Coarse-grained IE:Event Extraction

How To Predict The Stock Based On News Events

Zhang Haijun
MSc candidate

Department of Computer Science and Engineering
Shanghai Jiao Tong University

October 18, 2017

1. Introduction
2. Representation
3. Extraction
4. Embedding
5. Application(Stock Prediction)

1. Introduction
2. Representation
3. Extraction
4. Embedding
5. Application(Stock Prediction)

Example news [Ding et al., 2014]

Steve Jobs Death: Apple Stock (AAPL) Dips - ABC News
abcnews.go.com › Money
Oct 6, 2011 - Shares of **Apple** Inc. fell as trading began in New York on Thursday morning, the day after former CEO Steve Jobs passed away.

Google's stock falls after grim earnings come out early - Oct. 18, 2012
money.cnn.com/2012/10/18/technology/**google**-earnings/
Oct 18, 2012 - **Google's** third-quarter earnings results missed analysts' estimates on both sales and profit, in a report that was accidentally released early.

Figure: Example news for Apple Inc. and Google Inc.

Example news

The two news are discribed below:

1. Shares of Apple Inc. fell as trading began in New York on Thursday morning, the day after its former CEO Steve Jobs passed away.
2. Google's stock fell after grim earnings came out

Which inspires us that:

- If extracting events from unstructured text accurately, we can do better for downstream application.

Example news

The two news are discribed below:

1. Shares of Apple Inc. fell as trading began in New York on Thursday morning, the day after its former CEO Steve Jobs passed away.
2. Google's stock fell after grim earnings came out

Which inspires us that:

- If extracting events from unstructured text accurately, we can do better for downstream application.

Unstructured Features

Shallow Features:

1. Bags of words
2. Noun phrases
3. Named entities

eg : "Apple has sued Samsung Electronics for copying 'the look and feel' of its iPad tablet and iPhone smartphone."

Bags of words feature : "Apple", "sued", "Samsung", "Electronics","copying",. . .

Shortcoming:

1. Such unstructured representation can not indicate the actor(subject) and object of the event.
2. It can not focus on key events information embedded in free text, which introduces additional noise.

Unstructured Features

Shallow Features:

1. Bags of words
2. Noun phrases
3. Named entities

eg : "Apple has sued Samsung Electronics for copying 'the look and feel' of its iPad tablet and iPhone smartphone."

Bags of words feature : "Apple", "sued", "Samsung", "Electronics","copying",. . .

Shortcoming:

1. Such unstructured representation can not indicate the actor(subject) and object of the event.
2. It can not focus on key events information embedded in free text, which introduces additional noise.

Unstructured Features

Shallow Features:

1. Bags of words
2. Noun phrases
3. Named entities

eg : "Apple has sued Samsung Electronics for copying 'the look and feel' of its iPad tablet and iPhone smartphone."

Bags of words feature : "Apple", "sued", "Samsung", "Electronics","copying",...

Shortcoming:

1. Such unstructured representation can not indicate the actor(subject) and object of the event.
2. It can not focus on key events information embedded in free text, which introduces additional noise.

Structured Features

Representation	$E = (O_1, P, O_2, T)$
O_1	Actor(Subject) of event
P	Action(Predicate) of event
O_2	Object of event
T	Timestamp of event

eg : Sep 3, 2013, - Microsoft agrees to buy Nokia's mobile phone business for $7.2 billion.

Structured Representation:

- Actor =Microsoft
- Action = buy
- Object = Nokia's mobile phone business
- Time = Sep 3, 2013

Structured Features

Representation	$E = (O_1, P, O_2, T)$
O_1	Actor(Subject) of event
P	Action(Predicate) of event
O_2	Object of event
T	Timestamp of event

eg : Sep 3, 2013, - Microsoft agrees to buy Nokia's mobile phone business for $7.2 billion.

Structured Representation:

- Actor =Microsoft
- Action = buy
- Object = Nokia's mobile phone business
- Time = Sep 3, 2013

Event Extraction

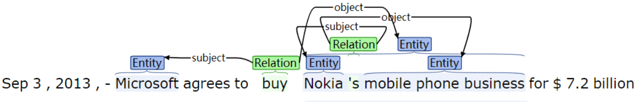


Figure: Event Extraction With Stanford Open IE tool

Candidate Tuples Of The Event:

1. (Microsoft, buy, Nokia's mobile phone business, Sep 3 2013)
2. (Nokia, 's, mobile phone business, Sep 3 2013)

Event Extraction

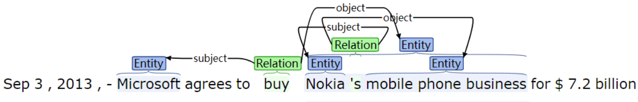


Figure: Event Extraction With Stanford Open IE tool

Candidate Tuples Of The Event:

1. (Microsoft, buy, Nokia's mobile phone business, Sep 3 2013)
2. (Nokia, 's, mobile phone business, Sep 3 2013)

Filter Out Bad Event Tuples

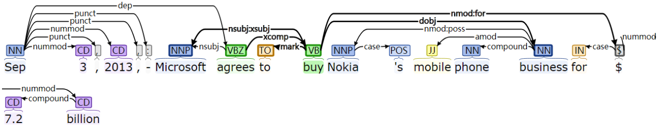


Figure: Filter Out Bad Tuples With tools like Dependency Parsing

Candidate Tuples Of The Event:

1. (Microsoft, buy, Nokia's mobile phone business, Sep 3 2013)
2. ~~(Nokia, 's, mobile phone business, Sep 3 2013)~~

Filter Out Bad Event Tuples

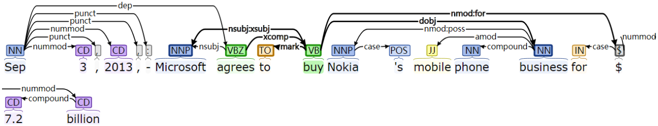


Figure: Filter Out Bad Tuples With tools like Dependency Parsing

Candidate Tuples Of The Event:

1. (Microsoft, buy, Nokia's mobile phone business, Sep 3 2013)
2. ~~(Nokia, 's, mobile phone business, Sep 3 2013)~~

Event Generalization

Example : **"Instant view: Private sector adds 114,000 jobs in July."**

$$(private\ sector, adds, 114\ 000\ jobs)$$

Event Generalization:

1. 'adds' and 'jobs' can be 'add' and 'job' respectively using some Stemmer tools
2. 'add' belongs to 'multiply' class With the help of VerbNet

$$(private\ sector, \underline{multiply\_class}, \underline{114\ 000\ job})$$

Event Generalization

Example : **"Instant view: Private sector adds 114,000 jobs in July."**

$$(private\ sector, adds, 114\ 000\ jobs)$$

Event Generalization:

1. 'adds' and 'jobs' can be 'add' and 'job' respectively using some Stemmer tools
2. 'add' belongs to 'multiply' class With the help of VerbNet

$$(private\ sector, \underline{multiply\_class}, \underline{114\ 000\ job})$$

Why Embedding?

Two events $(Nvidia\ fourth\ quarter\ results, miss, views)$ and $(Delta\ profit, didn't\ reach, estimates)$ has different structured representation, which increases the data sparsity

With Embedding these two things can have the similar vector. [Ding et al., 2015]

Embedding With Neural Tensor Network

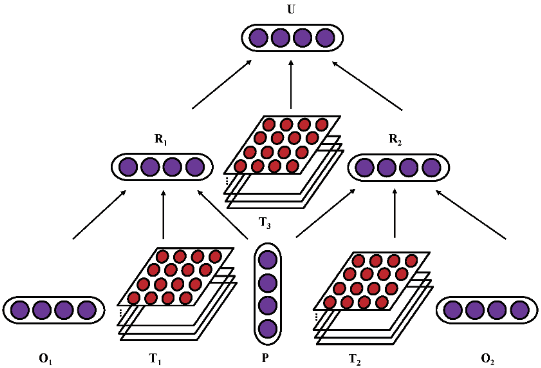


Figure: Neural tensor network for event embeddings

$$R_1 = f(O_1^T T_1^{[1:k]} P + W \begin{bmatrix} O_1 \\ P \end{bmatrix} + b_1) \text{ (the same with } R_2 \text{ and } U)$$

Neural Tensor Network Training

Training loss:

$$loss(E, E^{\tau}) = \max(0, 1 - f(E) + f(E^{\tau})) + \lambda \|\Phi\|_2^2$$

where $E = (O_1, P, O_2)$ and $E^{\tau} = (O_1^{\tau}, P, O_2)$

Algorithm 1: Event Embedding Training Process

Input: $\mathcal{E} = (E_1, E_2, \cdots, E_n)$ a set of event tuples; the model $EELM$

Output: updated model $EELM'$

random replace the event argument and got the corrupted event tuple

$\mathcal{E}^r \leftarrow (E_1^r, E_2^r, \cdots, E_n^r)$

while $\mathcal{E} \neq [\,]$ **do**

  $loss \leftarrow max\,(0\,,\,1 - f(E_i) + f(E_i^r) + \lambda \|\boldsymbol{\Phi}\|_2^2$

  if $loss > 0$ **then**

    $Update(\boldsymbol{\Phi})$

  else

    $\mathcal{E} \leftarrow \mathcal{E}/\{E_i\}$

return $EELM$

Embeddding With External Knowledge

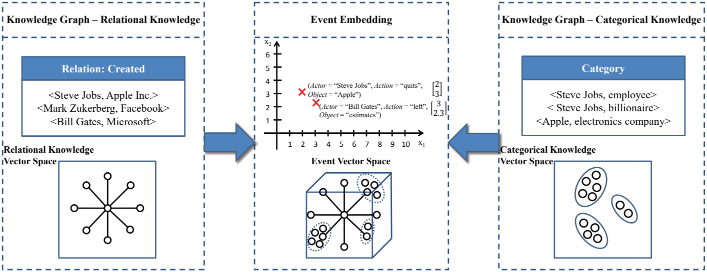


Figure: Incorporating knowledge graph into the learning process for event embeddings [Ding et al., 2016]

Trainging With External Knowledge

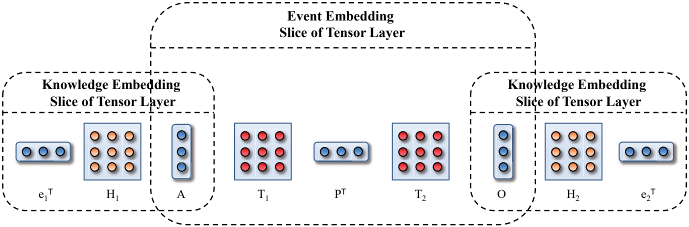


Figure: Architecture of the joint embedding model (only showing the tensor layer)

$$L = \alpha L_\epsilon + (1-\alpha) L_\kappa$$
$$L_\epsilon = loss(E, E^\tau) = \max(0, 1 - f(E) + f(E^\tau)) + \lambda \|\Phi\|_2^2$$
$$L_\kappa = \sum_{i=1}^{N} \sum_{m=1}^{M} \max(0, 1 - g(T^{(i)}) + g(T_c^{(i)})) + \lambda \|\Omega\|_2^2$$
where $E = (O_1, P, O_2)$, $T^{(i)} = (e_1^{(i)}, R, e_2^{(i)})$

Trainging With External Knowledge

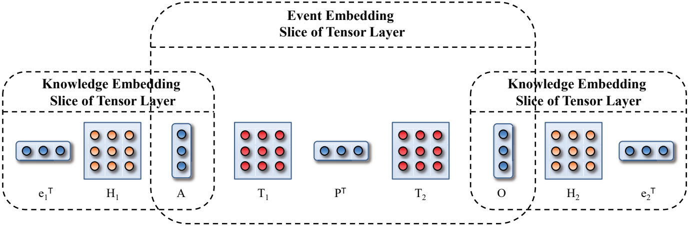


Figure: Architecture of the joint embedding model (only showing the tensor layer)

$$L = \alpha L_\epsilon + (1-\alpha) L_\kappa$$
$$L_\epsilon = loss(E, E^\tau) = \max(0, 1 - f(E) + f(E^\tau)) + \lambda \|\Phi\|_2^2$$
$$L_\kappa = \sum_{i=1}^{N} \sum_{m=1}^{M} \max(0, 1 - g(T^{(i)}) + g(T_c^{(i)})) + \lambda \|\Omega\|_2^2$$
where $E = (O_1, P, O_2)$, $T^{(i)} = (e_1^{(i)}, R, e_2^{(i)})$

Prediction Model

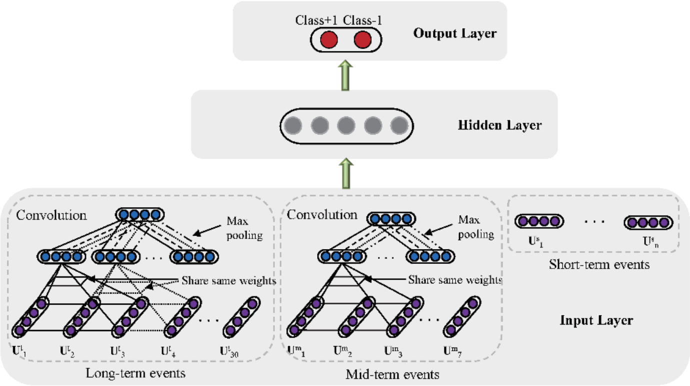
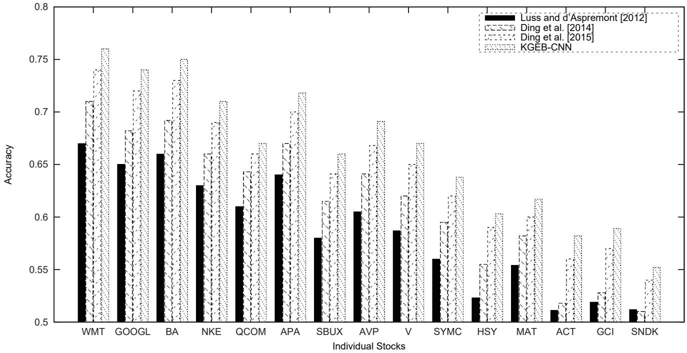


Figure: Architecture of the prediction model based on a deep convolutional neural network.

Table 3: Experimental results on index prediction.

	Acc	MCC
Luss and d'Aspremont (2012)	56.38%	0.0711
Ding et al. (2014)	58.83%	0.1623
WB-CNN	60.57%	0.1986
Ding et al. (2015)	64.21%	0.4035
KGEB-CNN	**66.93%**	**0.5072**



Thank You!

- Ding, X., Zhang, Y., Liu, T., and Duan, J. (2014).
  Using structured events to predict stock price movement: An empirical investigation.
  In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1415–1425, Doha, Qatar. Association for Computational Linguistics.

- Ding, X., Zhang, Y., Liu, T., and Duan, J. (2015).
  Deep learning for event-driven stock prediction.
  In Proceedings of IJCAI, BueNos Aires, Argentina.

- Ding, X., Zhang, Y., Liu, T., and Duan, J. (2016).
  Knowledge-driven event embedding for stock prediction.
  In Proceedings of COLING 2016, pages 2133–2142, Osaka, Japan. Association for Computational Linguistics.