MONASH University

# Effective Use of Graph Convolution Network and Contextual Sub-Tree for Commodity News Event Extraction

Meisin Lee, Lay-Ki Soon, Eu-Gene Siew

School of Information Technology, Monash University

# What is Event Extraction?

In line with ACE2005's definition of the task of Event Extraction:

1. Event **Trigger** extraction - identifying and classifying event triggers.
2. Event **Arguments** extraction - identifying arguments of event triggers and labeling their roles.

Information extracted is useful for:

1. Event sequences & narrative progression
2. Commodity Price Prediction


An example of a piece of commodity News (Figure 1 in paper)

| Event | Entity Mention | Argument Role |
|---|---|---|
| Trigger: **soared** | U.S. | Supplier |
| | crude | Item |
| | stockpiles | Attribute |
| | 1.350 million barrels | Difference |
| Event type: **movement_ up_gain** | December | Reference point |
| | 200 million barrels | Initial Value |
| | 438.9 million barrels | Final Value |
| | more than 50% | NONE |

Event extraction of the first event in example above (Table 1 in paper)

# What is different about Commodity News?

- Generally a **less researched area** as compared to
  - generic event extraction (ACE2005)
  - company-related financial events such as :
    - Merger & Acquisition
    - Quarterly results
    - Buy ratings / buy call
    - Dividend announcements, etc

- Events found in Commodity News are vastly different from the above, they can be grouped into:
  - **Geo-political,** eg: Trade war, civil unrest
  - **Macro-economic in nature,** eg: weak GDP growth
  - **Supply-demand-related,** eg: oversupply, increase in rig count
  - **Commodity Price Movements,** eg: price surged, price dropped.

# Details about the dataset...

- Dataset is introduced by Lee et al 2021 in "An annotated commodity news corpus for event extraction".
- **8,580 entities** and **3,949 events**, on average **3 events per sentence***.
- **21 Entity types*** (Nominal and Named Entities)
- **18 Event types***
- **19 Argument roles***

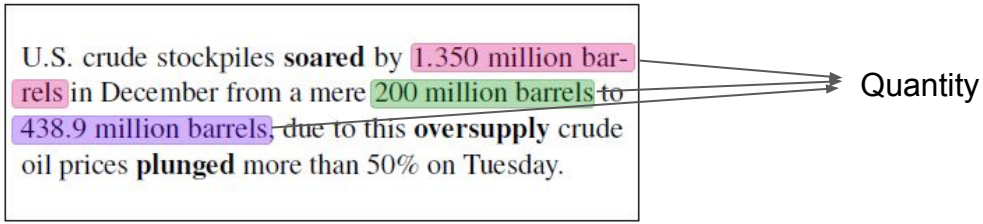
* See backup slides at the end for more information

| Event type | Type ratio | # sentence |
|---|---|---|
| 1. Cause-movement-down-loss | 13.35% | 524 |
| 2. Cause-movement-up-gain | 2.23% | 88 |
| 3. Civil-unrest | 2.53% | 100 |
| 4. Crisis | 0.76% | 30 |
| 5. Embargo | 3.75% | 148 |
| 6. Geopolitical-tension | 1.70% | 67 |
| 7. Grow-strong | 6.03% | 238 |
| 8. Movement-down-loss | 22.69% | 896 |
| 9. Movement-flat | 1.52% | 60 |
| 10. Movement-up-gain | 22.13% | 874 |
| 11. Negative-sentiment | 4.79% | 189 |
| 12. Oversupply | 2.63% | 104 |
| 13. Position-high | 3.82% | 151 |
| 14. Position-low | 3.11% | 123 |
| 15. Prohibiting | 1.06% | 42 |
| 16. Shortage | 1.04% | 41 |
| 17. Slow-weak | 5.47% | 216 |
| 18. Trade-tensions | 1.39% | 55 |
| Total | | 3949 |

Event type distribution and sentence level counts (Table 2 in paper)

# Unique characteristics

- Unique characteristics of events found in Commodity News:
  - **Number intensity**** - arguments are made up of numbers such as
    - Price - opening, closing price
    - Percentage - percentage of change
    - Dates
  - **Argument homogeneity** - arguments are made up of similar entity types
  - **Undifferentiated event types without its arguments** - hence event detection is insufficient.



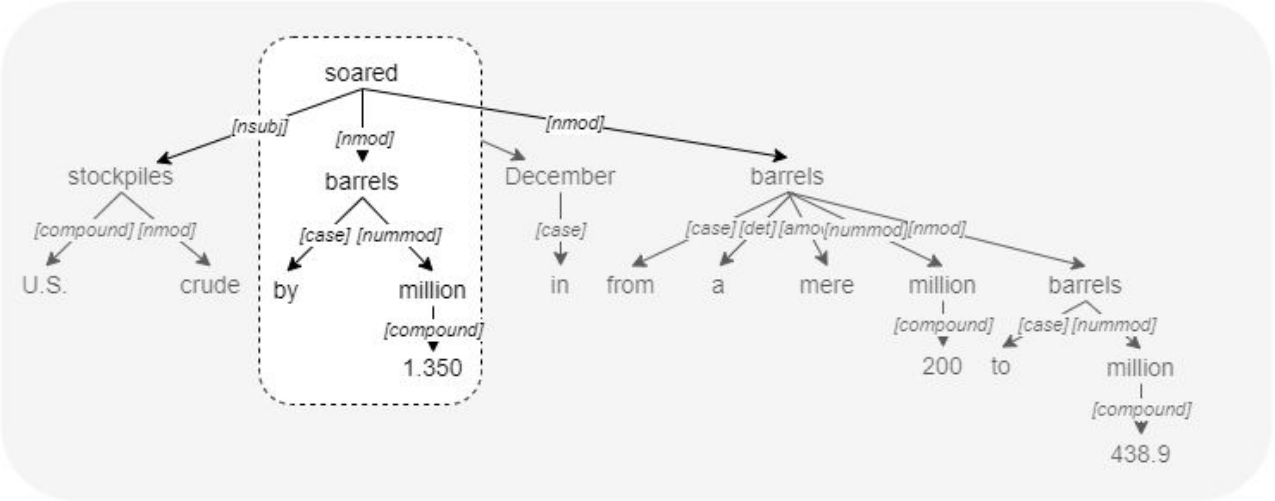

** values and temporal expressions are considered as entities in this work.

Question: How can we achieve accurate event extraction?

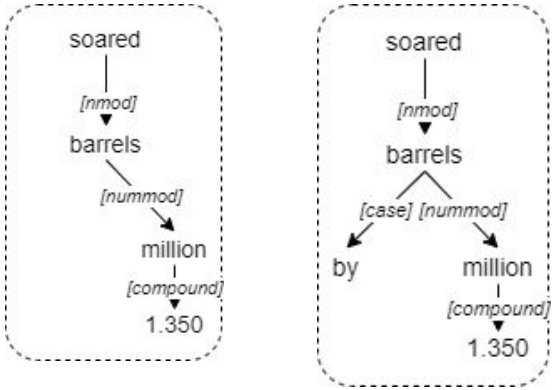
**Proposed solution: Graph Convolution Network with Contextual Sub-tree.**

# Contextual Sub-tree: a pruned dependency parse tree

Contextual Sub-tree is pruned leaving behind the **shortest path** between **candidate trigger** and **candidate argument** but also contains **off-path information** of distance DIST away from the shortest path.

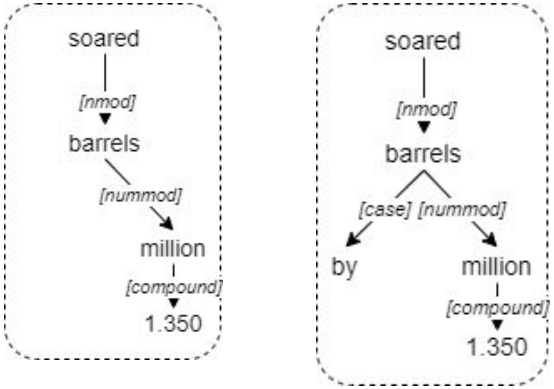



The dependency parse tree of the given example (Figure 3 in paper)




Left: Sub-tree with shortest path
Right: Sub-tree with off-path information DIST =1

# Contextual Sub-tree: a pruned dependency parse tree




Left: Sub-tree with shortest path
Right: Sub-tree with off-path information DIST =1
(Figure 4 in paper)

| Words in sub-dependency parse tree | Entity Type | Argument role |
|---|---|---|
| (1) [stockpiles] **soared** | Financial-Attribute | Attribute |
| (2) **soared** by [1.350 million barrels] | Quantity | Difference |
| (3) **soared** in [December] | Date | Reference time |
| (4) **soared** from a mere [200 million barrels] | Quantity | Initial value |
| (5) **soared** to [438.9 million barrels] | Quantity | Final value |

Table showing four pairings of trigger-argument and their words in context sub-tree of DIST = 1

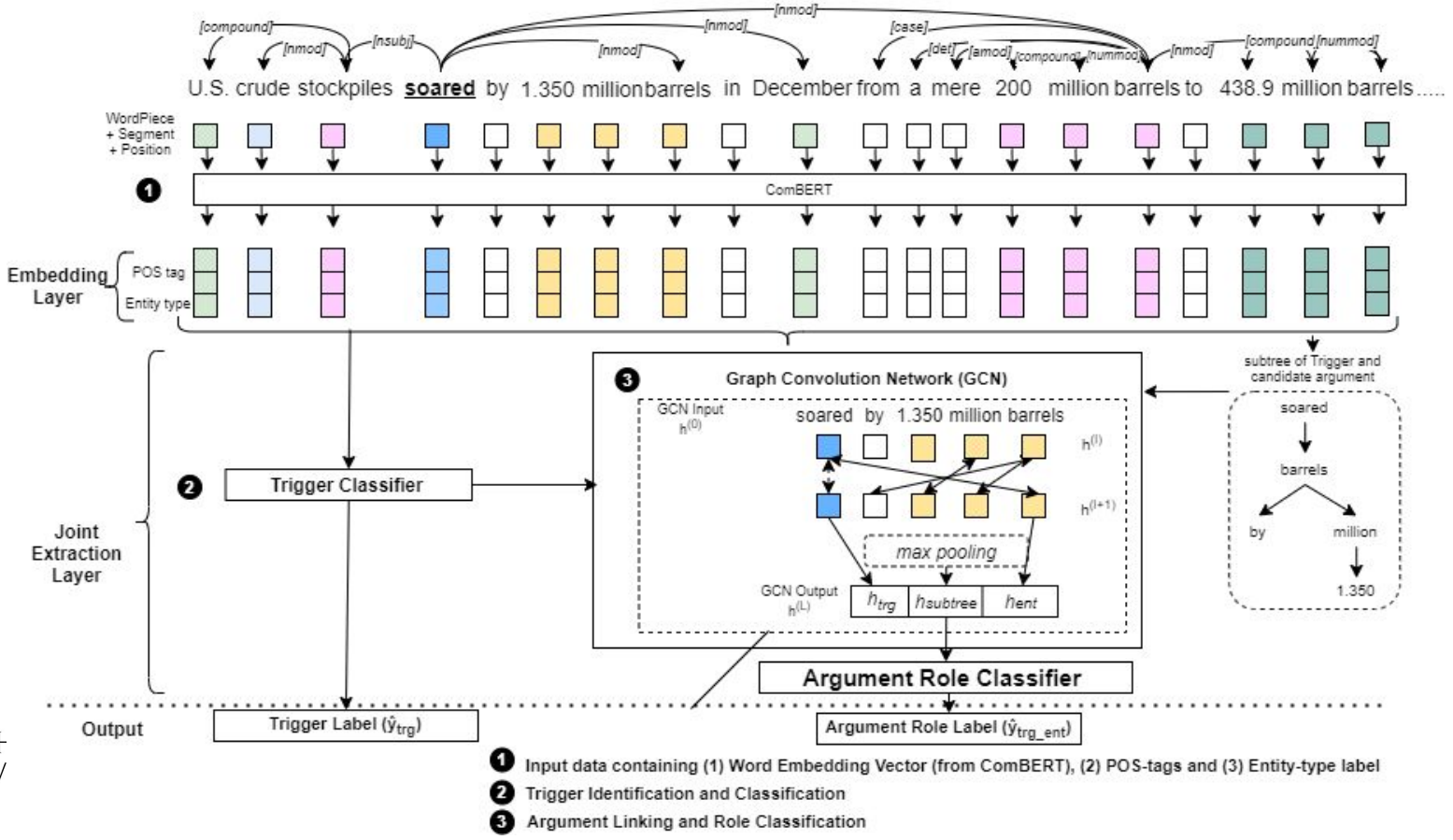
Why use DIST =1?

- Sub-tree with shortest path are rather shallow. DIST=1 is optimal.
- Also shown in (Zhang et al. 2018), DIST=1 produces the best results for relation extraction.



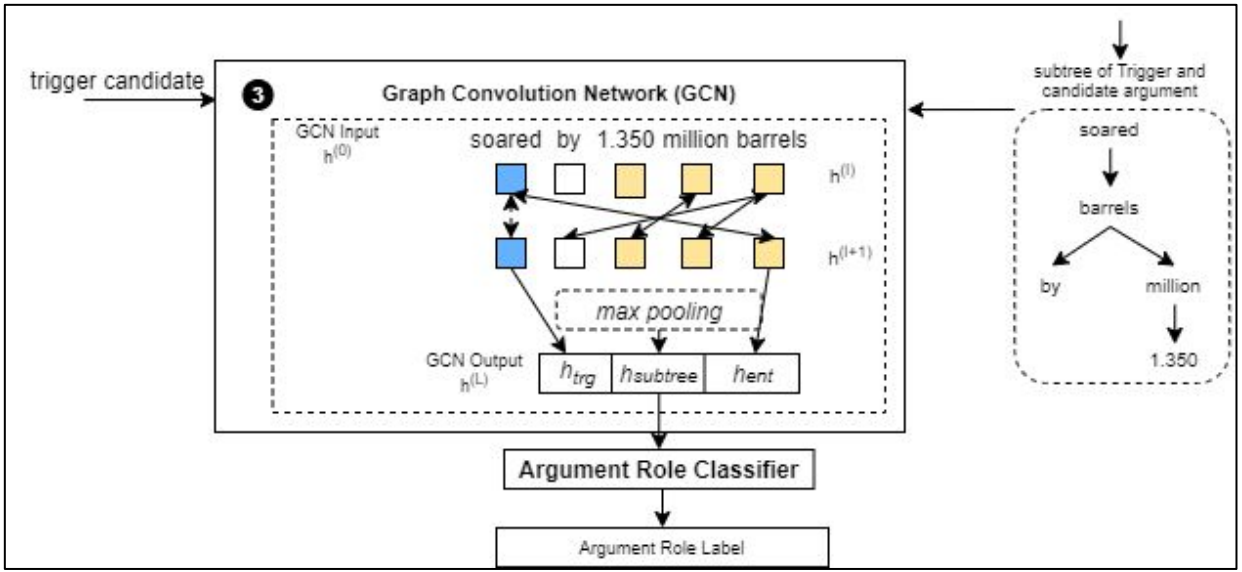
Reference:
Yuhao Zhang, Peng Qi, and Christopher D. Manning. 2018. Graph convolution over pruned dependency trees improves relation extraction

# Overall Solution Architecture




MONASH University

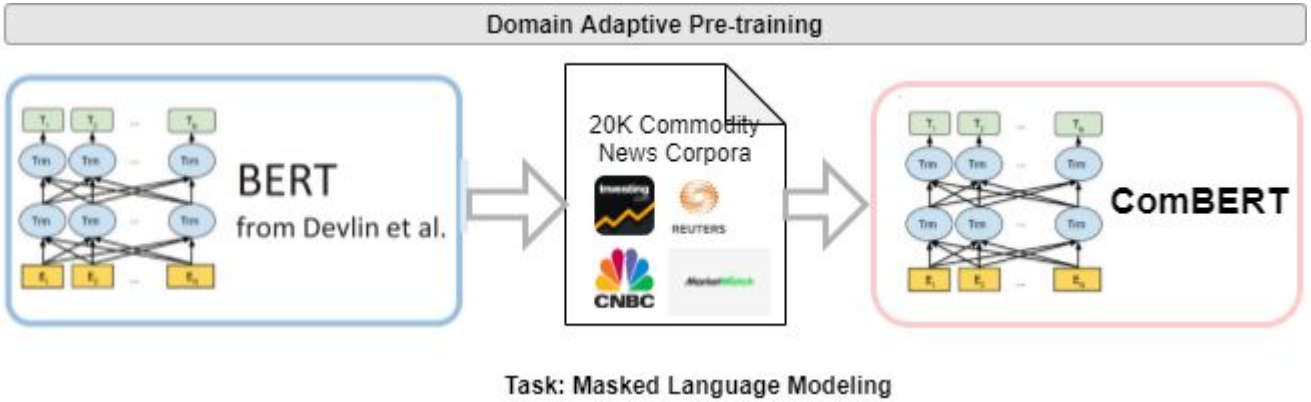
# Graph Convolution Network (GCN) over sub-dependency parse tree




A segment from the overall architecture on Graph Convolution operation on Sub-dependency parse tree

MONASH University

# ComBERT: Domain-Adaptive Pre-Training




Polysemous words in the Finance and Economics domain:

- `Stocks`: (1) inventory and (2) shares
- `Tank`: (1) Storage vessel (noun) and (2) market / price drop (verb)
- `Appreciates`: (1) recognize the full worth of and (2) increase in value

# Result and Analysis

**Analysis:**

- Contextual sub-tree produced better results both overall and also at argument role classification subtask.
- Higher F1 scores in classifying entities of the same type to the right argument role.

Group of Entities with same Entity type but plays different argument roles:

- ♣ Money / Price unit, production unit
- ♢ Date
- ♠ Country

| Argument Roles | Entity Type | Argument Role Classification F1 Score | | | | |
|---|---|---|---|---|---|---|
| | | Model A | Model B | Model C | Model D | Model E |
| NONE | - | 0.84 | 0.84 | 0.90 | 0.91 | **0.94** |
| Attribute | Financial Attribute | 0.40 | 0.65 | 0.79 | 0.75 | **0.83** |
| Item | Economic Item | 0.64 | 0.85 | **0.88** | 0.85 | **0.88** |
| Final_value ♣ | Money / Production unit / Price unit / Percentage / Quantity | 0.43 | 0.39 | 0.71 | 0.75 | **0.79** |
| Initial_value ♣ | Money / Production unit / Price unit / Percentage / Quantity | 0.56 | 0.56 | 0.73 | 0.69 | **0.77** |
| Difference ♣ | Money / Production unit / Price unit / Percentage / Quantity | 0.58 | 0.69 | 0.84 | **0.89** | **0.89** |
| Reference_point ♢ | Date | 0.54 | 0.69 | **0.80** | 0.71 | **0.80** |
| Initial_reference_point ♢ | Date | 0.40 | 0.63 | 0.63 | 0.60 | **0.66** |
| Contract_date ♢ | Date | 0.52 | 0.54 | 0.70 | 0.66 | **0.80** |
| Duration | Duration | 0.55 | 0.55 | 0.75 | 0.82 | **0.84** |
| Type | Location | 0.52 | 0.59 | 0.70 | 0.68 | **0.76** |
| Imposer ♠ | Country / State or province | 0.71 | 0.69 | **0.81** | 0.79 | **0.81** |
| Imposee ♠ | Country / State or province | 0.50 | 0.49 | 0.60 | **0.68** | **0.68** |
| Place ♠ | Country / State or province | 0.58 | 0.69 | **0.74** | 0.60 | **0.74** |
| Supplier_consumer ♠ | Country / State or provience / Nationality / Group | 0.49 | 0.71 | 0.73 | 0.73 | **0.79** |
| Impacted_countries ♠ | Country | 0.42 | 0.69 | 0.72 | 0.70 | **0.76** |
| Participating_countries ♠ | Country | 0.65 | 0.75 | 0.78 | 0.83 | **0.89** |
| Forecaster | Organization / Group | 0.62 | 0.75 | 0.78 | 0.80 | **0.82** |
| Forecast | Forecast_Target | 0.61 | 0.61 | 0.83 | 0.67 | **0.91** |
| Situation | Phenomenon / Other acitivites | 0.57 | 0.69 | **0.73** | 0.67 | 0.66 |

F1-scores Argument role classification (Table 4 in paper.)

# Any Questions?

# Commodity News Dataset - Example of events I

- **Embargo**: Trade or other commercial activity of the commodity is banned.
  *The Trump administration imposed a "strong and swift" economic sanctions trigger on Venezuela on Thursday.*
- **Shortage**: Situation where demand is more than supply.
  *Oil reserves are within "acceptable" range in most oil consuming countries and there is no shortage in oil supply globally, the minister added.*
- **Civil-unrest**: Violence or turmoil within the oil producing country.
  *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.*
- **Geo-political Tension**: Political tension between oil-producing nation with other nations.
  *Deteriorating relations between Iraq and Russia first half of 2016 ignited new fears of supply restrictions in the market.*

# Commodity News Dataset - Example of events II

- **Oversupply**: Situation where production goes into surplus.
  *Forecasts for an crude oversupply in West African and European markets early June help to push the Brent benchmark down more than 20% January.*
- **Movement-up-gain / Movement-down-loss/ Movement-flat**: Situation where commodity price rises, falls or no change.
  *Globally crude oil futures surged $2.50 to $59 per barrel on Tuesday.*
- **Slow-weak / Grow-strong**: Economic / GDP / Employment condition of a nation.
  *U.S. employment data strengthen with the euro zone.*
- **Position-high / Position-low:** Describes the position of the current commodity price.
  *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.*



For complete examples, please see "An annotated commodity news corpus for event extraction".

# Entity Types

| 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 with Entity types

MONASH University

# Event Schemas - Movement-up-gain

**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 | |

MONASH University

For complete examples, please see "An annotated commodity news corpus for event extraction".