HIT-SCIR

Harbin Institute of Technology
Research Center of
Social Computing and Information Retrieval

Deep Learning for Event-Driven Stock Prediction

Xiao Ding[1], Yue Zhang[2], Ting Liu[1], Junwen Duan[1]

[1]Research Center for Social Computing and Information Retrieval,
Harbin Institute of Technology, China

[2]Singapore University of Technology and Design, Singapore

[1]{xding, tliu, jwduan}@ir.hit.edu.cn; [2]yue_zhang@sutd.edu.sg

Outline

Introduction

- Traditional stock prediction
  - Using simple features from news documents, such as bags-of-words, noun phrases, and named entities
  - These features do not capture structured relations, which limits their potentials

Accuser

Microsoft sues Barnes & Noble → {“Microsoft”, “sues”, “Barnes”, “Noble”}

Defendant

Introduction

- Event-driven stock prediction
  - Using open information extraction (Open IE) to obtain structured events representations [Ding et al., 2014]
  - Improved stock market prediction using structured representation instead of words as features

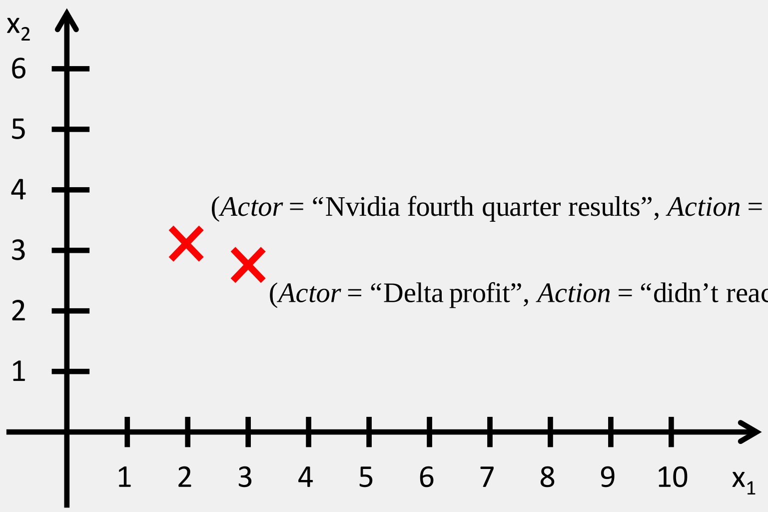
Microsoft sues Barnes & Noble →
(*Actor* = "Microsoft", *Action* = "sues", *Object* = "Barnes & Noble")

One disadvantage of structured representations of events is that they lead to increased sparsity, which potentially limits the predictive power.

Introduction

- Event embedding
  - Low-dimensional, dense, real-valued
  - In theory, embeddings are appropriate for achieving good results with a density estimator, which can misbehave in high dimensions

$$(\textit{Actor} = \text{“Microsoft”}, \textit{Action} = \text{“sues”}, \textit{Object} = \text{“Barnes \& Noble”}) = \begin{bmatrix} 0.286 \\ 0.792 \\ -0.177 \\ -0.235 \\ 0.348 \\ -0.784 \\ 0.963 \\ 0.128 \\ -0.289 \\ \vdots \end{bmatrix}$$

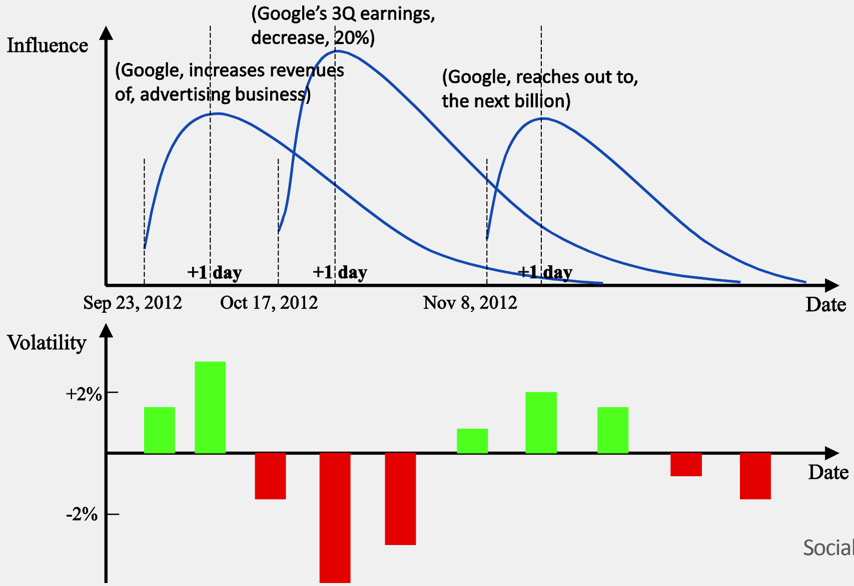


$(\textit{Actor} = \text{“Nvidia fourth quarter results”}, \textit{Action} = \text{“miss”}, \textit{Object} = \text{“views”})$ $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$

$(\textit{Actor} = \text{“Delta profit”}, \textit{Action} = \text{“didn’t reach”}, \textit{Object} = \text{“estimates”})$ $\begin{bmatrix} 3 \\ 2.5 \end{bmatrix}$

Introduction

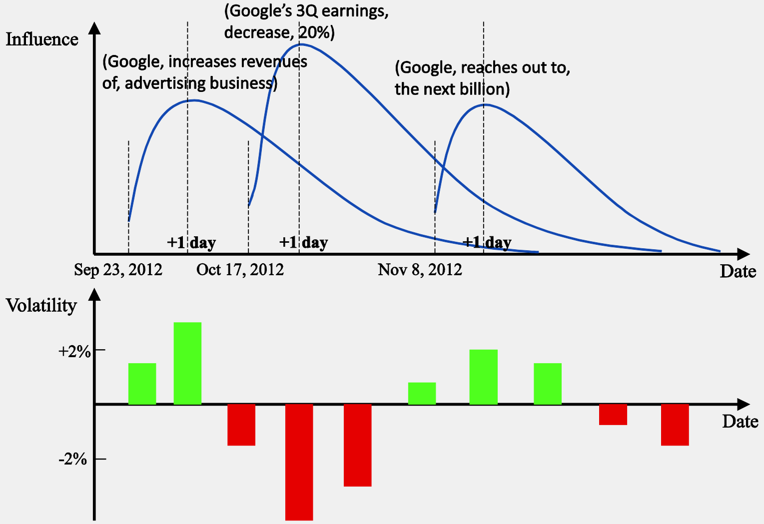
- Deep prediction model
  - Capture the influence of news events over a history that is longer than a day based on deep prediction model
  - Research shows diminishing effects of reported events on stock market volatility [Xie et al., 2013]



Harbin Institute of Technology
Research Center of
Social Computing and Information Retrieval

Deep Prediction Model

- The influences of three actual events for Google Inc. in the year 2012 was the highest on the second day, but gradually weakened over time
- Despite the relatively weaker effects of long-term events, the volatility of stock markets is still affected by them
- Little previous work quantitatively models combined short-term and long-term effects of events
- Treat history news as daily event sequences, using a convolutional neural network (CNN) to model short-term and long-term effects of events



Main Method

Neural Tensor Network for Learning Event Embeddings

- Event Representation and Extraction

Representation

Actor **Action** **Object** **Timestamp**

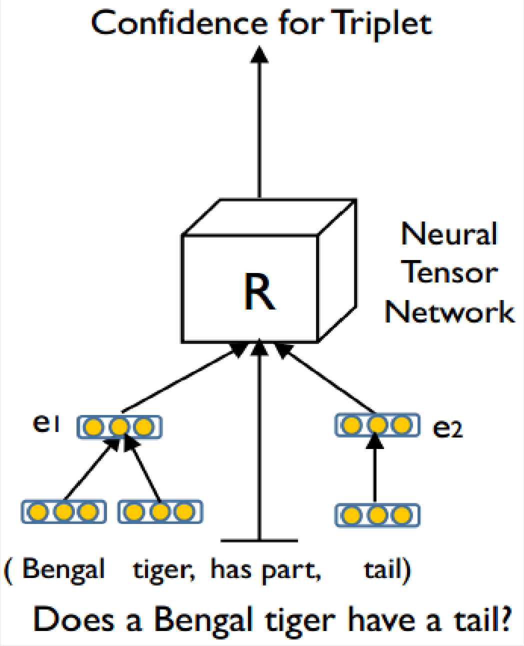
$$E = (O_1, P, O_2, T)$$

Extraction

Jan 13, 2014 – Google Acquires Thermostat Maker Nest for \$3.2 billion

Event Embedding

- Related previous work
  - Learning distributed representations of multi-relational data from knowledge bases, which learns the embedding of (e_1; R; e_2), where e_1 and e_2 are named entities and R is the relation type. (Socher *et al*., 2013)



Differences with Previous Work

- The number of relation types in knowledge bases is limited
  - Most previous work models a relation type by using a matrix or a tensor, and train a model for each specific relation type
  - The event types is an open set, so it is more difficult to train a specific model for each event type
- The goal of relational database embedding is to be able to state whether two entities (e_1; e_2) are in a certain relation R
  - When R is symmetric, e1 and e2 have interchangeable roles. In contrast, each argument of the event has a specific role, which is not interchangeable

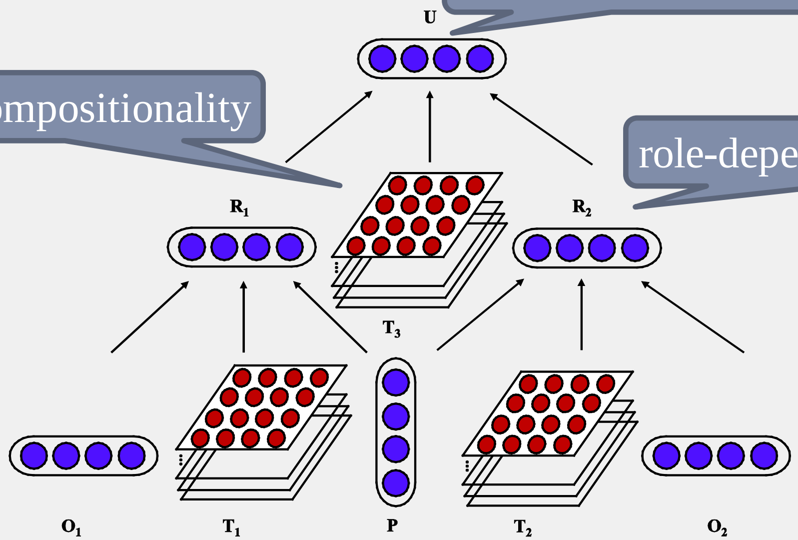
Neural Tensor Network for Event Embedding

- Input: word embeddings
- Output: event embeddings

structured embedding

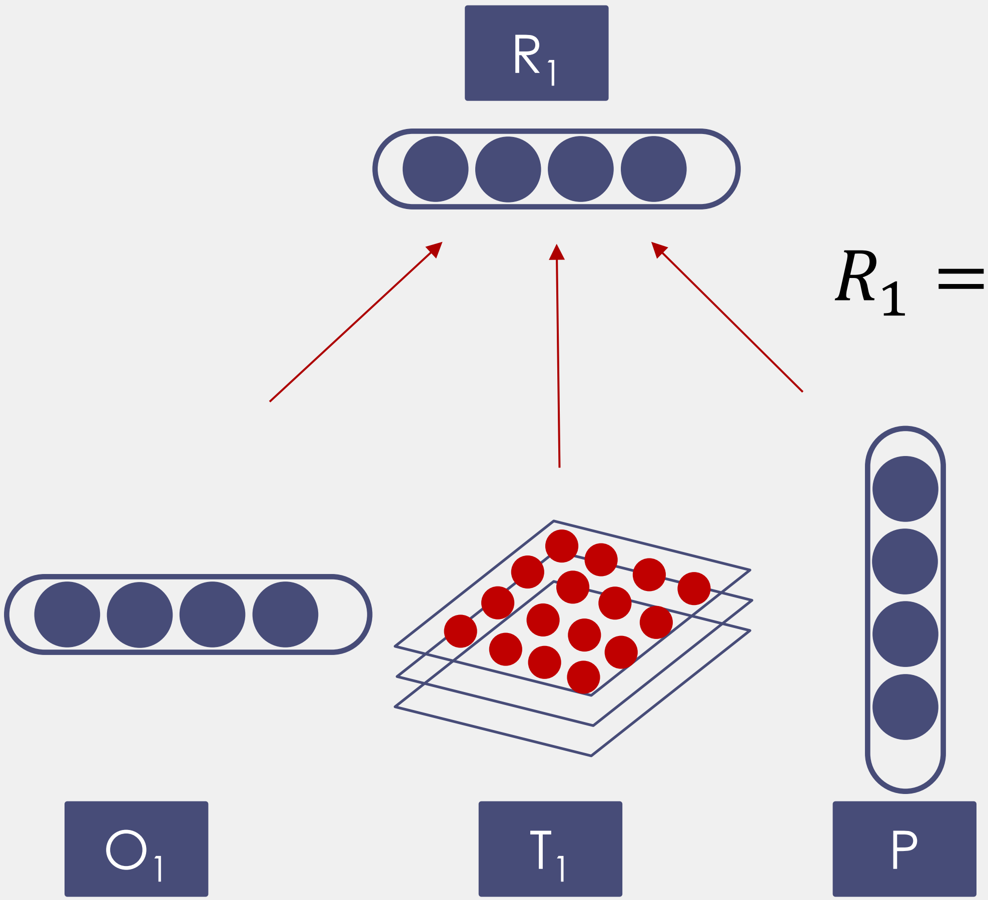
semantic compositionality

role-dependent embedding



As most event arguments consist of several words, we represent the actor, action and object as the average of its word embeddings, respectively

Neural Tensor Network for Event Embedding



tanh

weight matrix

$$R_1 = f\left(O_1^T T_1^{[1:k]} P + W \begin{bmatrix} O_1 \\ P \end{bmatrix} + b\right)$$

bilinear tensor product

bias

Training

- Assume that event tuples in the training set should be given a higher score than corrupted tuples, in which one of the event arguments is replaced with a random argument

Parameters

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

Random replace with an object

Regulation weight,set to 0.0001

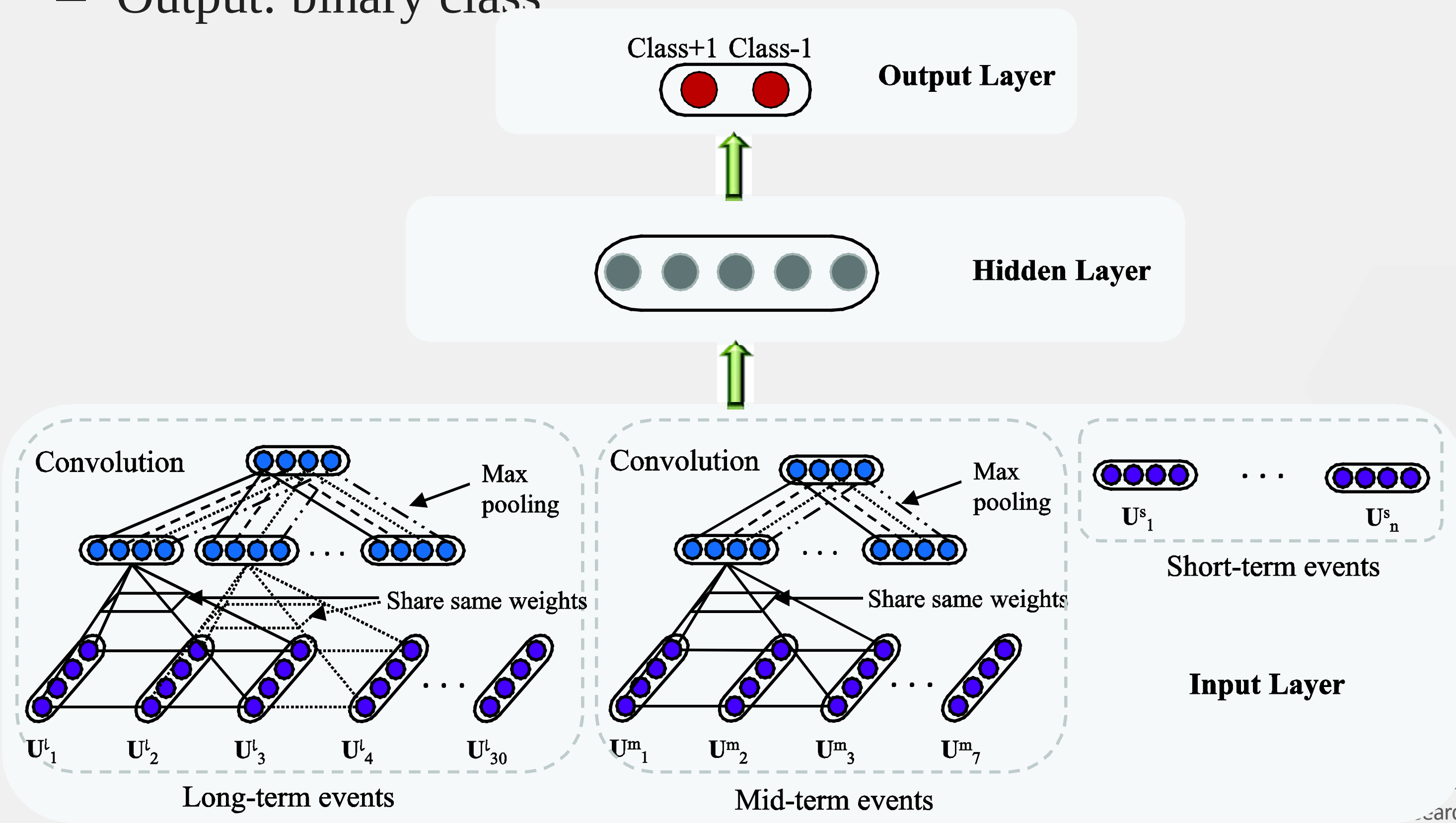
Deep Prediction Model

- Model long-, mid-, short-term events
  - Long-term events (Last month)
  - Mid-term events (Last week)
  - Short-term events (Last day)

- The prediction model learns the effect of these three different time spans on stock prices based on the framework of a CNN

Deep Prediction Model

- Architecture
  - Input: a sequence of event embeddings, arranged in chronological order
  - Output: binary class



Deep Prediction Model

- Convolution and Max-pooling
  - Convolutional layer to obtain local feature
    - Model the effect of each individual event

$$Q_j = W_1^T U_{j-l+1:j}$$

  - Max-pooling to determine the global representative feature
    - Model the combination effect of all events

$$V_j = \max Q(j, \cdot)$$

Note that the convolution operation is only applied to the long-term and mid-term event embeddings, because the unit of timing is one day

Experiment

- Dataset
  - Financial news are from Reuters and Bloomberg news
  - Predicting the Standard & Poor's 500 stock (S&P 500) index and its individual stocks

	Training	Development	Test
#documents	442,933	110,733	110,733
#words	333,287,477	83,247,132	83,321,869
#events	295,791	34,868	35,603
time interval	02/10/2006 - 18/06/2012	19/06/2012 - 21/02/2013	22/02/2013 - 21/11/2013

Table 1: Statistics of datasets.

Download URL: http://ir.hit.edu.cn/~xding/index_english.htm/

Index Prediction

- Baselines

	Feature	Model
Luss and d'Aspremont [2012]	Bag of words	SVM
Ding et al. [2014] (E-NN)	Structured event	NN
WB-NN	Word embedding	NN
WB-CNN	Word embedding	CNN
E-CNN	Structured event	CNN
EB-NN	Event embedding	NN
EB-CNN	Event embedding	CNN

Index Prediction

- Results
  - Events are better features than words for stock market prediction
  - Event embedding is useful for the task of stock market prediction
    - Low-dimensional dense vector can effectively alleviate the problem of feature sparsity
    - Deeper semantic relations between event embeddings can be learned by modeling the semantic compositionality over word embeddings

	Acc	MCC
Luss and d'Aspremont [2012]	56.42%	0.0711
Ding et al. [2014] (E-NN)	58.94%	0.1649
WB-NN	60.25%	0.1958
WB-CNN	61.73%	0.2147
E-CNN	61.45%	0.2036
EB-NN	62.84%	0.3472
EB-CNN	**65.08%**	**0.4357**

Table 2: Development results of index prediction.

Index Prediction

- Results
  - CNN-based prediction models are more powerful than NN-based prediction models
    - CNN can quantitatively analyze the influence of the history events over longer terms, and can extract the most representative feature vector for the prediction model

	Acc	MCC
Luss and d'Aspremont [2012]	56.42%	0.0711
Ding et al. [2014] (E-NN)	58.94%	0.1649
WB-NN	60.25%	0.1958
WB-CNN	61.73%	0.2147
E-CNN	61.45%	0.2036
EB-NN	62.84%	0.3472
EB-CNN	**65.08%**	**0.4357**

Table 2: Development results of index prediction.

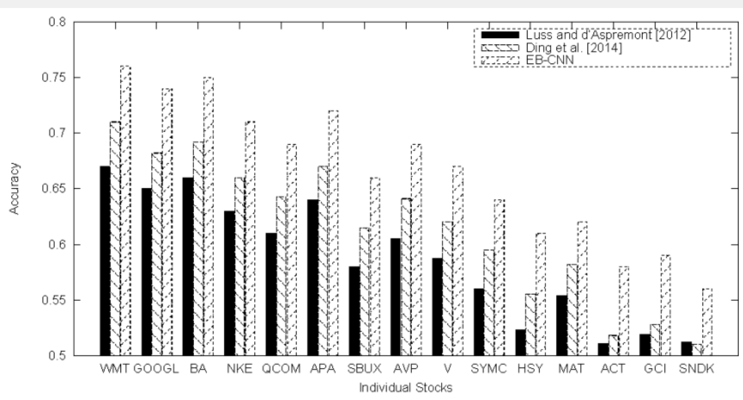
Individual Stock Prediction

- 15 companies from S&P 500
  - Consists of high-,mid- and low-ranking companies according to the Fortune Magazine
  - Evaluation metric: Accuarcy and MCC
    - Using MCC to avoid bias due to data skew

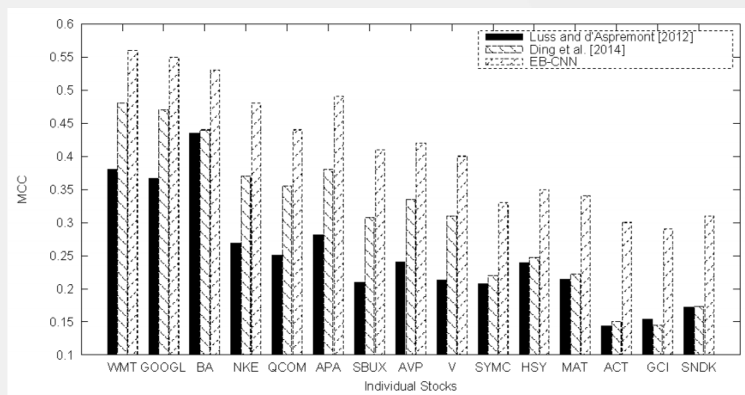
$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

Individual Stock Prediction

- Results
  - Our model achieves consistently better performance compared to the baseline methods, on both individual stock and index prediction
  - Our model achieves relatively higher improvements on those lower fortune ranking companies compared with baseline methods
    - Our model considers the diminishing influence of monthly news and weekly news, which are important features for individual stock prediction
    - Even without daily news, our model can also give relatively accurate prediction results



(a) Accuarcy



(b) MCC

Conlcusion

- Deep learning is useful for event-driven stock price movement prediction
- Event embeddings-based document representations are better than discrete events-based methods
- Deep CNN can help capture longer-term influence of news event

Acknowledgements

- We gratefully acknowledge the support of
  - The National Basic Research Program (973 Program) of China via Grant 2014CB340503
  - The National Natural Science Foundation of China (NSFC) via Grant 61133012 and 61472107
  - The TL SUTD grant IGDST1403012 and SRG ISTD 2012 038 from Singapore University of Technology and Design.

Thanks!
Q&A