and computes all relevant statistics for 3 classes.
  - *Events*
  - *TimeX3*
  - *Others*

Word	Actual	Classification
So	OTHERS	OTHERS
when	OTHERS	OTHERS
the	OTHERS	OTHERS
president	OTHERS	OTHERS
of	OTHERS	OTHERS
Cyprus	OTHERS	OTHERS
admitted	EVENT	OTHERS
this	TIMEX3	OTHERS
week	TIMEX3	TIMEX3
that	OTHERS	OTHERS
his	OTHERS	OTHERS
country	OTHERS	OTHERS
did	OTHERS	OTHERS
not	OTHERS	OTHERS
have	EVENT	EVENT
the	OTHERS	OTHERS
money	OTHERS	OTHERS
to	OTHERS	OTHERS
backstop	OTHERS	OTHERS
the	OTHERS	OTHERS
30	OTHERS	OTHERS
billion	OTHERS	OTHERS
euros	OTHERS	OTHERS
of	OTHERS	OTHERS
guaranteed	OTHERS	OTHERS
bank	OTHERS	OTHERS
deposits	OTHERS	OTHERS

Sample Output & Statistics

Entity	P	R	F1	TP	FP	FN
Event	0.8115	0.7715	0.7910	1128	262	334
Other	0.5999	0.5532	0.5756	952	635	769
TimeX3	0.9378	0.7101	0.8082	196	13	80
Total	0.7144	0.6580	0.6850	2276	910	1183

```
Six	OTHERS
of	OTHERS
the	OTHERS
pediatric	OTHERS
deaths	OTHERS
were	OTHERS
reported	EVENT
in	OTHERS
the	TIMEX3
last	TIMEX3
week	TIMEX3
,	OTHERS
and	OTHERS
it	OTHERS
's	OTHERS
possible	OTHERS
there	OTHERS
will	OTHERS
be	OTHERS
more	OTHERS
,	OTHERS
said	EVENT
the	OTHERS
CDC	OTHERS
's	OTHERS
Dr.	OTHERS
Michael	OTHERS
Jhung	OTHERS
said	EVENT
Friday	TIMEX3
.	OTHERS
```

Common Instances of TP, FP, FN

TIME EXPRESSIONS

TP			
Friday	TIMEX3	TIMEX3	TRUE
last	TIMEX3	TIMEX3	TRUE
May	TIMEX3	TIMEX3	TRUE
last	TIMEX3	TIMEX3	TRUE
year	TIMEX3	TIMEX3	TRUE
week	TIMEX3	TIMEX3	TRUE
FN			
2010	TIMEX3	OTHERS	FALSE
2008	TIMEX3	OTHERS	FALSE
eight	TIMEX3	OTHERS	FALSE
years	TIMEX3	OTHERS	FALSE
now	TIMEX3	OTHERS	FALSE
FP			
.	OTHERS	TIMEX3	FALSE

EVENTS

TP				
accepted	EVENT	EVENT	TRUE	
negotiations	EVENT	EVENT	TRUE	
FP/FN				
compared	OTHERS	EVENT		FALSE
havoc	EVENT	OTHERS		FALSE

FP/FN				
required	OTHERS	EVENT		FALSE
bailout	EVENT	OTHERS		FALSE

Variations in Training Set Size

For 10:						
Entity	P	R	F1	TP	FP	FN
EVENT	0.7929	0.2435	0.3726	356	93	1106
OTHERS	0.2202	0.0633	0.0984	109	386	1612
TIMEX3	0.875	0.1014	0.1818	28	4	248
Totals	0.5051	0.1425	0.2223	493	483	296

For 50:						
Entity	P	R	F1	TP	FP	FN
EVENT	0.7904	0.5157	0.6242	754	200	708
OTHERS	0.4246	0.2766	0.335	476	645	1245
TIMEX3	0.8882	0.5181	0.6545	143	18	133
Totals	0.614	0.3969	0.4822	1373	863	2086

For 100:						
Entity	P	R	F1	TP	FP	FN
EVENT	0.793	0.5923	0.6782	866	226	596
OTHERS	0.468	0.3481	0.3992	599	681	1122
TIMEX3	0.8866	0.6232	0.7319	172	22	104
Totals	0.638	0.4733	0.5434	1637	929	1822

For 200:						
Entity	P	R	F1	TP	FP	FN
EVENT	0.804	0.6676	0.7294	976	238	486
OTHERS	0.5351	0.4346	0.4796	748	650	973
TIMEX3	0.9362	0.6377	0.7586	176	12	100
Totals	0.6786	0.5493	0.6071	1900	900	1559

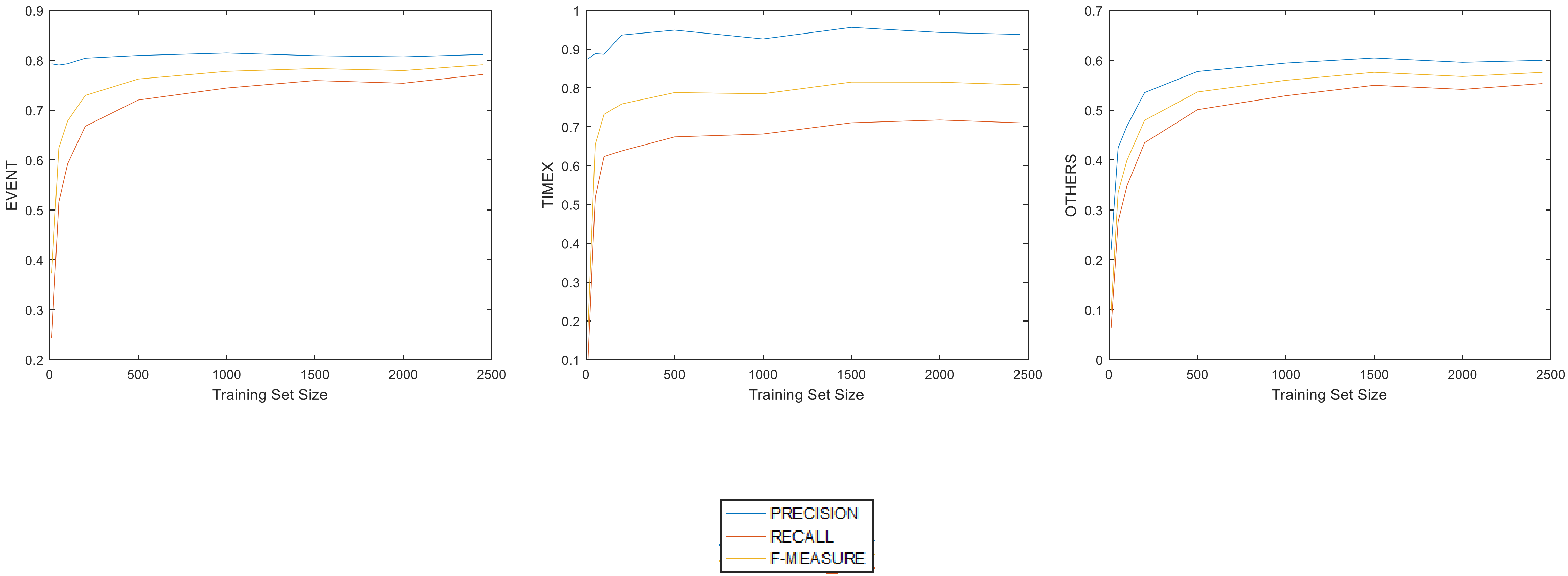
For 500:						
Entity	P	R	F1	TP	FP	FN
EVENT	0.8094	0.7202	0.7622	1053	248	409
OTHERS	0.5774	0.5009	0.5364	862	631	859
TIMEX3	0.949	0.6739	0.7881	186	10	90
Totals	0.7027	0.6074	0.6516	2101	889	1358

For 1000:						
Entity	P	R	F1	TP	FP	FN
EVENT	0.8144	0.7442	0.7777	1088	248	374
OTHERS	0.5944	0.5288	0.5597	910	621	811
TIMEX3	0.9261	0.6812	0.785	188	15	88
Totals	0.7121	0.632	0.6696	2186	884	1273

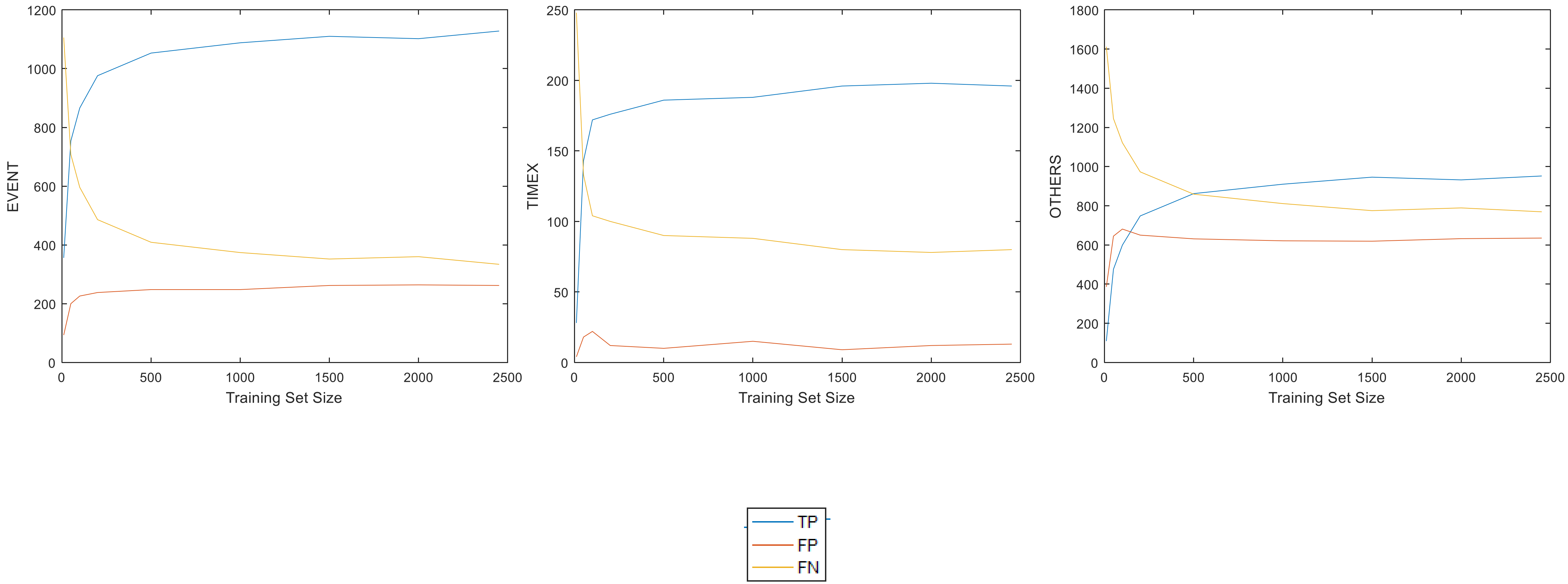
For 1500:						
Entity	P	R	F1	TP	FP	FN
EVENT	0.809	0.7592	0.7833	1110	262	352
OTHERS	0.6045	0.5497	0.5758	946	619	775
TIMEX3	0.9561	0.7101	0.815	196	9	80
Totals	0.7167	0.6511	0.6823	2252	890	1207

For 2000:						
Entity	P	R	F1	TP	FP	FN
EVENT	0.8067	0.7538	0.7793	1102	264	360
OTHERS	0.5959	0.5415	0.5674	932	632	789
TIMEX3	0.9429	0.7174	0.8148	198	12	78
Totals	0.7108	0.6453	0.6765	2232	908	1227

Graphs – Precision, Recall, F-Measure



Graphs – TP, FP, FN



Sample Run

python control.py -pre_train_skip -train_skip -test_skip –train_n <number>

```
File   Edit   View   Terminal   Tabs   Help
Converting data/TE3-Silver-data/TE3-Silver-data/XIN_ENG_20061109.0072.tml...
Converting data/TE3-Silver-data/TE3-Silver-data/XIN_ENG_20051110.0316.tml...
Converting data/TE3-Silver-data/TE3-Silver-data/XIN_ENG_20061129.0486.tml...
Converting data/TE3-Silver-data/TE3-Silver-data/AFP_ENG_20051216.0058.tml...
Converting data/TE3-Silver-data/TE3-Silver-data/XIN_ENG_20061122.0258.tml...
Converting data/TE3-Silver-data/TE3-Silver-data/AFP_ENG_20061201.0153.tml...
Converting data/TE3-Silver-data/TE3-Silver-data/XIN_ENG_20051112.0141.tml...
Converting data/TE3-Silver-data/TE3-Silver-data/XIN_ENG_20051125.0065.tml...
Converting data/TE3-Silver-data/TE3-Silver-data/XIN_ENG_20061110.0204.tml...
Converting data/TE3-Silver-data/TE3-Silver-data/XIN_ENG_20061103.0294.tml...
Converting data/TE3-Silver-data/TE3-Silver-data/XIN_ENG_20061130.0382.tml...
Converting data/TE3-Silver-data/TE3-Silver-data/XIN_ENG_20061109.0155.tml...
Converting data/TE3-Silver-data/TE3-Silver-data/XIN_ENG_20061119.0189.tml...
Converting data/TE3-Silver-data/TE3-Silver-data/XIN_ENG_20051114.0243.tml...
```

Preprocessing

```
File   Edit   View   Terminal   Tabs   Help
Iter 150 evals 171 <D> [M 1.000E0] 3.636E4 2430.38s |3.190E3| {1.574E-3} 1.603E-
Iter 151 evals 172 <D> [M 1.000E0] 3.568E4 2446.31s |1.615E3| {7.970E-4} 1.752E-
Iter 152 evals 173 <D> [2M 5.009E-1] 3.528E4 2478.54s |1.809E3| {8.929E-4} 1.662
Iter 153 evals 175 <D> [M 1.000E0] 3.475E4 2494.66s |1.329E3| {6.561E-4} 1.641E-
Iter 154 evals 176 <D> [M 1.000E0] 3.413E4 2509.36s |1.916E3| {9.459E-4} 1.655E-
Iter 155 evals 177 <D> [M 1.000E0] 3.367E4 2524.96s |2.532E3| {1.249E-3} 1.583E-
Iter 156 evals 178 <D> [M 1.000E0] 3.310E4 2540.74s |1.134E3| {5.598E-4} 1.512E-
Iter 157 evals 179 <D> [1M 4.698E-1] 3.278E4 2571.36s |2.595E3| {1.281E-3} 1.463
Iter 158 evals 181 <D> [M 1.000E0] 3.238E4 2586.15s |2.791E3| {1.378E-3} 1.383E-
Iter 159 evals 182 <D> [M 1.000E0] 3.201E4 2601.76s |1.403E3| {6.923E-4} 1.357E-
Iter 160 evals 183 <D> [M 1.000E0] 3.148E4 2617.41s |1.120E3| {5.527E-4} 1.334E-
Iter 161 evals 184 <D> [M 1.000E0] 3.128E4 2633.14s |3.170E3| {1.565E-3} 1.281E-
Iter 162 evals 185 <D> [M 1.000E0] 3.077E4 2648.81s |1.161E3| {5.731E-4} 1.293E-
Iter 163 evals 186 <D> [1M 2.596E-1] 3.063E4 2678.13s |3.201E3| {1.580E-3} 1.142
```

Training

```
File   Edit   View   Terminal   Tabs   Help
1106.0264.col,data/silver-col/inputCol/XIN_ENG_20051103.0054.col,data/silver-
ta/silver-col/inputCol/XIN_ENG_20061115.0208.col,data/silver-col/inputCol/XIN
putCol/XIN_ENG_20061127.0276.col,data/silver-col/inputCol/XIN_ENG_20051114.00
noMidNGrams=true
serializeTo=ner-model.ser.gz
maxNGramLeng=6
useNGrams=true
usePrev=true
useNext=true
maxLeft=1
map=word=0,answer=1
useWord=true
useTypeSeqs=true
[1000][2000]numFeatures = 841301
Time to convert docs to feature indices: 32.3 seconds
numClasses: 4 [0=O,1=OTHERS,2=EVENT,3=TIMEX3]
numDocuments: 2452
```

Feature Extraction

```
File   Edit   View   Terminal   Tabs   Help
Loading classifier from ner-model.ser.gz ... done [2.4 sec].
CRFClassifier tagged 341 words in 17 documents at 1409.09 words per second.
CRFClassifier invoked on Tue Dec 06 01:54:48 EST 2016 with arguments:
   -loadClassifier ner-model.ser.gz -testFile data/te3-platinum-col/inputCol/
testFile=data/te3-platinum-col/inputCol/
loadClassifier=ner-model.ser.gz
Loading classifier from ner-model.ser.gz ... done [2.4 sec].
CRFClassifier tagged 14018 words in 20 documents at 9395.44 words per second.
         Entity P      R      F1     TP     FP     FN
          EVENT 0.8115 0.7715 0.7910 1128   262    334
         OTHERS 0.5999 0.5532 0.5756 952    635    769
         TIMEX3 0.9378 0.7101 0.8082 196    13     80
         Totals 0.7144 0.6580 0.6850 2276   910    1183
```

Testing

Performance Comparison to Clinical Domain

- *Results Obtained from : Hitachi at SemEval-2016 Task 12: A Hybrid Approach for Temporal Information Extraction from Clinical Notes - Sarath P R, Manikandan R, Yoshiki Niwa*

TimeX3	P	R	F
Clinical	0.821	0.669	0.737
Ours	0.9378	0.7101	0.8082

- *Results Obtained from : CDE-IIITH at SemEval-2016 Task 12: Extraction of Temporal Information from Clinical documents using Machine Learning techniques - Veera Raghavendra Chikka*

Event	P	R	F
Clinical	0.835	0.797	0.815
Ours	0.8115	0.7715	0.791

Conclusions & Key Findings

- We integrated state-of-the-art Stanford NER system to work on the tempEval3 dataset news domain using general CRF classifier
- We performed an in-depth quantitative & qualitative analysis of the results for the Time & Event Extraction Tasks (TS and ES)
- The size of the training set affected FN (accepted more expressions) following a logarithmic decrease which is a mirror to the logarithmic increase in TP!
- The TP decrease although much slower followed to decrease with the increase in the Training Set size.
- The news domain was more conducive to the general CRF classifers in extraction of temporal information. - The news domain has more absolute expressions than the clinical domain! -
- The Clinical domain was better in the event extraction task due to the higher density of events and procedures as seen in such Corpora.

Future Work

- Normalization of values to the Identified temporal expressions.
- Identifying container relations between temporal and event expressions.
- Concentrate on the Duration Attribute of Temporal Expressions exploring novel ML techniques
- Open source our software under the GNU license.
- Sutton and McCallum suggest in A survey on Conditional Random Fields that “using HMM with CRF's in a hybrid approach are theoretically proven to give better f-measure values for the sequential Named Entity Recognition task” – We want to explore this line of thought as well!

Bibliography

- *A Cascaded Machine Learning Approach to Interpreting Temporal Expressions - David Ahn Joris van Rantwijk Maarten de Rijke*
- *An Introduction to Conditional Random Fields - Charles Sutton and Andrew McCallum*
- *Temporal Tagging on Different Domains: Challenges, Strategies, and Gold Standards -Jannik Str¨otgen, Michael Gertz*
- *Machine Learning Approaches for Temporal Information Extraction: A Comparative Study - Oleksandr Kolomiyets, Marie-Francine Moens*
- *Hitachi at SemEval-2016 Task 12: A Hybrid Approach for Temporal Information Extraction from Clinical Notes - Sarath P R1, Manikandan R1, Yoshiki Niwa2*
- *CDE-IIITH at SemEval-2016 Task 12: Extraction of Temporal Information from Clinical documents using Machine Learning techniques - Veera Raghavendra Chikka*
- *TempEval-3: Evaluating Events, Time Expressions, and Temporal Relations - Naushad UzZaman, Hector Llorensy, James Allen, Leon Derczynskiz, Marc Verhagen and James Pustejovsky*
- *A Corpus of Clinical Narratives Annotated with Temporal Information - Lucian Galescu Nate Blaylock*
- *Discovering Narrative Containers in Clinical Text - Timothy A. Miller1, Steven Bethard2, Dmitriy Dligach1, Sameer Pradhan1, Chen Lin1, and Guergana K. Savova1*

THANK YOU

Presented by Ankush Israney & Jaidev Ramakrishna for CS 613