Event Causality Detection using NLP techniques

Project Spring 2022

Supervisor(s): Raphaël Troncy, Pasquale Lisena

27 June 2022

EVENT DETECTION

- Detecting events and their relationships has become a crucial part of natural language processing.
- It has endless applications in the information research field while still being a highly challenging task.
- An event can be defined as something that happened. It can have many questions associated with it, for instance, when, where, how, by whom, etc.
- The idea of event detection is to identify the events and their relationships in unstructured data.

Introduction

Relationship Extraction $\subset$ Information Extraction

- Event : a course of action (expressed in natural language)
- Relation: connection within words (given the event)

- Types of semantic relations: Causal and Temporal

Example: Causal Relation

Event is Causal when there is an associated effect.

"Indian Ocean earthquake triggered a series of devastating tsunamis"

- **Information Extraction:** Identification of entities and relationship among them.

    Indian Ocean, earthquake, trigger, series, devastating, tsunamis

- **Relationship Extraction:** Classification of the relation and relation entity

    Cause-Effect, earthquake-tsunami

Example: Temporal Relation

Events are Temporal when they are related to time.

"I ate my food and then washed my hands."

- **Information Extraction:** ate, food, wash, hands

- **Temporal Extraction:** Event occur in a sequence,

  Eating of food happens before washing of hands

Pre-Existing Approaches and Issues

Relationship Extraction:

1. **Rule Based** (Non-Statistical) :
    - Manual Work to construct patten
    - Low recall and precision
2. **ML Based** (Statistical) :
    - relay on a large amount of label data
    - Requires sophisticated feature engineering
    - Require external NLP toolkits (error-prone)
    - Impractical for casual relation extraction due to linguistic complexities.

Proposed solution:
KNN and CATENA

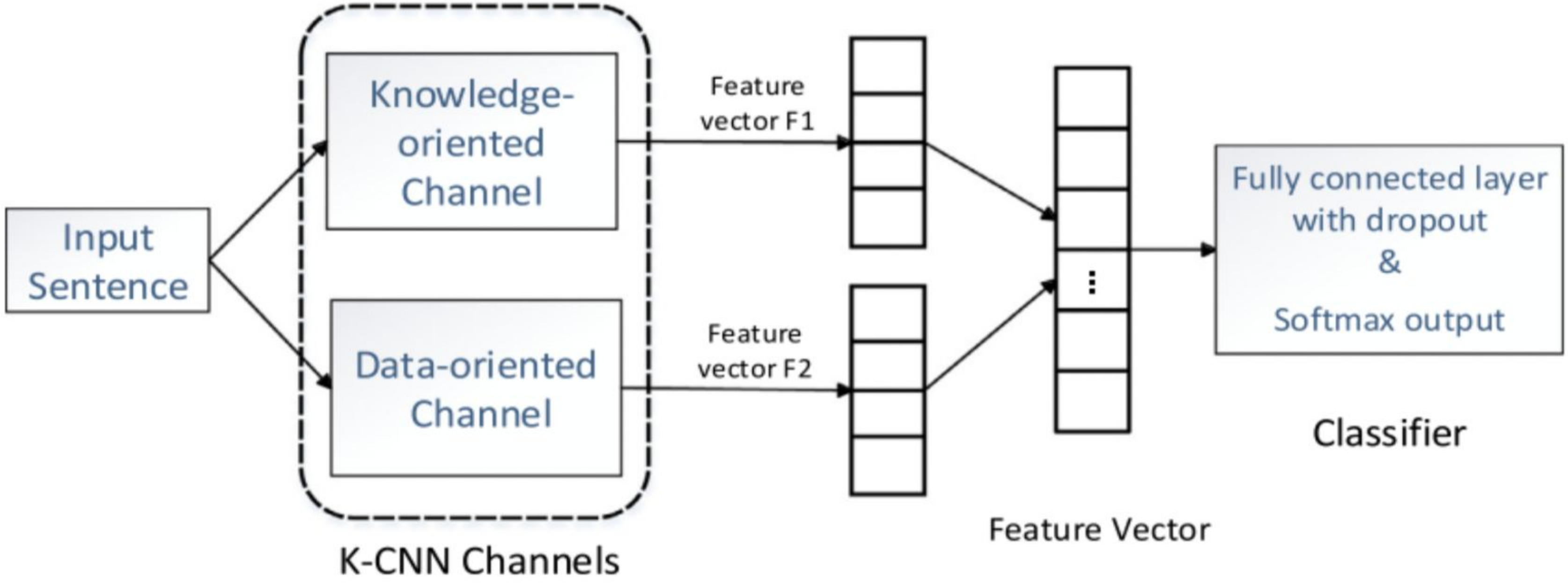
Sentence representation

Sentence: **The financial <e1>crisis<e1/> <u>results in</u> <e2>*unemployment*<e2/> in this country.**

- Target Entities: e1 and e2
- Cause: crisis
- Effect: unemployment
- Cue phrase: results in

K-CNN
Knowledge-Oriented Convolution Neural Network

Architecture



Architecture (Pengfei Li Kezhi Mao, 2019)

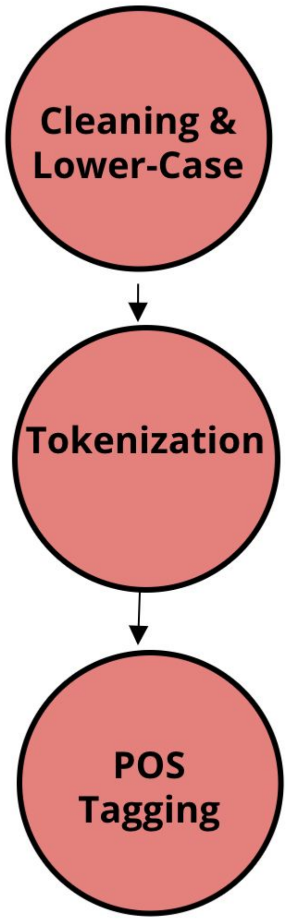
Flow Chart:

1. Pre-Processing
2. Modeling of Neural Network : Data-oriented Channel
3. Knowledge Embedding

Flow Chart:

1. **Pre-processing**

- Cleaning: removal of extra space, additional characters
- Lower-case
- Tokenization: break sentence into words
- Part-of-Speech tagging
- Created Answer-key file
- Created Word data-set
- Split the train-validation data



```
1	"The system as described above has its greatest application in an arrayed <e1>configuration</e1> of antenna
<e2>elements</e2>."
Component-Whole(e2,e1)
Comment: Not a collection: there is structure here, organisation.

2	"The <e1>child</e1> was carefully wrapped and bound into the <e2>cradle</e2> by means of a cord."
Other
Comment:

3	"The <e1>author</e1> of a keygen uses a <e2>disassembler</e2> to look at the raw assembly code."
Instrument-Agency(e2,e1)
Comment:

4	"A misty <e1>ridge</e1> uprises from the <e2>surge</e2>."
Other
Comment:

5	"The <e1>student</e1> <e2>association</e2> is the voice of the undergraduate student population of the State
University of New York at Buffalo."
```

Before pre-processing

```
1 Component-Whole(e2,e1) the system as described above has its greatest application in an arrayed e1_start configuration
e1_end of antenna e2_start elements e2_end .
2 Other the e1_start child e1_end was carefully wrapped and bound into the e2_start cradle e2_end by means of a cord .
3 Instrument-Agency(e2,e1) the e1_start author e1_end of a keygen uses a e2_start disassembler e2_end to look at the raw
assembly code .
6 Other this is the sprawling e1_start complex e1_end that is peru 's largest e2_start producer e2_end of silver .
7 Cause-Effect(e2,e1) the current view is that the chronic e1_start inflammation e1_end in the distal part of the stomach
caused by helicobacter pylori e2_start infection e2_end results in an increased acid production from the non-infected
upper corpus region of the stomach .
8 Entity-Destination(e1,e2) e1_start people e1_end have been moving back into e2_start downtown e2_end .
9 Content-Container(e1,e2) the e1_start lawsonite e1_end was contained in a e2_start platinum crucible e2_end and the
counter-weight was a plastic crucible with metal pieces .
```

After pre-processing

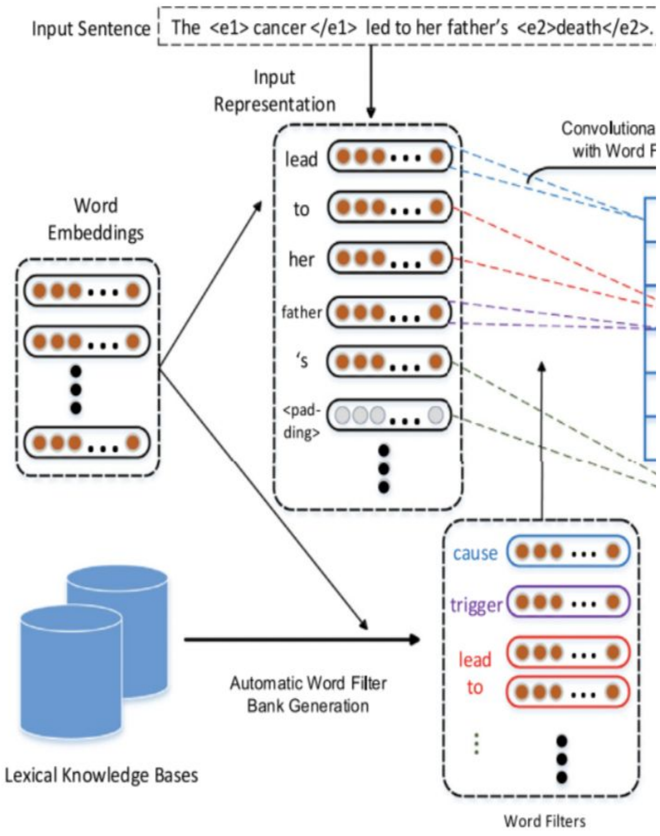
Flow Chart:

2. Pre-trained Embeddings:

trained on several datasets

1. Google's Word2Vec

   Word2vec creates vectors that are numerical representations of word features.

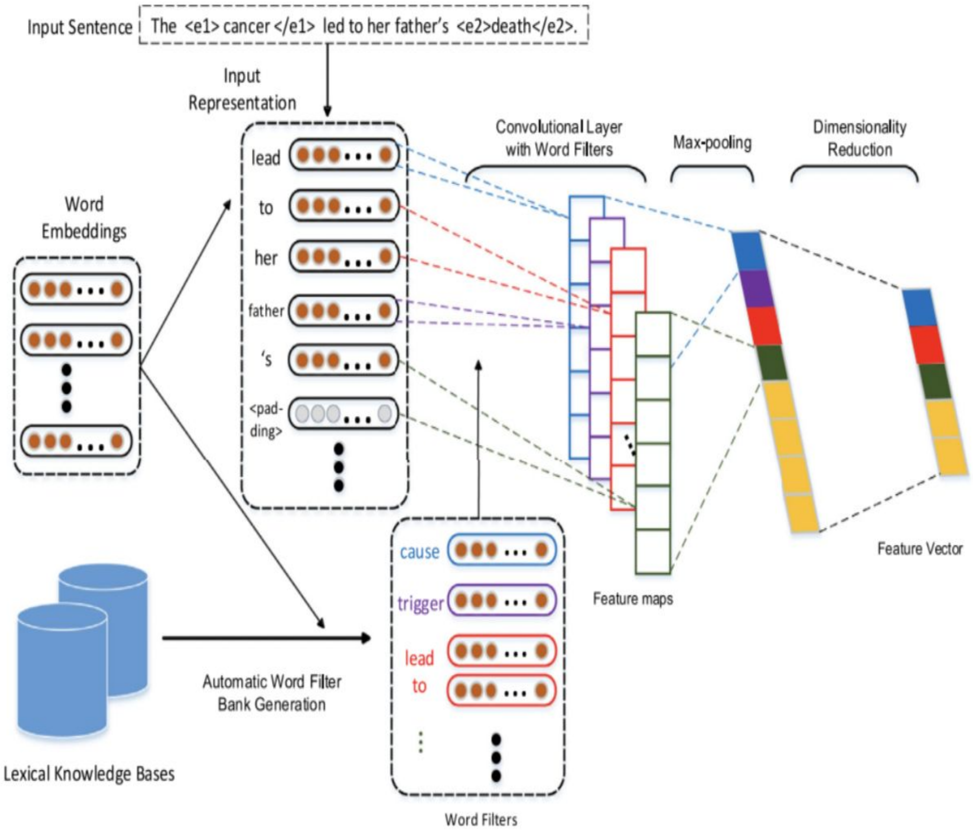


Flow Diagram (Pengfei Li Kezhi Mao, 2019)

Flow Chart:

3. Neural Network Modelling

- **Data-oriented channel**



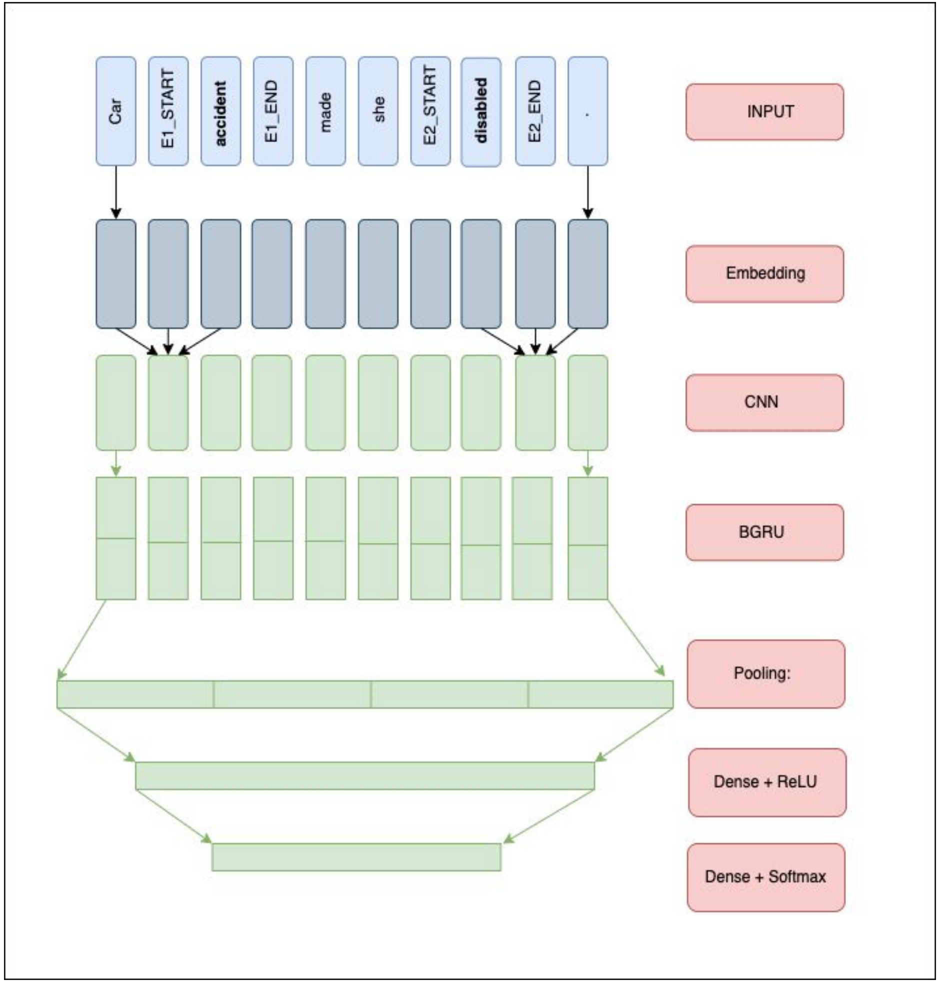
Flow Diagram (Pengfei Li Kezhi Mao, 2019)

Data-Oriented Channel

- Processed data sentence
- Pre-Trained Embedding
- Neural Network:

  CNN + Bidirectional Gate Recurrent Unit

- Multiple Pooling
- ReLu ⟶ Softmax



Layers of Neural Network

```
Model: "model_2"
__________________________________________________________________________________________________
 Layer (type)                   Output Shape         Param #     Connected to
==================================================================================================
 words_input (InputLayer)       [(None, 101)]        0           []

 words_Embedding (Embedding)    (None, 101, 300)     7035000     ['words_input[0][0]']

 dropout_6 (Dropout)            (None, 101, 300)     0           ['words_Embedding[0][0]']

 conv1d_2 (Conv1D)              (None, 101, 256)     230656      ['dropout_6[0][0]']

 dropout_7 (Dropout)            (None, 101, 256)     0           ['conv1d_2[0][0]']

 bidirectional_2 (Bidirectional)  (None, 101, 128)   123648      ['dropout_7[0][0]']

 activation_4 (Activation)      (None, 101, 128)     0           ['bidirectional_2[0][0]']

 dense_6 (Dense)                (None, 101, 1)       129         ['activation_4[0][0]']

 permute_2 (Permute)            (None, 1, 101)       0           ['dense_6[0][0]']

 attn_softmax (Activation)      (None, 1, 101)       0           ['permute_2[0][0]']

 words_input_mask (InputLayer)  [(None, 101)]        0           []

 lambda_2 (Lambda)              (None, 1, 128)       0           ['attn_softmax[0][0]',
                                                                  'activation_4[0][0]']

 global_max_pooling1d_2 (GlobalMaxPool  (None, 128)  0           ['activation_4[0][0]']
 ing1D)

 mask_max_pooling_layer_2 (MaskMaxPool  (None, 128)  0           ['activation_4[0][0]',
 ingLayer)                                                        'words_input_mask[0][0]']

 flatten_2 (Flatten)            (None, 128)          0           ['lambda_2[0][0]']

 concatenate_2 (Concatenate)    (None, 384)          0           ['global_max_pooling1d_2[0][0]',
                                                                  'mask_max_pooling_layer_2[0][0]',
                                                                  'flatten_2[0][0]']

 dropout_8 (Dropout)            (None, 384)          0           ['concatenate_2[0][0]']

 dense_7 (Dense)                (None, 300)          115500      ['dropout_8[0][0]']

 dense_8 (Dense)                (None, 19)           5719        ['dense_7[0][0]']

 activation_5 (Activation)      (None, 19)           0           ['dense_8[0][0]']
```

Knowledge-Oriented Channel

1. **WordNet:**
- It is a huge lexical database to represent various **meanings of words**
- Arranges English words into set of synonyms (synsets)

2. **FrameNet:**
- It is a huge lexical database organises English **words and Phrases** into high level semantic frames that describes various concepts.

Results

- **Relationship are most confused with "others"**
- **Large number of wrong directions for M-T**
- **Confused group: CW-MC**

<<< (9+1)-WAY EVALUATION TAKING DIRECTIONALITY INTO ACCOUNT -- OFFICIAL >>>:

Confusion matrix:

	C-E	C-W	C-C	E-D	E-O	I-A	M-C	M-T	P-P	_O_	-SUM-	xDIRx	skip	ACTUAL
C-E	301	0	0	0	10	0	0	0	4	12	327	1	0	328
C-W	1	242	4	2	1	9	13	5	2	29	308	4	0	312
C-C	0	0	174	3	1	0	1	0	0	9	188	4	0	192
E-D	0	2	12	262	0	1	0	1	0	13	291	1	0	292
E-O	1	0	3	2	233	2	0	3	2	11	257	1	0	258
I-A	0	5	0	1	0	120	0	0	5	23	154	2	0	156
M-C	0	4	0	1	1	0	205	0	1	19	231	2	0	233
M-T	0	1	0	0	2	0	1	234	0	18	256	5	0	261
P-P	4	4	0	2	6	11	2	4	178	19	230	1	0	231
_O_	14	29	28	17	38	20	32	29	21	226	454	0	0	454
-SUM-	321	287	221	290	292	163	254	276	213	379	2696	21	0	2717

(<-- classified as)

Coverage = 2717/2717 = 100.00%

Accuracy (calculated for the above confusion matrix) = 2175/2717 = 80.05%

CE: Cause-Effect ,**CW:**Component-Whole, **CC:**Content-Container, **ED:**Entity-Destination, **EO:**Entity-Origin, **IA:**Instrument-Agency, **MC:**Member-Collection, **MT:**Message-Topic, **PP:**Product-Producer, **O:**Others

Results

```
Results for the individual relations:
             Cause-Effect :    P =  301/( 321 +   1) =  93.48%     R =  301/ 328 =  91.77%     F1 =  92.62%
          Component-Whole :    P =  242/( 287 +   4) =  83.16%     R =  242/ 312 =  77.56%     F1 =  80.27%
        Content-Container :    P =  174/( 221 +   4) =  77.33%     R =  174/ 192 =  90.62%     F1 =  83.45%
       Entity-Destination :    P =  262/( 290 +   1) =  90.03%     R =  262/ 292 =  89.73%     F1 =  89.88%
            Entity-Origin :    P =  233/( 292 +   1) =  79.52%     R =  233/ 258 =  90.31%     F1 =  84.57%
        Instrument-Agency :    P =  120/( 163 +   2) =  72.73%     R =  120/ 156 =  76.92%     F1 =  74.77%
        Member-Collection :    P =  205/( 254 +   2) =  80.08%     R =  205/ 233 =  87.98%     F1 =  83.84%
            Message-Topic :    P =  234/( 276 +   5) =  83.27%     R =  234/ 261 =  89.66%     F1 =  86.35%
         Product-Producer :    P =  178/( 213 +   1) =  83.18%     R =  178/ 231 =  77.06%     F1 =  80.00%
                   _Other :    P =  226/( 379 +   0) =  59.63%     R =  226/ 454 =  49.78%     F1 =  54.26%

Averaged result (excluding Other):
P = 1949/2338 =  83.36%     R = 1949/2263 =  86.12%     F1 =  84.72%


<<< The official score is (9+1)-way evaluation with directionality taken into account: micro-averaged F1 = 84.72% >>>
```

Comparison

As we can observe from the comparison table,

K-CNN perform better than our model, which was expected.

	K-CNN	CNN+BGRU
Precision	93.12	83.36
Recall	90.73	86.12
F1 Score	91.82	84.72

In our future work we plan to add hopefully match if not improve the K-CNN results.

K-CNN Results: Knowledge-oriented convolutional neural network for causal relation extraction from natural language texts

Extra: Hyperparameters

Units - 64

Batch_size - 128

Dropout_emb - 0.64

Dropout_model - 0.32

Dropout_pen - 0.32

Learning_rate - 1.0

Activation_fn - tanh

Nb_epoch - 256

Es_epoch_stop - 20

CATENA

CATENA

- How about a model that works on both temporal and causal event extraction?
- Individual components for temporal and causal event extraction.
    - Temporal module, a combination of rule-based and supervised classifiers, with a temporal reasoner module in between.
    - Causal module, a combination of a rule-based classifier according to causal verbs, and supervised classifier taken into account syntactic and context features, especially causal signals appearing in the text.
- Each component can improve the other!

System architecture

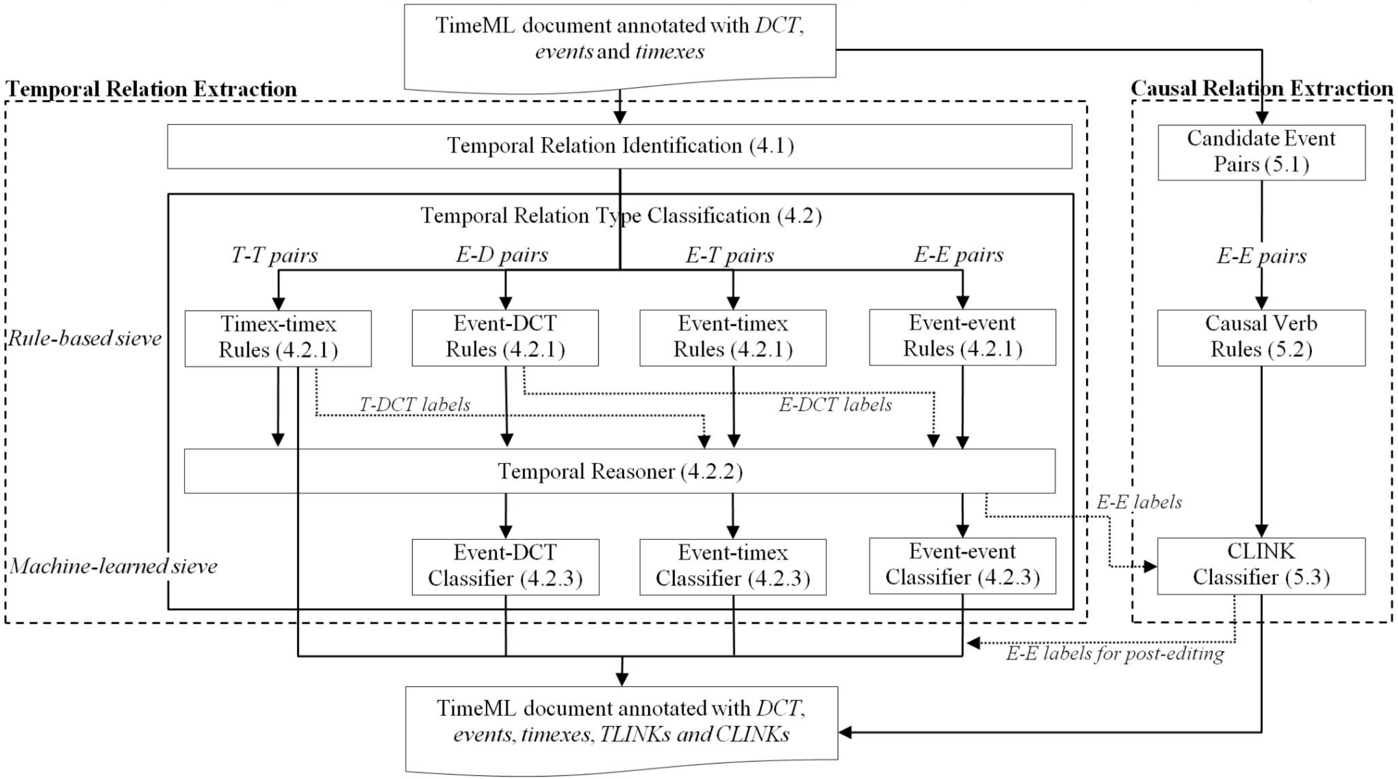


Image taken from the original research paper CATENA(Paramita Mirza)

TimeML Document

- TimeML- A markup language with annotated DCT(document creation time), events and Timexes.


```
<?xml version="1.0" ?><TimeML xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="http://timeml.org/timeMLdocs/TimeML_1.2.1.xsd">

<DOCID>bbc_20130322_721</DOCID>

<DCT><TIMEX3 functionInDocument="CREATION_TIME" temporalFunction="false" tid="t0" type="DATE" value="2013-03-22">2013-03-22</TIMEX3></DCT>

<TITLE>Russian police raid rights group Memorial and other NGOs</TITLE>

<TEXT>

Russian police and tax inspectors have <EVENT class="OCCURRENCE" eid="e1">raided</EVENT> the offices of the human rights group Memorial and other civil society groups which get foreign

Memorial is famous for <EVENT class="OCCURRENCE" eid="e2">documenting</EVENT> human rights <EVENT class="OCCURRENCE" eid="e3">abuses</EVENT> in Russia.

The US embassy in Moscow has <EVENT class="REPORTING" eid="e4">voiced</EVENT> concern and <EVENT class="I_ACTION" eid="e5">asked</EVENT> the Russian government for an <EVENT class="OCCU

A new Russian law <EVENT class="STATE" eid="e6">says</EVENT> foreign-funded non-governmental groups (NGOs) linked to politics must register as "foreign agents" - a term which suggests s

In the worst <EVENT class="OCCURRENCE" eid="e1000016">repressions</EVENT> of the Soviet period the label "foreign agents" was used to <EVENT class="OCCURRENCE" eid="e10">denounce</EVENT

Memorial <EVENT class="REPORTING" eid="e12">says</EVENT> inspectors <EVENT class="OCCURRENCE" eid="e13">returned</EVENT> to its Moscow offices on <TIMEX3 tid="t1" type="DATE" value="201

A statement on the Memorial website <EVENT class="REPORTING" eid="e15">said</EVENT> the <EVENT class="OCCURRENCE" eid="e1000018">inspections</EVENT> were directly <EVENT class="OCCURREN

Memorial director Arseny Roginsky, <EVENT class="OCCURRENCE" eid="e17">quoted</EVENT> by the Russian news website Vesti, <EVENT class="REPORTING" eid="e18">said</EVENT> it <EVENT class=

He <EVENT class="I_ACTION" eid="e21">insisted</EVENT> that the NGO law "will not <EVENT class="OCCURRENCE" eid="e22">change</EVENT> our position at all". "We won't <EVENT class="OCCURRE

</TEXT>
```

Temporal Relation Identification

- Pairing up temporal entities based on rules of TempEval-3 task description
- Rules are
  - two main events of consecutive sentences
  - two events in the same sentence
  - an event and a timex in the same sentence
  - an event and a document creation time
  - pairs of all possible timexes (including document creation time)
- Now we have e-e, e-d, e-t and t-t pairs
- Example: He threw a [party]$^{(e1)}$ on his [Birthday]$^{(e2)}$ -> (party,birthday)

Temporal Relation Classification

- Temporal Rule-Based Sieve: We use handcrafted rules designed for each type of pair.
  - Timex-Timex rules: “7 PM tonight” (2015-12-12T19:00) IS INCLUDED in “today”
  - Event-DCT rules: “(had) fallen” is BEFORE
  - Event-Timex rules:
    - “Since birth, he like candies”-BEGUN_BY
    - “Police [confirmed] E [Friday] T that the body was found…”-IS_INCLUDED
    - “Between January and February”-BEGUN_BY and END_BY
  - Event-Event rules: “He got $100 by winning the lottery”.

Temporal Relation Classification- contd.

- Temporal Reasoner: Output from previous sieve is mapped into a constraint problem and processed by automated temporal reasoner.
- Mapped according to Allen relations:
  - < and > for BEFORE and AFTER
  - o and o^{-1} for DURING and DURING_INV
  - d and d^{-1} for IS INCLUDED and INCLUDES
  - s and s^{-1} for BEGINS and BEGUN BY
  - f and f^{-1} for ENDS and ENDED BY
- We use Generic Qualitative Reasoner (GQR) as:
  - Fast and efficient
  - Works great with Allen problems

Temporal Relation Classification- contd.

- We build three supervised classification models:
  - event-DCT (E-D)
  - event-timex (E-T)
  - event-event (E-E) pairs.

  We use LIBLINEAR (Fan et al., 2008) L2-loss linear SVM (default parameters), and one-vs-rest strategy for multi-class classification.

Causal Event Extraction

- Like Temporal Event Identification, we began by making candidate event pairs.
- Unlike Temporal Event Extraction, we take into account every possible combination of events in a sentence in a forward manner as candidate event pairs
- We even consider the next line
- Example: “The bank robber [robbing]^{e1} the bank was [arrested]^{e2} and [charged]^{e3} for 2 years”
- Pairs: (robbing,arrested) , (robbing,charged) ,(arrested, charged)

Causal Relation Classification

- Rule based sieve:
  - Manually cluster verbs with same syntactic behaviour and define set of rules for each verb. For instance, causing, resulting, belong to the same cluster and results in a CLINK.
- Supervised Classifier:
  - Directions play a huge role, for instance, “The building [collapsed] T because of the [earthquake] S” vs “Because of the [earthquake] S the building [collapsed] T”. S and T denote the source (cause) and target (effect) of the causal relation.
  - We build a classification model using LIBLINEAR (Fan et al., 2008) L2-loss linear SVM (default parameters), and one-vs-rest strategy for multi-class classification. The classifier has to label an event pair (e1, e2) with CLINK, CLINK-R or O for others.

Testing on temporal

```
the value of the Indonesian stock market has <EVENT eid="e7" class="OCCURRENCE">fallen</EVENT>
by twelve percent. The Indonesian currency has <EVENT eid="e9" class="OCCURRENCE">lost</EVENT>
twenty six percent of its value.
```

Screenshot taken from the input file ABC19980108.1830.711.tml

```
ABC19980108.1830.0711.col    e3   e4   SIMULTANEOUS
ABC19980108.1830.0711.col    e66  e368      BEFORE
ABC19980108.1830.0711.col    e7   e9   AFTER
ABC19980108.1830.0711.col    e65  e66  BEFORE
ABC19980108.1830.0711.col    e65  e67  IS_INCLUDED
```

Screenshot taken from the console output

Testing on Causal

```
the <EVENT class="OCCURRENCE" eid="e1000018">inspections</EVENT>
were directly <EVENT class="OCCURRENCE" eid="e16">linked</EVENT>
to the new law on NGOs and the targeted groups'
<EVENT class="STATE" eid="e1000019">compliance</EVENT> with it.
```

Screenshot taken from the input file ABC19980108.1830.711.tml

```
bbc_20130322_721.col    e1000019    e1000018    CLINK
bbc_20130322_721.col    e10 e1000017    NONE
wsj_1014.col    tmx2201 tmx0    AFTER
wsj_1014.col    tmx261  tmx0    INCLUDES
wsj_1014.col    tmx354  tmx0    BEFORE
```

Screenshot taken from the console output

Results

Relation	Precision	Recall	F1
Before	.366	.275	.314
After	.6	.26	.37
Simultaneous	0	0	0
Includes	.66	.1	.17
is_included	.66	.14	.24
Vague	.57	.85	.685
Average	.47	.27	.29
Weighted-Avg	.53	.545	.494
W-Avg Original	.512	.510	.511
Total Recall	0.545		

Result after running EvaluateTimeBankDenseCrossVal

Conclusion

- We studied two different models, CATENA and K-CNN based data-oriented model.
- We studied the architecture of both systems.
- We were able to fully replicate CATENA and analyse the results with respect to the actual model.
- For K-CNN, we were able to build different data-oriented model, CNN+BGRU (than paper), with small margin in results compared to CNN with Knowledge-oriented channel.