EURECOM
Sophia Antipolis

Semester Project Report

Spring 2022

Event Causality Detection using NLP Techniques

AGARWALLA, Yash

27 June 2022

1. Introduction

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. The relationships between two events can be further classified into temporal and causal. The events are temporal when they are time-related. For instance, “We went to the school today”, “he slept at 9 pm tonight”. Here, “9 pm tonight” is included in “today” and hence going to school and sleeping are temporally related. Whereas, when one relation is caused by the other, the type of the relationship is called a causal relationship, for example, “His brakes failed and he had an accident”. Here, failure of the brakes caused the accident hence the events are causally related.

Several techniques have been proposed so far for the detection of relationships that are either exclusively based on machine learning, linguistic markers, or rule-based. In this project, we have studied a sieve-based model CATENA [2], which deals with both machine learning and a rule-based sieve. CATENA has a multiclass architecture for the extraction and classification of both temporal and causal events.

CATENA takes any annotated document with DCT, events, and timexs as an input and outputs the same document with temporal and causal links between the events in CATENA. DCT stands for Document Creation Time for the respective document, an event can be interpreted as any occurrence such as “eating”, “establishing”, etc and timexs represent any temporal information, for example, “2AM”, “friday”, “2022”,etc. The model outputs the same document with temporal and causal links between each event for CATENA.

Our end goal is to perform relation extraction along with the classification of the relation by assigning labels to entity pairs using different approaches. We are mainly working on causal and temporal relations. In addition, we want to study the performance of the model and fine-tune it.

This study makes us capable of implementing temporal as well as causal event extraction. Firstly, we understand the working architecture of how each component works in depth. Then, We replicate the model, and finally, we try to compare the results with the original work. The link to the code is available at https://github.com/yashagr911/catena-finall

2. CATENA

The approach CATENA is inspired by "temporal relation extraction” [3], which introduced 437 hand-coded rules coupled with supervised learning, and “CAEVO” [4], which combined rule-based and data-driven classifiers in a sieve-based architecture for temporal ordering. However, they both focus entirely on temporal links.

There are some works done on the causal side as well but they only focus on specific types of event pairs and causal expressions in the text.

CATENA is the first integrated system available to perform temporal and causal relation extraction at the same time! and none of the other present systems can deal with this task. To achieve this, CATENA has two separate classification modules for temporal and causal relations respectively.

It takes a TimeML[1] document which is a formal specification language for events and temporal expressions as input with DCT, events, and timexs as input(as discussed earlier). While the expected output is the temporal and causal links on the same document. TLINKS or temporal links denote the temporal ordering such as BEFORE, INCLUDES, etc while CLINKS or causal links denote the cause-effect relationship between the events.

The general idea behind the architecture of each component (temporal relation extraction and causal relation extraction) is to use a sieve-based approach where we start by labeling the data using a rule-based component or a transitive reasoner and feed the remaining unlabeled pairs into a supervised classifier like support vector machines. Moreover, the model is designed in such a way that each module can be used to improve the other as temporal and causal relations go hand in hand in real-life scenarios. For instance, TLINK labels for event-event pairs generated after running the rule-based sieve and the temporal reasoner modules can be fed into the CLINK classifier as a feature, and similarly, the CLINK labels can be used to correct wrongly identified pairs by the TLINK classifier.

Let us begin with understanding each step of both components. As in figure 1, For temporal relation extraction, we have two main components. 1) Temporal Relation Identification, which is based on a set of rules and 2) Temporal Relation Type Classification which is based on both rule-based and supervised learning combined. We also see that Temporal Relation Classification has three sub-components. These components are ordered by the precision of each of them. This allows us to label a few links using high precision based on rules, then use the reasoner to infer new links, and finally use the supervised learning to increase the recall based on previous steps.

[1] Further Details:
https://www.researchgate.net/publication/241318399_TimeML_A_Specification_Language_for_Temporal_and_Event_Expressions

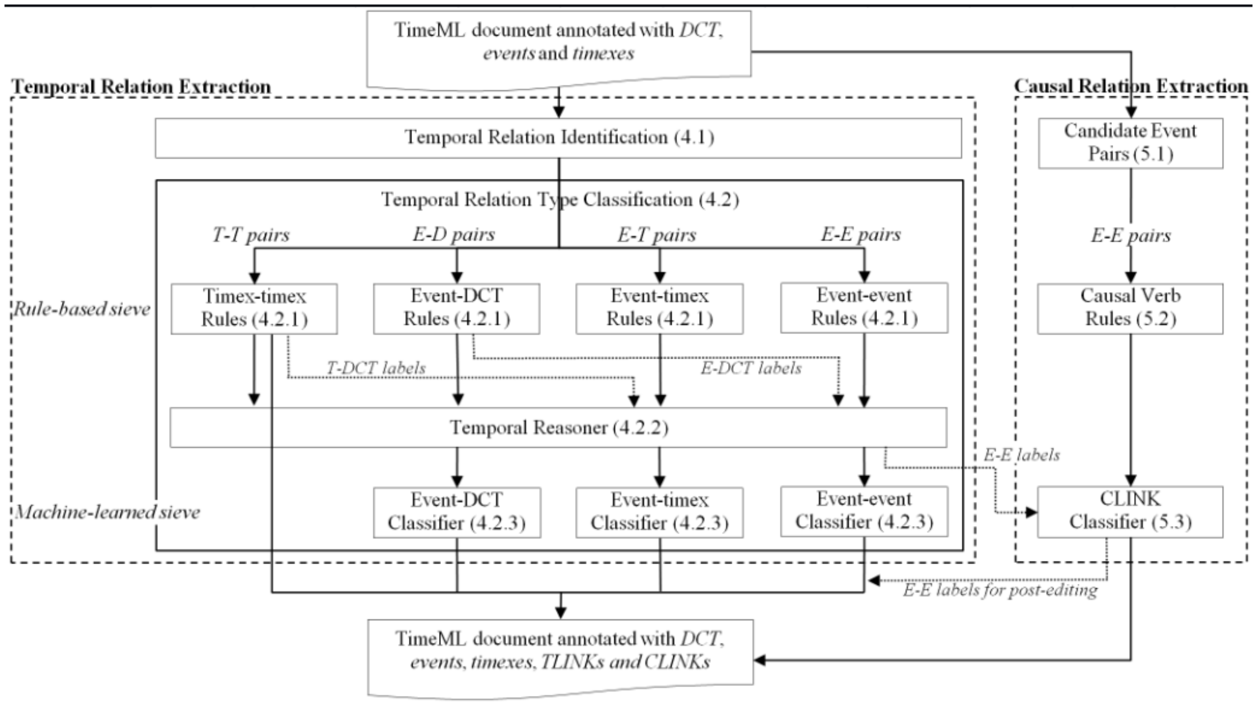

Figure 1: System architecture of CATENA

1. Let us try to understand what **Temporal Relation Identification** does. Basically, we try to map the temporal links between pairs of the entities based TempEval-3[2] task description. Their rules are as follows:

(i) two main events of consecutive sentences,
(ii) two events in the same sentence,
(iii) an event and a timex in the same sentence
(iv) an event and a document creation time and
(v) pairs of all possible timexes (including document creation time) linked with each other.

After this step, we have entity pairs grouped together into four different groups: timex-timex (T-T), event-DCT (E-D), event-timex (E-T), and event-event (E-E).

2. So now we move on to **Temporal Relation Classification** which is actually based on CAEVO(chambers et al., 2014) as we mentioned before. Since we have both rule-based and machine learning, it makes more sense to group them into individual sieves, that is, merging all the rule-based classifiers into a rule-based sieve and similarly, merging all support vector machine classifiers into a machine-learning sieve. We also run our temporal reasoner module only once instead of running it after each classifier for efficiency. Most importantly, the results from the rule-based sieve are used as features in the machine-learning sieve. For instance, for event-timex SVM, timex-DCT by timex-timex rules are used as a feature.

[2] Further details: https://paperswithcode.com/dataset/tempeval-3

2.1. Let us look in-depth at how the Temporal Rule-based sieve actually works. Previously we worked with rules defined by TempVal-3, instead, here we use hand-crafted rules designed for each entity pair since we already have the entity pairs from the last step.

2.1.1 timex-timex rules: we normalize the temporal expressions such as DATE and TIME. For instance, “10 PM today” is normalized to (2022-6-24T22:00) IS_INCLUDED in “today” which is normalized to (2022-6-24).

2.1.2 Event-DCT Rules: We label these pairs by taking into account the tense and/or aspect of the event word. For example, for the event mention “(had) eaten”, which has a perfective aspect and is in the past tense. So, according to the rule, it should be labeled “BEFORE”.

2.1.3 Event-timex Rules: we take into account prepositional temporal senses(Litkowski and Hargraves, 2006; Litkowski, 2014)[5]. A preposition of time is a preposition that allows us to discuss a specific time period such as a date on the calendar, one of the days of the week, or the actual time something takes place; for instance,”(in) January”, “(Since) birth”, here we see that “since” is a temporal preposition expressing a StartTime sense, thus, the relation is labeled as BEGUN_BY. Here, the timex “Since” is a *temporal modifier* as it acts as additional information for some temporal expressions such as Early or Late in the expression?

In absence of a temporal preposition, the timex can simply be a temporal modifier for the event as in the example, “I [read][Tuesday] about history”. Here, we can see that there is no preposition of time connecting “read” and “Tuesday”. So we consider it under IS_INCLUDED.

Moreover, sometimes events are modified by temporal expressions marking the starting time and ending time in a duration pattern such as ‘between TBEGIN and TEND’. We define additional rules as follows: (i) If the timex matches TBEGIN then the label is BEGUN BY, and (ii) if T matches TEND then the label is ENDED BY.

2.1.4 Event-Event Rules: there are two sets of rules for event-event labeling. The first set is using the verb information encoded between two events, that is, between the first(e1) and the second event(e2) and the verb information is encoded in the first event e1. “Brakes [failure] lead to an [accident]” Here, e1 and e2 are connected by an AFTER relation.

The second set of rules is inherited from CAEVO, including: (i) rules for linking a reporting event and another event syntactically dominated by the first, based on tense and aspect; and (ii) rules based on the role played by various tenses of English verbs in conveying temporal discourse (Reichenbach, 1947).

3. Temporal Reasoner: rule-based sieve is typical of high precision and low recall. Therefore, to alleviate the same, we run a transitive reasoner layer based on the output of the previous layer. This layer maps an annotated TimeML document into a constraint problem according to how TLINKS are mapped into Allen relations. (Allen, 1983)[6]. We apply the following mapping:

- < 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,

After mapping, we run an automated temporal reasoner to compute their deductive disclosure, and globally reasoning on them. We use the Generic Qualitative Reasoner[3] (GQR) (Westphal et al., 2010), a generic qualitative constraint problem solver which is fast and reliable, it works great on Allen constraint problems. The reason for choosing GQR over other solutions, such as fast Boolean Satisfiability Problem (SAT)[4] solvers, is due to its scalability, simplicity of use, and efficient performance (Westphal and Wolfl, 2009).

4. Temporal Supervised Classifiers: everything done till now was rule-based. Now, we use machine learning to improve our results. We start by building three supervised classification models, one for event-DCT (E-D), one for event-timex (E-T), and one for 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.

Several external **Tools and Resources** are used to extract features from each temporal entity pair, including

- **MorphoPro[5]** (Pianta et al., 2008), to get PoS tags and phrase chunks for each token.
- **Mate tools[6]** (Bjorkelund et al., 2010), to extract the dependency path between words.
- **WordNet similarity module[7]** to compute semantic similarity (Lin, 1998) between words.
- **Temporal signal lists**(Mirza and Tonelli 2014b), further expanded using the Paraphrase Database(Ganitkevitch et al., 2013), and manually clustered e.g. {before, prior to, in advance of}.

We implemented a set of features, listed in Table 1, on the basis of the best-performing systems in TempEval-2 (Verhagen et al., 2010)[7] and TempEval-3 (UzZaman et al., 2013)[8] campaigns. We used the simplified possible values of some features as follows:

---

[3] Further details: http://gki.informatik.uni-freiburg.de/events/aaai09-bench/tools/gqr-the-generic-qualitative-reasoner.html
[4] https://www.geeksforgeeks.org/2-satisfiability-2-sat-problem/
[5] https://textpro.fbk.eu/modules/morphopro
[6] https://www.ims.uni-stuttgart.de/en/research/resources/tools/matetools/
[7] https://metacpan.org/pod/WordNet::Similarity

Feature	TLINK E-D	TLINK E-T	TLINK E-E	CLINK E-E	Rep.	Description
Morphosyntactic information						
PoS	x	x	x	x	one-hot	Part-of-speech tags of e_1 and e_2.
phraseChunk	x	x	x	x	one-hot	Shallow phrase chunk of e_1 and e_2.
samePoS		x	x	x	binary	Whether e_1 and e_2 have the same PoS.
Textual context						
entityOrder		x			binary	Appearance order of e_1 and e_2 in the text.[5]
sentenceDistance		x	x	x	binary	0 if e_1 and e_2 are in the same sentence, 1 otherwise.
entityDistance		x	x	x	binary	0 if e_1 and e_2 are adjacent, 1 otherwise.
EVENT attributes						
class	x	x	x	x	one-hot	EVENT attributes as specified in TimeML.
tense	x	x	x	x	one-hot	
aspect	x	x	x	x	one-hot	
polarity	x	x	x	x	one-hot	
sameClass			x	x	binary	Whether e_1 and e_2 have the same EVENT attributes.
sameTenseAspect			x	x	binary	
samePolarity			x	x	binary	
TIMEX3 attributes						
type	x	x			one-hot	TIMEX3 attributes as specified in TimeML.
Dependency information						
dependencyPath			x	x	one-hot	Dependency path between e_1 and e_2.
isMainVerb	x	x	x	x	binary	Whether e_1/e_2 is the main verb of the sentence.
Temporal signals						
tempSignalTokens		x	x	x	one-hot	Tokens (cluster) of temporal signal around e_1 and e_2.
tempSignalPosition		x	x	x	one-hot	Temporal signal position w.r.t e_1/e_2 (BETWEEN, BEFORE, BEGIN, etc.)
tempSignalDependency		x	x	x	one-hot	Temporal signal dependency path between signal tokens and e_1/e_2.
Causal signals						
causSignalTokens				x	one-hot	Tokens (cluster) of causal signal around e_1 and e_2.
causSignalPosition				x	one-hot	Causal signal position w.r.t e_1/e_2 (BETWEEN, BEFORE, BEGIN, etc.)
causSignalDependency			x	x	one-hot	Causal signal dependency path between signal tokens and e_1/e_2.
Lexical semantic information						
wnSim			x	x	one-hot	WordNet similarity computed between the lemmas of e_1 and e_2.
TLINK labels from the rule-based sieve						
timex-DCT label		x			one-hot	The TLINK type of the e_2 (timex) and DCT pair (if any).
event-DCT label			x		one-hot	The TLINK types of the e_1/e_2 and DCT pairs (if any).

Table 1: Feature sets for TLINK classification of event-DCT (E-D), event-timex (E-T) and event-event (E-E) pairs, and for CLINK classifier (E-E pairs), with corresponding feature representation (Rep).

In the table, the dependencyPath corresponds to the dependency paths between an event pair only if it describes coordination, subordination, subject or object relation. For signalTokens, we include the clusterID of the clusters with the synonymous signals, for instance, {before, prior to, in advance of} instead of the features set tokens. Finally, for wnSim, the value of the WordNet similarity measure is discretized as: sim ≤ 0.0, 0.0 < sim ≤ 0.5, 0.5 < sim ≤ 1.0 and sim > 1.0.

We simplify the labels by combining some of the types, i.e., IBEFORE into BEFORE, IAFTER into AFTER, DURING and DURING INV into SIMULTANEOUS, considering only 10 out of 14 relation types due to the sparse annotation of such labels in the datasets.

Furthermore, we also exclude lexical features such as token/lemma from the feature set of temporal entities to increase the robustness of the classifier and decrease variance, therefore, we include WordNet similarity to capture the semantic connection between event words.

Causal Relation Extraction System:

Similar to temporal relation extraction, we still use a hybrid approach combining both rule-based and supervised classifiers. However, dealing with causal relationships is a bit more challenging because they can be expressed by different syntactic and semantic features and involve both situations specific information and world knowledge, whereas, temporal order has a clear formalization in the NLP community.

So, we use the concept of causality proposed in the annotation guidelines of the CausalTimeBank (Mirza et al., 2014; Mirza and Tonelli, 2014a)[9], which includes CAUSE, ENABLE, and PREVENT phenomena (Wolff, 2007; Wolff and Song, 2003)[10]. Particularly, when (i) causal verbs(effect, link, and causative verbs) include CAUSE-, ENABLE- and PREVENT-type verbs, For example, “I ran *because* I saw a ghost”. (ii)causal signals, which can be easily understood by the same example but with a different direction, “*because* I saw the ghost I ran”. We require different approaches to deal with each type because the first one is very simple and can be easily solved by rule-based approaches. However, the latter can be very ambiguous as it can appear in different syntactic constructions. Therefore, we tackle it using supervision, using the freely available Causal-TimeBank.

1. Causal Relation Identification: causal relation classification begins with the identification of candidate event pairs like the temporal processing model. Since our document is already annotated for event-event relations, we tend to build pairs for every possible combination of the events. For instance, if “e1, triggered by e2, cause them to e3,” the candidate event pairs are (e1,e2), (e1,e3) and (e2,e3). We also consider the following sentence as causality may be expressed in two consecutive sentences.

2. Causal Rule-Based Sieve: in rule-based, we mainly focus on causal verbs. In a typical causal verb v, the first event e1 is usually the subject of v and the second event e2 is either the object or the predicative complement of v. We take the list of 56 effect, link, and causative verbs presented in Mirza et al. (2014)[9] as the causal verb list, then augment the list with Paraphrase Database (Ganitkevitch et al., 2013)[11] and original verbs as seeds, resulting in a total of 97 verbs. We then manually cluster the causal verbs sharing the same syntactic behavior in groups and define a set of rules for each verb group, taking into account the possible existing dependency paths between v and e1/e2

3. Causal Supervised Classifier: A supervised approach is taken when it comes to determining the causal direction of the CLINKS that are signaled by causal signal. The classification model comprises LIBLINEAR (Fan et al., 2008)[12] L2-loss linear SVM (default parameters), and a one-vs-rest[8] strategy for multi-class classification. The classifier has to label an event pair (e1, e2) with CLINK, CLINK-R, or O for others. Note: we only care about those candidate pairs in which the causal signal is connected via dependency path to either e1, e2, or both.

3. Replication of the system

Luckily the code was already available for CATENA[9] but with some issues to address.

[8] More details: https://machinelearningmastery.com/one-vs-rest-and-one-vs-one-for-multi-class-classification/
[9] Github Link: https://github.com/paramitamirza/CATENA

3.1 Dataset

I started by trying to obtain the dataset from the links mentioned in the research paper by the original writer. However, the links were broken. So, I got the dataset from her GitHub page. I began the analysis of the data and discovered that the data is in two formats, COL and TML.

COL data is when the data is represented as a table, i.e. the first column has the entire dataset and the other columns have associated features.

```
DCT_1998-01-08  O   O   O   O   O   O   tmx0   DATE   1998-01-08   O   O   NP0   B-NP   dct_1998-01-08   LS   O   O

On       t1   1   on       O   O   O+O+O   O   O   O   O   O   PRP   B-PP   on   IN   t4:PMOD O
the      t2   1   the      O   O   O+O+O   O   O   O   O   O   AT0   B-NP   the   DT   O   O
other    t3   1   other    O   O   O+O+O   O   O   O   O   O   AJ0   I-NP   other   JJ   O   O
hand     t4   1   hand     O   O   O+O+O   O   O   O   O   O   NN1   I-NP   hand   NN   t2:NMOD||t3:NMOD   O
,        t5   1   ,        O   O   O+O+O   O   O   O   O   O   PUN   O   ,   ,   O   O
it       t6   1   it       O   O   O+O+O   O   O   O   O   O   PNP   B-NP   it   PRP   O   O
's       t7   1   be       O   O   O+O+O   O   O   O   O   O   VBZ   B-VP   's   VBZ   t1:ADV||t5:P||t6:SBJ||t8:VC||t19:P   mainVb
turning  t8   1   turn     e1   OCCURRENCE   PRESENT+PROGRESSIVE+POS O   O   O   O   O   VVG   I-VP   turn   VBG   t9:PRT||t10:OPRD
out      t9   1   out      O   O   O   O   O   O   O   O   AVP   B-PRT   out   RP   O   O
to       t10  1   to       O   O   O   O   O   O   O   O   TO0   B-VP   to   TO   t11:IM  O
be       t11  1   be       O   O   O   O   O   O   O   O   VBI   I-VP   be   VB   t16:PRD O
another  t12  1   another  O   O   O   O   O   O   O   O   DT0   B-NP   another DT   O   O
very     t13  1   very     O   O   O   O   O   O   O   O   AV0   I-NP   very   RB   O   O
bad      t14  1   bad      O   O   O   O   O   O   O   O   AJ0   I-NP   bad   JJ   t13:AMOD   O
financial        t15   1   financial   O   O   O   O   O   O   O   O   AJ0   I-NP   financial   JJ   O   O
week     t16  1   week     O   O   O   tmx83   DURATION   P1W   O   O   NN1   I-NP   week   NN   t12:NMOD||t14:NMOD||t15:NMOD||t17:NMOD
for      t17  1   for      O   O   O   O   O   O   O   O   PRP   B-PP   for   IN   t18:PMOD   O
```

Data in COL format

While TML files have the data presented in an XML format, where all the annotations are separate tags, for instance, EVENT has its own tag marking any events.

```
<?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>wsj_1014</DOCID>

<DCT><TIMEX3 tid="t0" type="DATE" value="1989-10-26" temporalFunction="false" functionInDocument="CREATION_TIME">10/26/89</TIMEX3></DCT>

<EXTRAINFO>
WSJ891026-0042 = 891026 891026-0042.
Anheuser-Busch, Slugging It Out, Plans Beer Price Cuts @ ---- @ By James P. Miller @ Staff Reporter of The Wall Street Journal 10/26/89 WALL STREET JOURNAL (J) BUD BEVERAGES (
</EXTRAINFO>

<TEXT>
Anheuser-Busch Cos. <EVENT eid="e1" class="REPORTING">said</EVENT> it <EVENT eid="e2" class="I_STATE">plans</EVENT> to aggressively <EVENT eid="e4" class="OCCURRENCE">discount

Anheuser, the world's largest brewer and U.S. market leader, has historically been <EVENT eid="e203" class="I_STATE">reluctant</EVENT> to <EVENT eid="e12" class="OCCURRENCE">e

Over the <TIMEX3 tid="t246" type="DATE" value="1989-SU" temporalFunction="true" functionInDocument="NONE" anchorTimeID="t0">summer</TIMEX3>, Anheuser competitors <EVENT eid="e

The St. Louis company <EVENT eid="e34" class="REPORTING">said</EVENT> major rivals, Philip Morris Co.'s Miller Brewing unit and Adolph Coors Co. "have been <EVENT eid="e35" cl

Anheuser-Busch <EVENT eid="e47" class="REPORTING">announced</EVENT> its <EVENT eid="e1991" class="I_STATE">plan</EVENT> at the same time it <EVENT eid="e50" class="REPORTING">
```

Data in TML format

For the evaluation of the temporal relation extraction module based on TempEval-3, we use TBAQ-cleaned(cleaned and improved version of the TimeBank 1.2 and the AQUAINT[10] corpora) and TempEval-3-platinum[11], respectively. The TimeBank 1.2 corpus contains 183 documents coming from a variety of news reports, specifically from the ACE[12] program and PropBank[13],

[10] More details: https://catalog.ldc.upenn.edu/LDC2002T31
[11] More details: https://aclweb.org/aclwiki/TempEval-3_Platinum_TimeML_annotations_(Repository)
[12] More details: https://www.ldc.upenn.edu/collaborations/past-projects/ace
[13] More details: https://propbank.github.io/

while the AQUAINT corpus contains 73 news report documents and is often known as the Opinion corpus(as related to news). The TempEval-3-platinum corpus, containing 20 news articles, was annotated by the TempEval-3 organizers.

The TimeBank-Dense corpus (Chambers et al., 2014)[13] is designed to address the problem of sparsity in TimeML corpora available. The resulting corpus contains 12,715 temporal relationships in 36 documents taken from TimeBank 1.2. For a TimeBank-Dense test, we follow the test setup Chambers et al. (2014), when the TimeBank-Dense corpus was divided into a 22-document training set of 5, 5 a text development set, and a 9.9 text review set.

To test the module to extract the causal relationship, we use the Causal-TimeBank corpus(Mirza and Tonelli, 2014a)[9] for training. For a TimeBank-Dense test, the test set is a part of TimeBank, so we do not submit 9 test documents to Causal-TimeBank during training. With the TempEval-3 test, we personally annotated 20 TempEval-3-platinum texts with the leading links following the annotation Causal-TimeBank guidelines. The causal relationship is much smaller than a temporary one, and we found only 26 CLINKs.

3.2 Working with the code

We already got the code from the Github of the original author of the CATENA. Since the code was relatively old, was designed to work on her system and lastly, was done in JAVA (we were not proficient in this language), we faced a lot of challenges. We used IntelliJ to run our code.

During the first run of the code, we faced an issue on one of the vectors used, it was an exception "IndexOutOfBoundException". After reading a bit about the exception and locating the error(using many print statements), We were able to locate the error and with the help of the supervisors, we understood how to fix the error, we fixed it by traversing the variable, found the original function that was changing the variable, discovered that the original function was not called, then found that there was a boolean which is actually responsible to enable that function and is set using arguments. Similarly, after addressing small bugs, like, invalid syntaxes, and removing some assert statements causing errors, we met another problem.

The problem was the absence of some resource files, in particular, "jawjaw.conf[14]" and "similarity.conf" was missing(according to the console output). These files are configurations to use the wordnet in our project. Luckily, a quick google search was enough to get the required files. However, after getting the files, we realized the wordnet itself was missing. It was specifically asking for "wnjpn.db[15]". Which is the combination of English and Japanese wordnet. Again we did a google search and found a GitHub link with the file. However, the file did not work so we had to get it from another source (which was extremely difficult). Finally, we were able to get our hands on the file (from a shady website). And the problem was fixed.

[14] Links to jawjaw.conf and similarity.conf: https://github.com/Sciss/ws4j/tree/master/config
[15] Link to wnjpn: https://qiita.com/tchih11/items/ba6873b194fa50306823

The final hurdle with the code was missing TML files. For our code to work, we need two versions of the same file. However, only COL format files were available in the author's project repository. Since the format TML is made by her, it was impossible to find it on the internet. We found a project on her Github which converted from TML to COL. We tried running her code (hoping there will be an option to get TML from COL), however, the code itself was buggy. After fixing the bugs and running it, we discovered that it was not capable of converting from COL to TML. After a deep analysis of both the formats, I started to understand the similarity and started to write the conversion code myself. It was quite challenging as there were certain problems like the presence of quotes(") in the words which broke the code. Meanwhile, our supervisor discovered that the author had tml files in her previous commits and we were able to get the files and the code was running perfectly after.

After running the code, we discovered that it took col and tml files from the directory "Catena-train_COL". It has 276 pairs of col and tml files. For testing, the files in the directory "example_COL" were used. We used 3 documents(instead of 2 used by the author) to test.

Example: TLINKS

```
the       t43    3    the       O    O           O
value     t44    3    value     O    O           O
of        t45    3    of        O    O           O
the       t46    3    the       O    O           O
Indonesian       t47  3         indonesian       O
stock     t48    3    stock     O    O           O
market    t49    3    market    O    O           O
has       t50    3    have      O    O           O
fallen    t51    3    fall      e7   OCCURRENCE
by        t52    3    by        O    O           O
twelve    t53    3    twelve    O    O           O
percent   t54    3    percent   O    O           O
.         t55    3    .         O    O           O

The       t56    4    the       O    O           O
Indonesian       t57  4         indonesian       O
currency  t58    4    currency  O    O           O
has       t59    4    have      O    O           O
lost      t60    4    lose      e9   OCCURRENCE
twenty    t61    4    twenty    O    O           O
six       t62    4    six       O    O           O
percent   t63    4    percent   O    O           O
of        t64    4    of        O    O           O
its       t65    4    its       O    O           O
value     t66    4    value     O    O           O
.         t67    4    .         O    O           O
```

Input data in COL(filename: ABC19980108.1830.711.col)

```
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.
```

Input data in TML(filename: 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
```

Output after running the model

Here, we know that **after** the currency **falls**(e7) the currency **loses**(e9) its value. This was predicted perfectly by our model. We can confirm that the model works when it comes to TLINKs

Example CLINKS:

```
A         t137      7         a         O         O         O
statement t138      7         statement O         O         O
on        t139      7         on        O         O         O
the       t140      7         the       O         O         O
Memorial  t141      7         memorial  O         O         O
website   t142      7         website   O         O         O
said      t143      7         say       e15       REPORTING PAST+NONE+
the       t144      7         the       O         O         O
inspections         t145      7         inspection          e1000018
were      t146      7         be        O         O         O
directly  t147      7         directly  O         O         O
linked    t148      7         link      e16       OCCURRENCE
to        t149      7         to        O         O         O
the       t150      7         the       O         O         O
new       t151      7         new       O         O         O
law       t152      7         law       O         O         O
on        t153      7         on        O         O         O
NGOs      t154      7         ngo       O         O         O
and       t155      7         and       O         O         O
the       t156      7         the       O         O         O
targeted  t157      7         target    O         O         O
groups    t158      7         group     O         O         O
'         t159      7         '         O         O         O
compliance          t160      7         compliance          e1000019
with      t161      7         with      O         O         O
it        t162      7         it        O         O         O
.         t163      7         .         O         O         O
```

Input data in COL(filename: bbc_20130322_721.col)

```
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.
```

Input data in TML (filename: bbc_20130322_721.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
```

Output after running the model

Here, we can see that the **inspection**(e1000018) was **caused** by the **compliance**(e1000019) which is predicted correctly by our model. We can confirm that our model was able to establish TLINKS and CLINKS.

Evaluating the code was simple as that part of the code had no bugs. We use 10 folds to evaluate your model against TimeBankDense. We consider the result for the 10th fold.

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

Compared to the original CATENA model, we observe that our model performed slightly worse. This may be due to the fact that we included 3 documents instead of 2 as used by the author.

4. Conclusion

We studied CATENA, a combined system for the extraction and classification of temporal and causal relations in text. We were able to replicate the model and had almost the same performance. We also understood how the sieve-based architecture works for both temporal and causal event classification and understood the individual sieves, i.e., machine learning and rule-based. We also tried to understand the interaction between both modules. Since, CATENA works on the already established notion of TimeML which makes it super convenient to generate the output, i.e., putting temporal and causal information.

However, we did not consider implicit causality because it was not annotated in the Casual-TimeBank corpus. So, our future plan is to get it annotated (by Mturkers or use a small dataset annotated by ourselves) and try to implement it. Finally, we also plan to launch a web tool that uses CATENA to increase its accessibility.

Acknowledgments: The research is based on an already ongoing research CATENA and is done under the supervision of Raphael Troncy, Youssra Rebboud, and Pasquale lisena.

5. References

[1] Pengfe iLi, Kezhi Mao. “knowledge-oriented convolutional neural network for causal relation extraction from natural language texts” - *School of Electrical and Electronic Engineering, Nanyang Technological University, 50 Nanyang Avenue, 639798, Singapore*

[2] Paramita Mirza and Sara Tonelli. 2016. “CATENA: CAusal and TEmporal relation extraction from NAtural language texts”. In Proceedings of *COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pages 64–75, Osaka, Japan*. *The COLING 2016 Organizing Committee.*

[3] D’Souza, Jennifer, and Vincent Ng. "Sieve-based spatial relation extraction with expanding parse trees." *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. 2015.

[4] Nathanael Chambers, Taylor Cassidy, Bill McDowell, and Steven Bethard. 2014. Dense event ordering with a multi-pass architecture. Transactions of the Association for Computational Linguistics 2:273– 284.

[5] Litkowski, Kenneth & Hargraves, Orin. (2006). Coverage and Inheritance in The Preposition Project. 10.3115/1621431.1621437.

[6] James F. Allen. 1983. Maintaining knowledge about temporal intervals. Commun. ACM, 26(11):832–843, November.

[7] Verhagen, Marc, Roser Sauri, Tommaso Caselli, and James Pustejovsky. "SemEval-2010 Task 13: TempEval-2." In Proceedings of the 5th international workshop on semantic evaluation, pp. 57-62. 2010.

[8] UzZaman, Naushad, Hector Llorens, Leon Derczynski, James Allen, Marc Verhagen, and James Pustejovsky. "Semeval-2013 task 1: Tempeval-3: Evaluating time expressions, events, and temporal relations." In Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), pp. 1-9. 2013.

[9] Mirza, Paramita. (2014). Extracting Temporal and Causal Relations between Events. 10-17. 10.3115/v1/P14-3002.

[10] Wolff, Phillip & Song, Grace. (2003). Models of Causation and the Semantics of Causal Verbs. Cognitive psychology. 47. 276-332. 10.1016/S0010-0285(03)00036-7.

[11] Juri Ganitkevitch, Benjamin Van Durme, and Chris Callison-Burch. 2013. PPDB: The Paraphrase Database. In Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 758–764, Atlanta, Georgia. Association for Computational Linguistics.

[12] Fan, Rong-En, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. "LIBLINEAR: A library for large linear classification." the Journal of Machine Learning research 9 (2008): 1871-1874.

[13] Cassidy, Taylor, Bill McDowell, Nathanel Chambers, and Steven Bethard. An annotation framework for dense event ordering. Carnegie-Mellon Univ Pittsburgh PA, 2014.