Event Annotation Guidelines

for Geo-political, Macro-economic and Commodity Price Movement Events

Version 1.0

Contents

Chapter 1

Introduction

This annotation guide contains the details of the Commodity News Corpus, covering topics like data collection, annotation guidelines, and next phase of annotation dealing with complicated event structure. The resulting annotated dataset is made publicly available and this guide serves as a detailed documentation for the effective usage of this dataset.

1.1 Corpus

The commodity news is scrapped from notable news agency including Reuters, Investing.com, Wall Street Journal and etc. Specifically only commodity news relating to **crude oil** are targetted at this stage. Potentially other traded commodities such as *gold, silver, natural gas* might be included in subsequent phrases of this project.

From the collated news articles, about 100 pieces of news articles were chosen based on their headlines to ensure a balanced dataset with each identified event type represented. Effort has been put in to make sure the dataset is as balanced as possible to avoid any skewness of data.

1.2 Source of Reference

The event schemas are designed referencing (1) ACE (Automatic Content Extraction) English Annotation Guidelines for Events [1]and (2) TAC-KBP Event track shared task running from 2015-2017[2]. . Details of events and their corresponding list of arguments is found in Section 3.2.

1.3 Annotation Tool

The dataset is annotated using Brat rapid annotation tool (Stenetorp et al., 2012), a web-based tool for text annotation[3]. The annotation task is explained using the example shown in Figure **??**. It is a snippet taken from a piece of commodity news. This annotated version of the sentence is shown in Figure1.1.

1.4 Quality Assurance

At this stage of the project, only straightforward events are annotated. Complicated events as highlighted in Section 5 such as nested events are left out for now, but will be addressed in the subsequent phases of the project. The plan is to involve more annotators and to ensure quality annotation through high inter-annotators' agreement.

[1] `https://www.ldc.upenn.edu/sites/www.ldc.upenn.edu/files/english-events-guidelines-v5.4.3.pdf`
[2] `https://tac.nist.gov//2015/KBP/`
[3] Installation `https://brat.nlplab.org/installation.html`

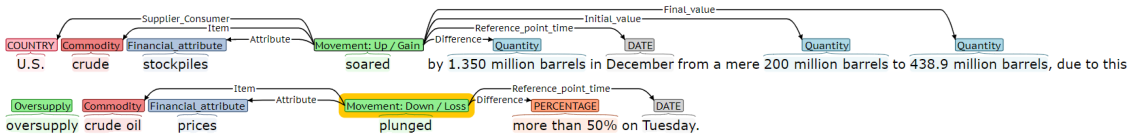

Figure 1.1: An annotated example using Brat annotation tool

Chapter 2

Entity Types

An entity mention is a reference to an object or a set of objects in the world, including named entities, nominal entities and pronouns. In the example in Figure **??**, entity mentions are *"U.S."*, *"crude"*, *"stockpiles"*. For convenience, values such as time/date (e.g., *"December"*, *"Tuesday"*) and value expressions (e.g., *"1.350 million barrels"*) are also considered as entity mentions in this work. There are 21 entity types identified and annotated in the dataset, the list of entity types are listed in Table 2.1 below:

Entity Type	**Examples**
1. Commodity	*oil, crude oil, Brent, West Texas Intermediate (WTI), fuel, U.S Shale, light sweet crude, natural gas*
2. Country**	*Libya, China, U.S, Venezuela, Greece*
3. Date**	*1998, Wednesday, Jan. 30, the final quarter of 1991, the end of this year*
4. Duration**	*two years, three-week, 5-1/2-year, multiyear, another six months*
5. Economic Item	*economy, economic growth, market, economic outlook, employment data, currency, commodity-oil*
6. Financial attribute	*supply, demand, output, production, price, import, export*
7. Forecast target	*forecast, target, estimate, projection, bets*
8. Group	*global producers, oil producers, hedge funds, non-OECD, Gulf oil producers*
9. Location**	*global, world, domestic, Middle East, Europe*
10. Money**	*$60, USD 50*
11. Nationality**	*Chinese, Russian, European, African*
12. Number**	(any numerical value that does not have a currency sign)
13. Organization**	*OPEC, Organization of Petroleum Exporting Countries, European Union, U.S. Energy Information Administration, EIA*
14. Other activities	(free text)
15. Percent**	*25%, 1.4 percent*
16. Person**	*Trump, Putin* (and other political figures)
17. Phenomenon	(free text)
18. Price unit	*$100-a-barrel, $40 per barrel, USD58 per barrel*
19. Production Unit	*170,000 bpd, 400,000 barrels per day, 29 million barrels per day*
20. Quantity	*1.3500 million barrels, 1.8 million gallons, 18 million tonnes*
21. State or province**	*Washington, Moscow, Cushing, North America*

Table 2.1: Entity Types

2.1 Named Entities

Entity type marked with ** in Table 2.1 are made up of mostly Named Entities. These named entities are identitical to the NER-tagging in Stanford CoreNLP. In the corpus, named entities that are part of an event are annotated. This includes both the Entity Type and as well as the Argument Role the entity plays in the event. Named entities which are not part of an event, on the other hand, are not annotated but they can be easily tagged by running Stanford CoreNLP NER tagging over them.

2.2 Nominal Entities

Nominal entities relating to Finance and Economics are annotated. This includes (1) entities which are linked to an event or (2) standalone entities which are not linked to an event. Nominal entities are more general as compared to Named Entities. Here's a list of nominal entities found in the Commodity News Corpus and was duly annotated:

1. oil-related : *oil, crude oil, crude, fuel, petroleum, gasoline, shale, oil fields, rig count, light sweet crude, reserves, oil and gas, petrochemicals, diesel, distillates,*
2. commodity-related items: *rig count*
3. attributes : *price, futures, contract, imports, exports, consumption, inventory, supply, production, usage, deliveries*
4. economic entity : *economic growth, economy, market(s), economic outlook, growth,*
5. others: *dollar, green back, employment data*
6. location: *domestic, global, world, worldwide*
7. keywords indicating forecast: *forecast, target, estimate, projection, bets*

Below is an example sentence with nominal entities annotated:
As [oil]$_{commodity}$ [production]$_{attribute}$ swells, [demand]$_{attribute}$ falters and [prices]$_{attribute}$ slide, the [global]$_{location}$ [oil]$_{commodity}$ [market]$_{economic\_item}$ appears on the verge of a pivotal shift from an era of scarcity to one of abundance.

Chapter 3

Event Types & Event Arguments

3.1 Event Triggers

According to TAC-KBP Event Guidelines and also highlighted in (?, ?), an event trigger (known as event nugget in the shared task) can be either a single word (main verb, noun, adjective, adverb) or a continuous or discontinuous multi-word phrase as shown in the list below:

1. **Verb**:
   - Houthi rebels **attacked** Saudi Arabia..
   - US **sanctioned** Iran.
2. **Noune**:
   - The government slapped **sanctions** against its petroleum....
   - ....supply and demand **fluctuations** in the international oil market.
3. **Adjectives**:
   - Interest rates were **unchanged**.....
   - A fast **growing** economy has....
4. **Adverb**:
   - The banks **increasingly** expect oil price to stay low.
5. **Multi-verb**:
   - The market **bounced back** ....
   - The company **laid** their workers **off**....

3.2 Event Type

Types of Event The list of annotated events are aligned to Ravenpack's event taxanomy categeories which is made publicly available. There are 19 event types identified and annotated, the event types are listed in Table 3.1 below:

Event Type	Example Trigger Word(s)
1. Cause-movement-down-loss	*cut, trim, reduce, disrupt, curb, squeeze, choked off*
2. Cause-movement-up-gain	*boost, revive, ramp up, prop up, raise*
3. Civil-unrest	*violence, turmoil, fighting, civil war, insurgent attacks, conflicts*
4. Crisis	*crisis, crises*
5. Embargo	*embargo, sanction*
6. Geopolitical-tension	*war, clashes, tensions, deteriorating relationship*
7. Grow-strong	*grow, picking up, boom, recover, expand, strong, rosy, improve, solid*
8. Movement-down-loss	*fell, down, less, drop, tumble, collapse, plunge, downturn, slump, slide, decline*
9. Movement-flat	*unchanged, flat, hold, no change, maintained*
10. Movement-up-gain	*up, gain, rise, surge, soar, swell, increase, rebound*
11. Negative-sentiment	*worries, concern*
12. Oversupply	*ample supply, glut, oversupply, bulging stock level, excess supplies*
13. Position-high	*high, highest, peak, highs*
14. Position-low	*low, lowest, lows, trough*
15. Prohibiting	*ban, bar, prohibit*
16. Shortage	*shortfall, shortage, under-supplied*
17. Situation-deteriorate	*deepen, worsen, intensified, exacerbated*
18. Slow-weak	*slow, weak, tight, lackluster, falter, weaken, bearish, slowdown, crumbles*
19. Trade-tensions	*price war, trade war, trade tensions, economic fallout, trade dispute*

Table 3.1: List of Event Types

3.2.1 Geo-political News

1. Civil unrest (**civil-unrest**)- Violence or turmoil within the oil producing country.
2. Other forms of Crisis (**Crisis**)- A time of intense difficulty.
3. Embargo (**Embargo** / **Prohibiting**) - Trade or other commercial activity of the commodity is banned.
4. Geo-political tension - Political tension between oil-producing nation with other nations.
5. Trade tensions (**Trade-tensions**)- Trade-related tension between oil-producing and oil-consuming nations.

3.2.2 Macro-economic News

1. Employment (**Grow-strong** / **Slow-weak**) - Status of US Employment Data, which is an indicator of economic situation.
2. Economy / GDP (**Grow-strong** / **Slow-weak**) - Economic / GDP growth of a nation.
3. Bearish technical view / outlook (**Negative-sentiment**) - Bearish sentiment or outlook

3.2.3 Commodity Supply (includes exports)

1. Oversupply (**Oversupply**)- Situation where production goes into surplus.
2. Shortage (**Shortage**)- Situation where demand is more than supply.
3. Supply increase (**Movement-up-gain**) - Situation where supply increased.
4. Supply increase (**Cause-movement-up-gain**) - Deliberate action to increase supply.
5. Supply decrease (**Movement-down-loss**) - Situation where supply decreased.
6. Supply decrease (**Cause-movement-down-loss**) - Deliberate action to decrease supply.

3.2.4 Commodity Demand (includes imports)

1. Demand increase (**Movement-up-gain**) - Situation where demand increased.
2. Demand decrease (**Movement-down-loss**) - Situation where demand decreased.

3.2.5 Commodity Price Movement

Commodity price here includes *spot price*, *futures* and *futures contract*.

1. Price increase (**Movement-up-gain**) - Situation where commodity price rises.
2. Price decrease (**Movement-down-loss**) - Situation where commodity price drops.
3. Price movement flat (**Movement-flat**) - Situation where no or little change to commodity price.
4. Price target /forecast increase (**Caused-movement-up-gain**) - Commodity forecasted / target price is raised.
5. Price target /forecast decrease (**Caused-movement-down-loss**) - Commodity forecasted / target price is lowered.
6. Price position (**Position-high**, **Position-low**) - Describes the position of the current commodity price.

3.3 Event Arguments

3.3.1 Embargo

Example sentence: The [Trump administration] imposed a "strong and swift" economic **sanctions** on [Venezuela] on [Thursday].

Role	Entity Type	Argument Text
Imposer	Organization, Country, Nationality, State or province, Person, Group, Location	Trump administration
Imposee[1]	Organization, Country, Nationality, State or province, Group	Venezuela
Reference_point_time	Date	Thursday

3.3.2 Prohibiting

Example sentence: [Congress] **banned** most [U.S.] [crude oil] [exports] on [Friday] after price shocks from the 1973 Arab oil embargo.

Role	Entity Type	Argument Text
Imposer	Organization, Country, Nationality, State or province, Person, Group, Location	Congress
Imposee	Organization, Country, Nationality, State or province, Group	U.S.
Item	Commodity, Economic_item	crude oil
Attribute	Financial_attribute	exports
Reference_point_time	Date	Friday
Activity	Other_activities	

3.3.3 Shortage

Example Sentence: Oil reserves are within "acceptable" range in most oil consuming countries and there is no **shortage** in [oil] [supply] [globally], the minister added.

Role	Entity Type	Argument Text
Place	Country, State or province, Location, Nationality	Congress
Item	Commodity	crude oil
Attribute	Financial_attribute	exports
Reference_point_time	Date	

3.3.4 Civil Unrest

Example sentence: The drop in oil prices to their lowest in two years has caught many observers off guard, coming against a backdrop of the worst **violence** in [Iraq] [this decade].

Sentence:		
Role	**Entity Type**	**Argument Text**
Place	Country, State or province, Location, Nationality	Iraq
Reference_point_time	Date	this decade

3.3.5 Crisis

Example Sentence: Asia 's diesel consumption is expected to recover this year at the second weakest level rate since the [2014] [Asian] [financial] **crisis**.

Role	Entity Type	Argument Text
Place	Country, State or province, Location, Nationality	Asian
Reference_point_time	Date	this year
Item	Commodity, Economic_item	financial

3.3.6 Geo-political Tension

Example sentence: **Deteriorating relations** between [Iraq] and [Russia] [first half of 2016] ignited new fears of supply restrictions in the market. **Trade Tension**

Role	Entity Type	Argument Text
Participating_countries	Country, Group, Organization, Location, State or province, Nationality	U.S., China
Reference_point_time	Date	early June

3.3.7 Oversupply

Example sentence: [Forecasts] for an [crude] **oversupply** in [West African] and [European] [markets] [early June] help to push the Brent benchmark down more than 20% January.

Role	Entity Type	Argument Text
Place	Country, Group, Organization, Location, State or province, Nationality	West African, European
Reference_point_time	Date	[this year]
Item	Commodity	crude
Attribute	Financial_attribute	markets
Difference	Production_unit	
Forecast	Forecast_target	forecasts

3.3.8 Caused-movement-down-loss, Caused-movement-up-gain

Example sentence: The [IMF] earlier said it **reduced** its [2018] [global] [economic growth] [forecast] to [3.30%] from a [July] forecast of [4.10%].

Role	Entity Type	Argument Text
Type	Nationality, Location	global
Place	Country, Group, Organization, Location, State or province, Nationality	West African, European
Supplier_consumer	Organization, Country, State_or_province, Group, Location	
Reference_point_time	Date	2018
Initial_reference_point	Date	July
Final_value	Percentage, Number, Money, Price_unit, Production_unit, Quantity	3.30%
Initial_value	Percentage, Number, Money, Price_unit, Production_unit, Quantity	4.10%
Item	Commodity, Economic_item	economic growth
Attribute	Financial_attribute	
Difference	Percentage, Number, Money, Production_unit, Quantity	
Forecast	Forecast_target	forecast
Duration	Duration	
Forecaster	Organization	IMF

3.3.9 Movement-down-loss, Movement-up-gain, Movement-flat

Example sentence: [Globally] [crude oil] [futures] **surged** [$2.50] to [$59 per barrel] on [Tuesday].

Role	Entity Type	Argument Text
Type	Nationality, Location	globally
Place	Country, Group, Organization, Location, State or province, Nationality	
Supplier_consumer	Organization, Country, State_or_province, Group, Location	
Reference_point_time	Date	Tuesday
Initial_reference_point	Date	
Final_value	Percentage, Number, Money, Price_unit, Production_unit, Quantity	$59 per barrel
Initial_value	Percentage, Number, Money, Price_unit, Production_unit, Quantity	
Item	Commodity, Economic_item	crude oil
Attribute	Financial_attribute	futures
Difference	Percentage, Number, Money, Production_unit, Quantity	$2.50
Forecast	Forecast_target	
Duration	Duration	
Forecaster	Organization	

3.3.10 Slow-weak, Grow-strong

Example sentence: [U.S.] [employment data] **contrasts** with the euro zone.

Role	Entity Type	Argument Text
Type	Nationality, Location	
Place	Country, Group, Organization, Location, State or province, Nationality	U.S.
Supplier_consumer	Organization, Country, State_or_province, Group, Location	
Reference_point_time	Date	
Initial_reference_point	Date	
Final_value	Percentage, Number, Money, Price_unit, Production_unit, Quantity	
Initial_value	Percentage, Number, Money, Price_unit, Production_unit, Quantity	
Item	Commodity, Economic_item	employment data
Attribute	Financial_attribute	
Difference	Percentage, Number, Money, Production_unit, Quantity	
Forecast	Forecast_target	
Duration	Duration	
Forecaster	Organization	

3.3.11 Position-high, Position-low

Example sentence: The IEA estimates that U.S. crude oil is expected to seek higher ground until reaching a [5-year] **peak** in [late April] of about [17 million bpd]. **Position-high / Position-low**

Role	Entity Type	Argument Text
Reference_point_time	Date	late April
Initial_reference_point	Date	
Final_value	Percentage, Number, Money, Price_unit, Production_unit, Quantity	17 million bpd
Initial_value	Percentage, Number, Money, Price_unit, Production_unit, Quantity	
Item	Commodity, Economic_item	
Attribute	Financial_attribute	
Difference	Percentage, Number, Money, Production_unit, Quantity	
Duration	Duration	5-year

3.3.12 Negative Sentiment

Example sentence: Oil futures have dropped due to **concern** about softening demand growth and awash in crude.

Role	Entity Type	Argument Text
Item	Commodity, Economic_item	
Attribute	Financial_attribute	

Note: **Negative Sentiment** is a special type of event, where majority of the time it contains just the trigger words such as *concerns, worries, fears* and 0 event arguments.

Chapter 4

Event Polarity and Modality

This section compares how event modality and polarity are defined and annotated across the benchmark datasets. Modality and Polarity cue words are in *italics*, the cue words found in commodity news dataset are annotated closely following these benchmark:

1. ACE 2005 :
In addition to defining event types and subtypes, ACE dataset also has additional properties defined and annotated for each event. These properties are **Polarity**, **Tense**, **Genericity** and **Modality**. For the purposes and scope of this work, only **Polarity** and **Modality** will be discussed. According to 2005 English Annotation Guidelines, the definitions are as follows:

- **Polarity:** An Event has the value *positive* unless there is an explicit indication that the event did not take place, in which case *Negative* is assigned. According to ACE2005, there are two categories:
  - **Syntactically** - using negative cue words or the standard negative syntax
    didn't say, *don't* think, *not* hurt, *not to be* extradited to California.
  - **Lexically** - using context
    refused to **sack** him, they backed out of the **purchase**, he *denied* **killing** the man.
- **Modality**: Determines whether the event represents a "real" occurrence. There are two possible values: *Asserted* if the author or speaker refers to it as though it were a real occurrence, and *other* otherwise. *Others* could include:
  - Believed Events
    rumors of **arrest**, *suspected* of **giving** money
  - Hypothetical Events
    if...., should he not **pay**
  - Commanded and Requested Events
    He was *asked* to **return**, Iraq was *ordered* to **cut**
  - Threatened, Proposed, and Discussed Events.
    US *threatened* to **sanction** Iran.
  - Desired Events
    China *wanted* to **increase** production.
  - Promised Events
    OPEC *agreed* to **cut** supplies.
  - Unclear or complicated sentence constructs.

2. Rich ERE : Realis
Rich introduced the Realis attribute in TAC-KBP 2015 Event Nugget Detection task. In the corpus, each event is tagged with one of each of Realis attributes:

- **ACTUAL**: events that actually occurred / attested events.
- **GENERIC**: events that are not specific events with a (known or unknown) time and/or place
- **OTHER**: all other events, including failed events, future events, and conditional statements, and all other non-generic variations.

The task in TAC-KBP 2015 task not only involves standard event nugget detection but also classification of the event's Realis attribute. However, TAC-KBP considers negated (failed) events, future events and conditional statements under the same Realis value, hence losing fine-grained epistemic status of the events. With just the Realis attribute, this event *US plans to lift sanctions on Iran* (future event) cannot be differentiated from *US lifted sanctions on Iran* (negated event) since both have the same Realis value.

3. Other materials on Modality and Polarity :

More rigorous definition of Modality and Polarity and in-depth analysis on Modality and Polarity can be found in the various sources below:

1. Modality and negation: An introduction to the special issue
2. ACL 2011 conference tutorial materials accessible in this link: `http://mirror.aclweb.org/ijcnlp11/downloads/tutorial/tu3_present.pdf`. According to this tutorial, types of expressions with modal meanings:
   - **Modal auxiliaries**
     Sandy must/should/might/may/could be home
   - **Semimodal verbs**
     Sandy has to/ought to/needs to be home
   - **Adverbs**
     Perhaps Sandy is home
   - **Nouns**
     There is a slight possibility that Sandy is home
   - **Adjectives** *Is it far from necessary that Sandy is home*
   - **Conditionals** *If the light is on, Sandy is home*

Chapter 5

Complicated Event Structure

Describe that on average sentences contain many events.

Even though the event schema used here follow ACE2005 and TAC-KBP 2015 guidelines closely, there are complicated events that are not specified in the guides and hence may not have the best annotation schema at this stage. These complicated events are considered out of scope for this work. As future enhancement, a new set of event schemas can be designed to cater for these special types of events. Below here are some examples:

1. Nested Events:
   - Delay a planned [easing of [output cuts]$_{event1}$]$_{event2}$.
   - OPEC is considering a [reduction in its [supply cuts]$_{event1}$ to 5.7 million bpd]$_{event2}$.
   - OPEC+ may hesitate to implement a planned [loosening of [output curbs]$_{event1}$]$_{event2}$.
2. Sentences without clear-cut events: Here is an example of a sentence written differently than a standard news article, the information is conveyed in an indirect way that makes it difficult to pinpoint clear-cut events
   - Then spent the rest of the week trying to defend those gains as market optimism over the vaccine gave way to concerns over the logistics of its eventual roll-out, though other factors contributed to the two-sided trade.
3. Reversal of events: The reversal of events are currently being annotated as **Negation** (NEGATIVE Polarity) but semantically they may not mean negation. Events with negative Polarity means the event did not take place, however in this situation, the said events did occur but was later reversed.
   - Oil prices climbed on Monday, recouping some losses from the previous session as hopes that OPEC+ will hold current output curbs *offset concerns* over weaker fuel demand due to growing coronavirus infections and higher production in Libya.
   - Oil *pared gains / paring losses*.

Hence more thought need to go into determining the best way to annotate events like these so that the events can be accurately represented.

In the next phase of the project, event schema and refined annotation guidelines will be developed to effectively capture the event details. More annotators will also be involved in annotating complex events described above. High annotator inter-annotator agreement will be used as criteria for determining the best annotation.

Chapter 6

Citation

This dataset is made available for anyone, please cite :

[Citation details will be made available later.]

References

Stenetorp, P., Pyysalo, S., Topić, G., Ohta, T., Ananiadou, S., & Tsujii, J. (2012). Brat: a web-based tool for nlp-assisted text annotation. In *Proceedings of the demonstrations at the 13th conference of the european chapter of the association for computational linguistics* (pp. 102–107).