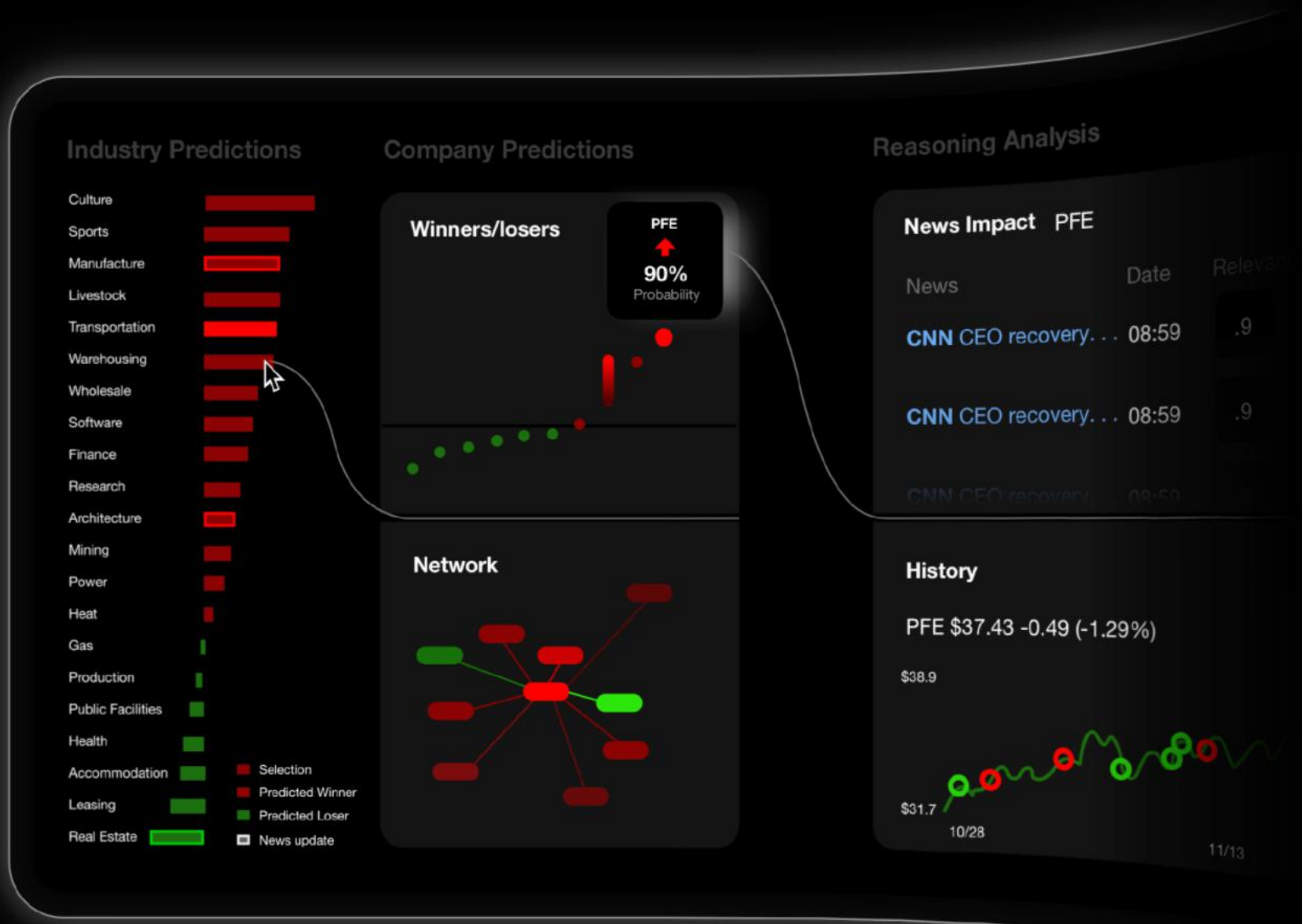
OPEN

INSIDE THE BLACK BOX

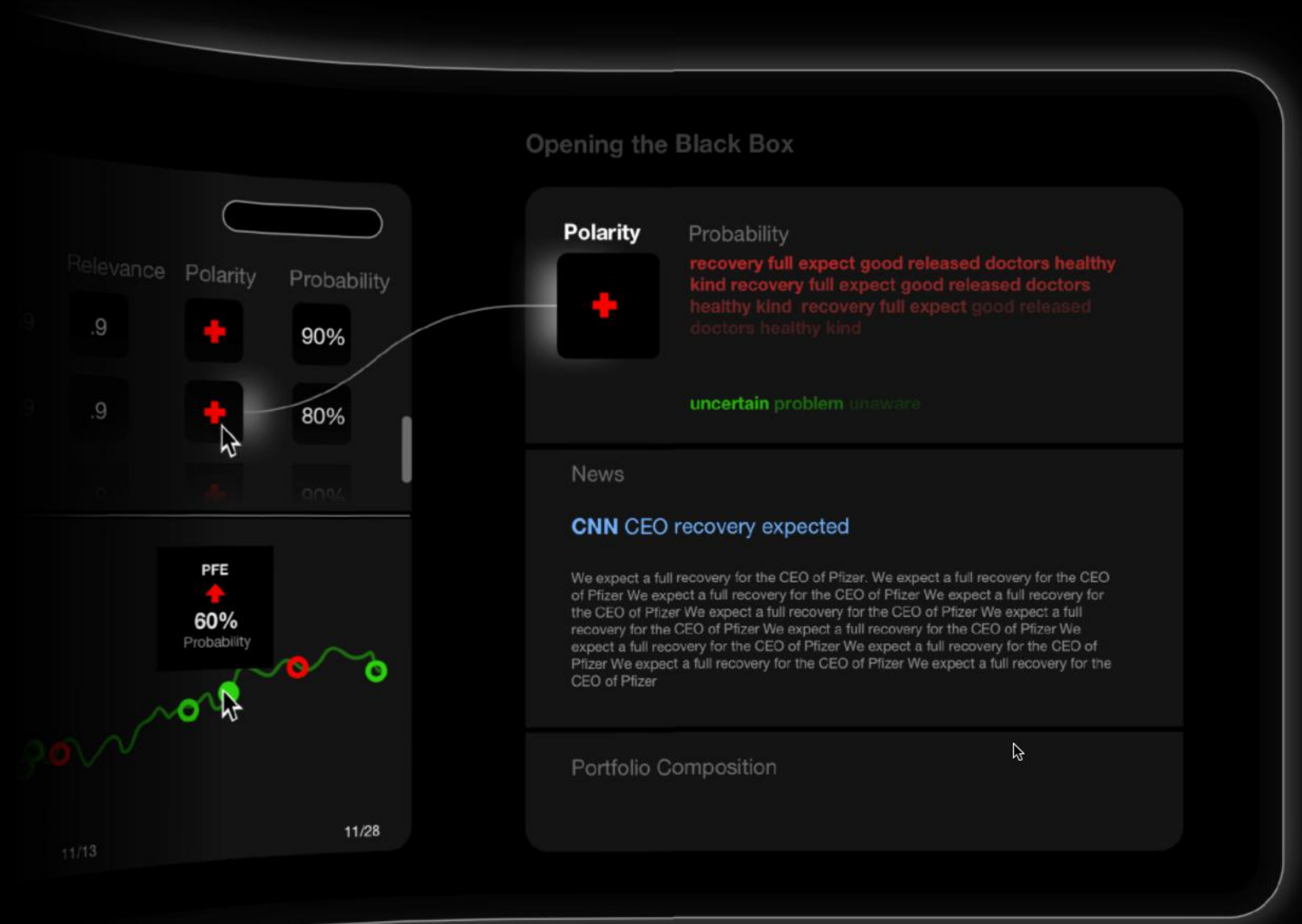
清華大學
Tsinghua University

This presentation contains live data

Press the right key to play the presentation


Industry Predictions
Culture
Sports
Manufacture
Livestock
Transportation
Warehousing
Wholesale
Software
Finance
Research
Architecture
Mining
Power
Heat
Gas
Production
Public Facilities
Health
Accommodation
Leasing
Real Estate
Selection
Predicted Winner
Predicted Loser
News update
Company Predictions
Winners/losers
PFE
90%
Probability
Network
Reasoning Analysis
News Impact PFE
News
Date
CNN CEO recovery. . .
08:59
CNN CEO recovery. . .
08:59
History
PFE $37.43 -0.49 (-1.29%)
$38.9
$31.7
10/28
11/13


Reasoning Analysis
PFE
90%
Probability
News Impact PFE
News
Date
Relevance
Polarity
Probability
CNN CEO recovery. . .
08:59
.9
90%
CNN CEO recovery. . .
08:59
.9
80%
History
PFE $37.43 -0.49 (-1.29%)
PFE
60%
Probability
$38.9
$31.7
10/28
11/13
11/28
Polarity
News
CNN CEO
Portfolio


Opening the Black Box
Relevance
Polarity
Probability
.9
90%
.9
80%
PFE
60%
Probability
11/13
11/28
Polarity
Probability
recovery full expect good released doctors healthy kind recovery full expect good released doctors healthy kind recovery full expect good released doctors healthy kind
uncertain problem unaware
News
CNN CEO recovery expected
We expect a full recovery for the CEO of Pfizer. We expect a full recovery for the CEO of Pfizer We expect a full recovery for the CEO of Pfizer We expect a full recovery for the CEO of Pfizer We expect a full recovery for the CEO of Pfizer We expect a full recovery for the CEO of Pfizer We expect a full recovery for the CEO of Pfizer We expect a full recovery for the CEO of Pfizer We expect a full recovery for the CEO of Pfizer We expect a full recovery for the CEO of Pfizer We expect a full recovery for the CEO of Pfizer
Portfolio Composition

Problem Background

Innacurate

Traditional statistical and ML models fail prediction accuracy for quantitative investment

Irrelevant

Current financial terminals merely deliver news and stock performance in an independent manner

Unexplainable

Investors and analysts face the black–box problem and the predicted result lacks credibility

Motivations

Innacurate

Develop a deep model to enhance prediction performance

Irrelevant

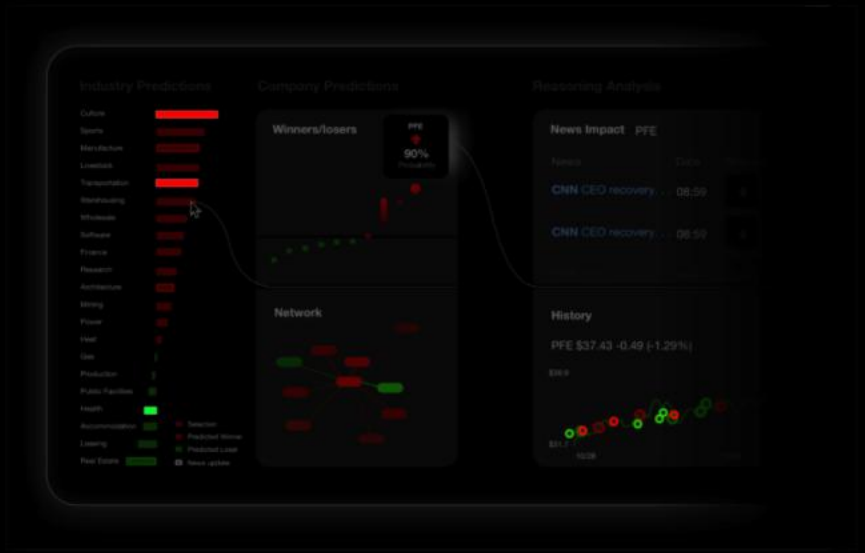
Build the bridge between the news and stock price trend

Unexplainable

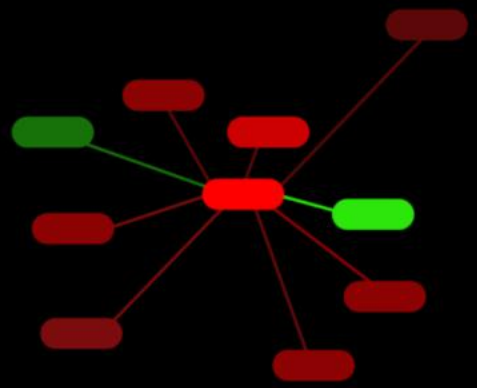
Interprete the model output for investors

Product Introduction

Features



Emphasize market signals



Locate the Most Influencial Firms

Probability

recovery full expect good released doctors healthy kind recovery full expect good released doctors healthy kind recovery full expect good released doctors healthy kind

uncertain problem unaware

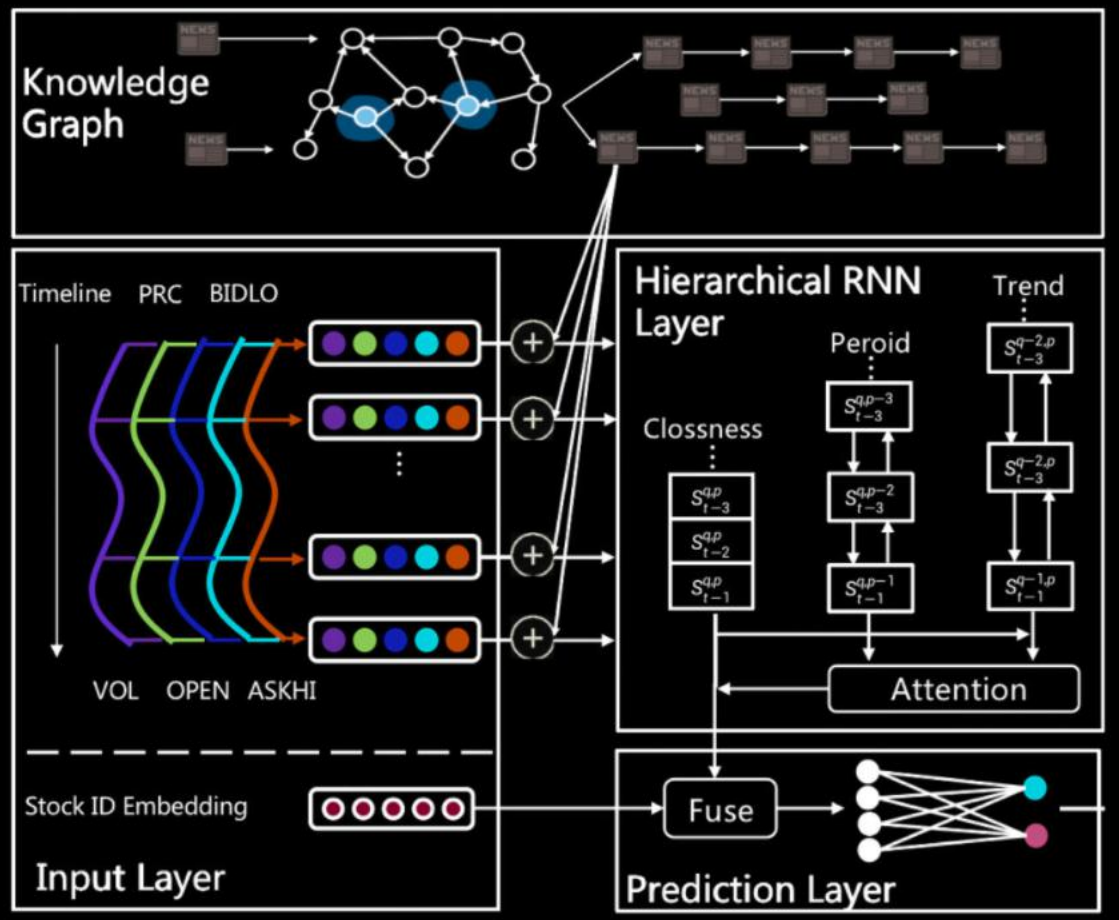
Key Information Extraction

Model Structure

Hierarchical-LSTM

Knowledge-Enhanced Model

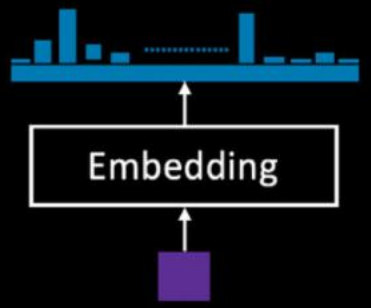
Pearson-based Distance Relation



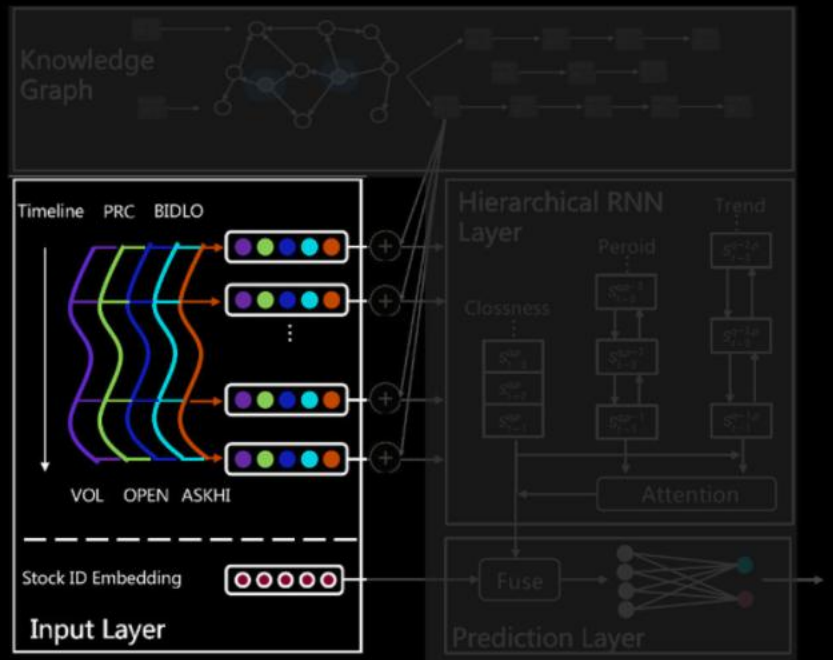
Model Structure

Feature Aggregation

- **Min-Max Normalization**
  - Multi-dimension (VOL, PRC, OPEN, BIDLO, ASKHI)
  - $x = \frac{x - \min}{\max - \min}$
- **Batch Normalization**
  - Normalization across multiple features
- **Stock ID Embedding**
  - vector



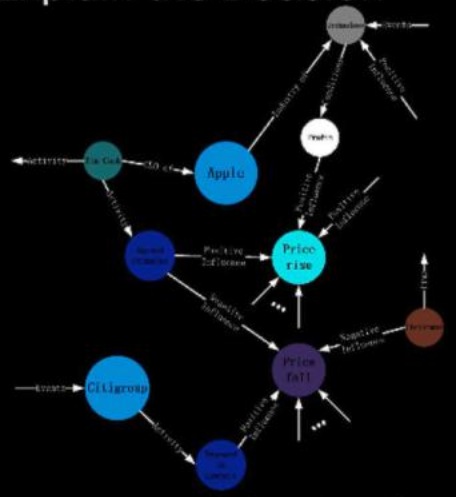
Muiti-dimention feature aggregation takes key features into account and then normalize these feature into [0,1]
Stock ID embedding tries to find inter-correlations between different stocks



Model Structure

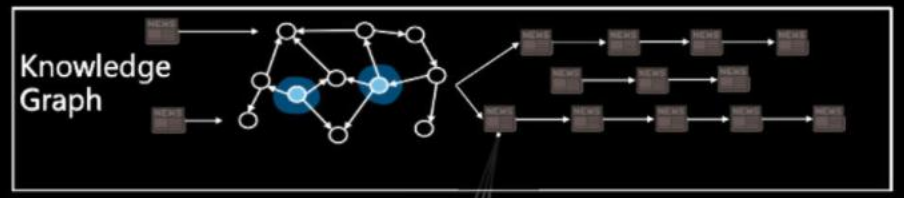
Knowledge Graph

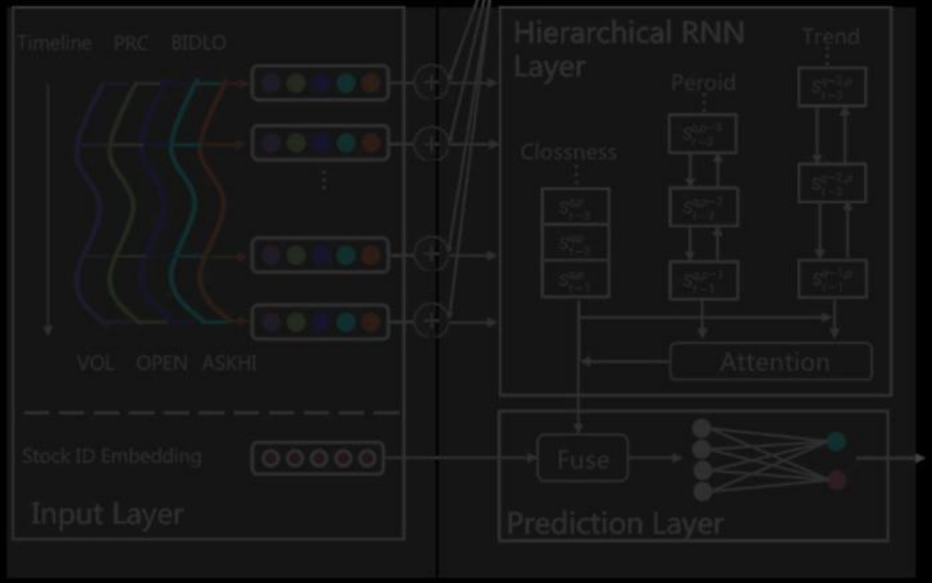
- Understanding News
- Events Relation
- Logic behind the stock fluctuations
- Explain the Decision



Knowledge Graph

News: Tim Cook sells large amount of his stocks.
Knowledge Graph:
Tim Cook -> Apple Inc.->sells stocks--> stocks declines



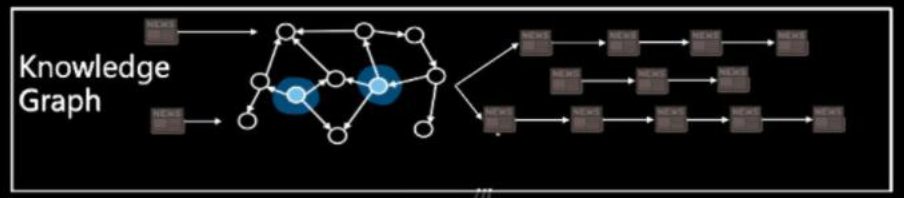


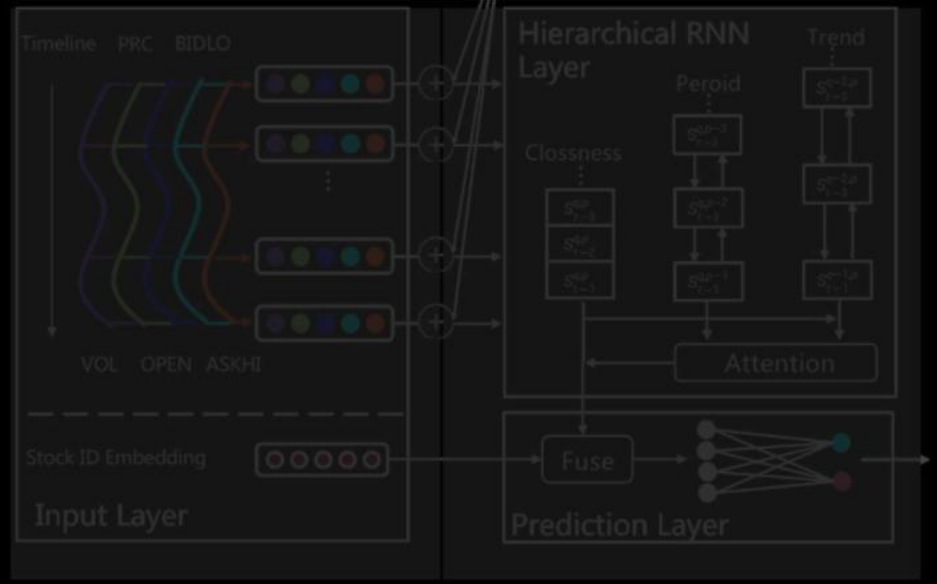
Model Structure

Knowledge Graph

- Named Entity Recognition
- Meta-path Recognition
- Meta-path Encoding
- Feed to Model

News: Tim Cook sells large amount of his stocks.
Knowledge Graph:
Tim Cook -> Apple Inc.->sells stocks--> stocks declines

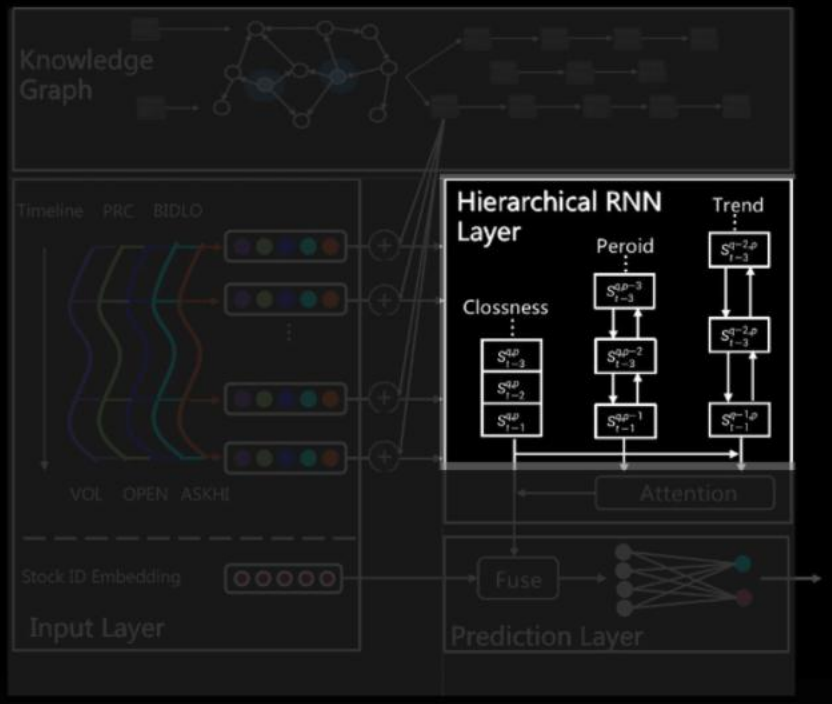




Model Structure

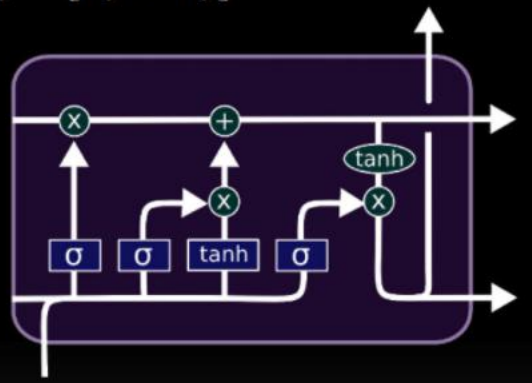
Hierarchical LSTM

Bidirectional Long Short Term Memory (BiLSTM) captures high-order and complex sequential information in stock sequence. Hirarchical RNN learns quite long-term sequential information



- Hierarchical RNN Layer
  - BiLSTM module
  - Attention module

$$i_t = \sigma_i(x_t W_{xi} + h_{t-1} W_{hi} + b_i)$$
$$f_t = \sigma_f(x_t W_{xf} + h_{t-1} W_{hf} + b_f)$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \sigma_c(x_t W_{xc} + h_{t-1} W_{hc} + b_c)$$
$$o_t = \sigma_o(x_t W_{xo} + h_{t-1} W_{ho} + b_o)$$
$$h_t = o_t \odot \sigma_h(c_t)$$
$$h_t = [\overrightarrow{h_t} \oplus \overleftarrow{h_t}]$$

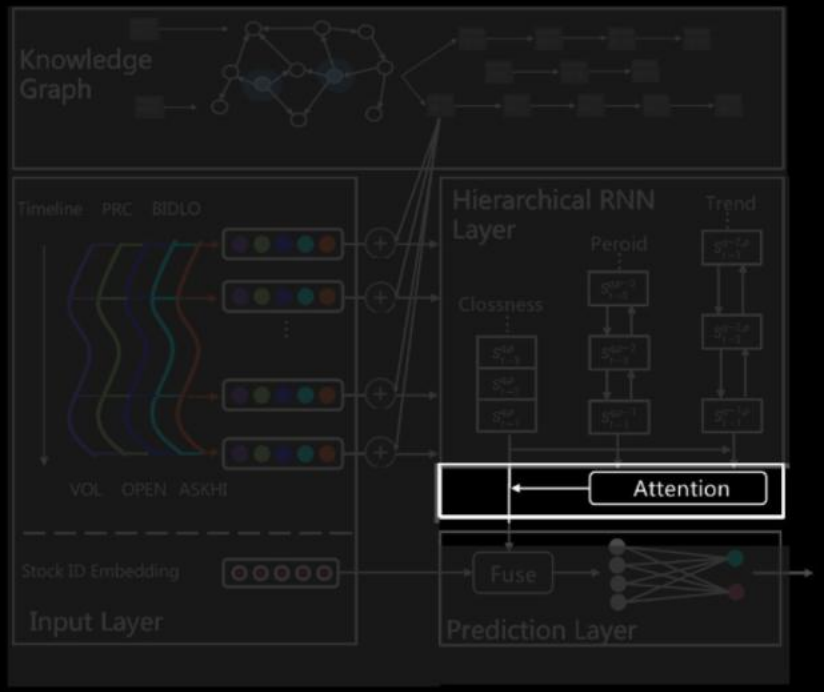


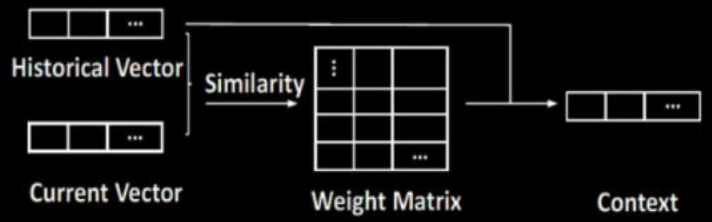
Model Structure

Attention

Attention module combines clossness, period and trend information, and learns multi-periodic nature of stock price to augment the BiLSTM for stock prediction.

- Hierarchical RNN Layer
  - BiLSTM module
  - Attention module
    - Clossness with Period and trend





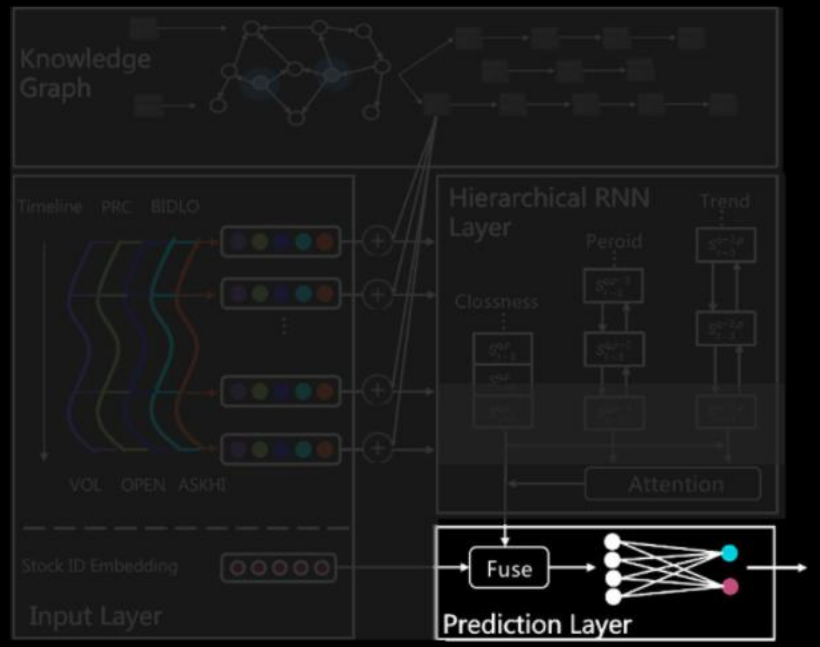
$$a_{i,j} = \frac{exp(f(h_t, h_{i,j}))}{\sum_{i1} \sum_{j1} exp(f(h_t, h_{i1,j1}))}$$

$$c = \sum_i \sum_j a_{i,j} h_{i,j}$$

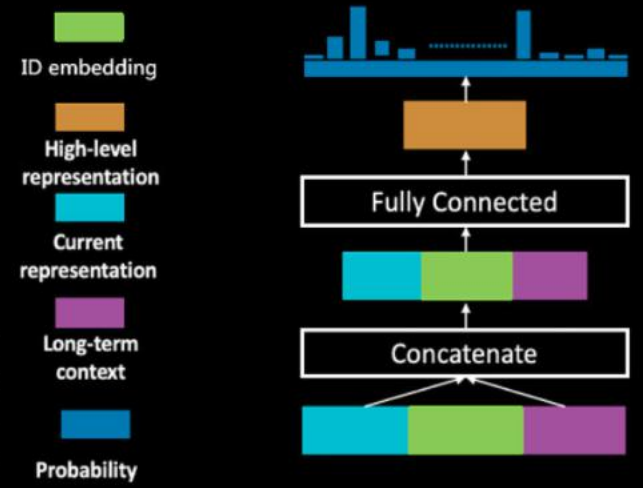
$$f(h_t, h_{i,j}) = h_t^T W_q$$

Model Structure

Prediction Layer



- **Prediction Layer**
  - Concatenate (Current representation, long-term context and Stock type)
  - Fully connection
  - Softmax



Current representation, long-term context and stock embedding are concatenated to get more high-level representations with full-connection layer.
We project them into an output layer associated with the probability to conduct prediction.

Model Structure

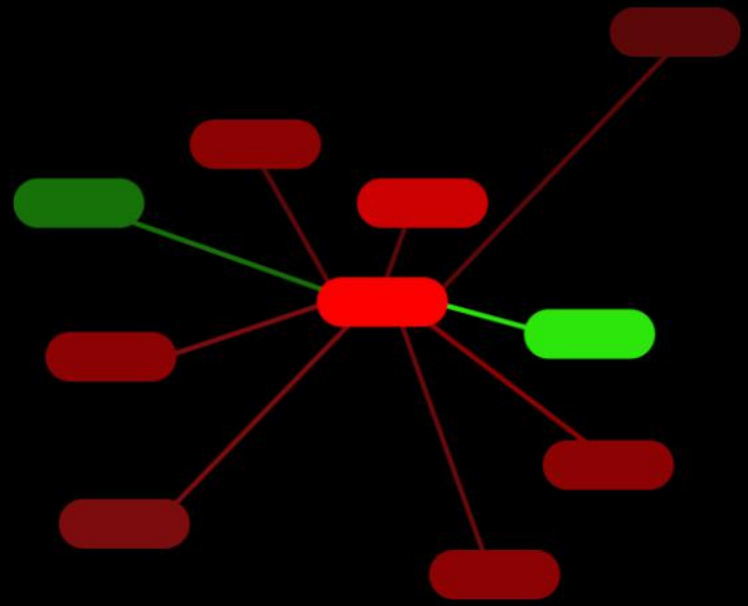
Inter–firm Distance

$$r_{i,j}^{PRC} = \frac{Cov(\Delta PRC_i, \Delta PRC_j)}{\sqrt{Var(\Delta PRC_i)Var(\Delta PRC_j)}}$$

$$r_{i,j}^{VOL} = \frac{Cov(\Delta VOL_i, \Delta VOL_j)}{\sqrt{Var(\Delta VOL_i)Var(\Delta VOL_j)}}$$

$$d_{i,j} = \sqrt{(1 - \text{rank}(r_{i,j}^{PRC}))^2 + (1 - \text{rank}(r_{i,j}^{PRC}))^2}$$

Relation Map



Strategy Performance

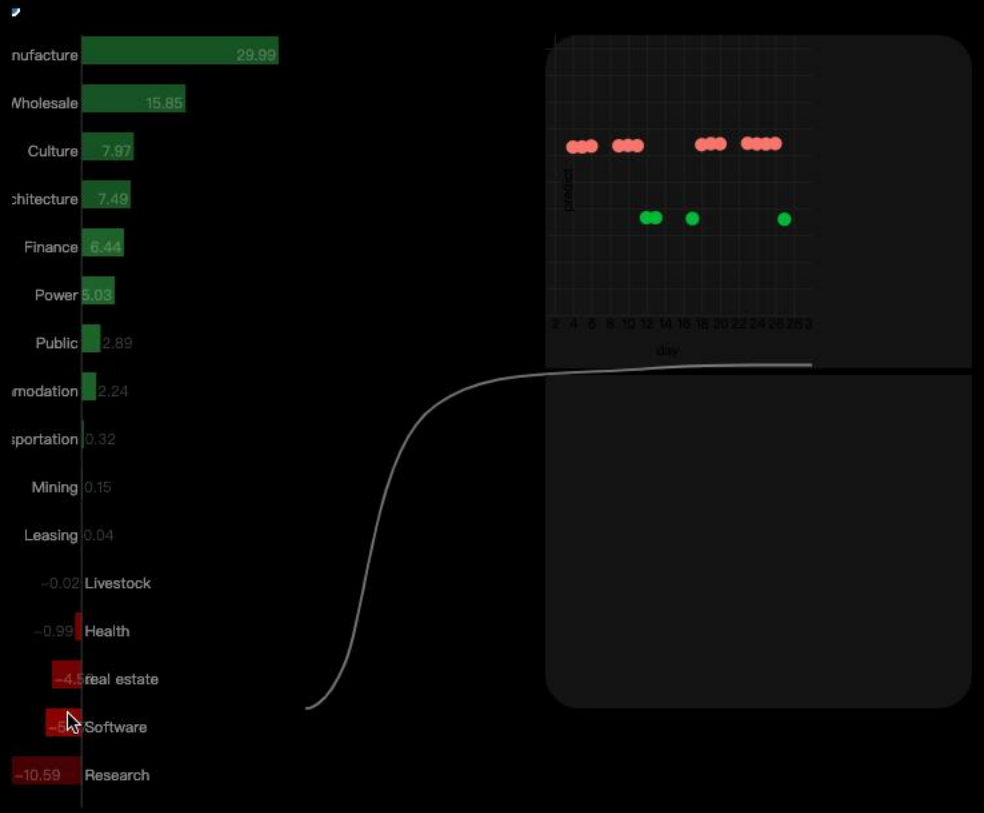
Sample

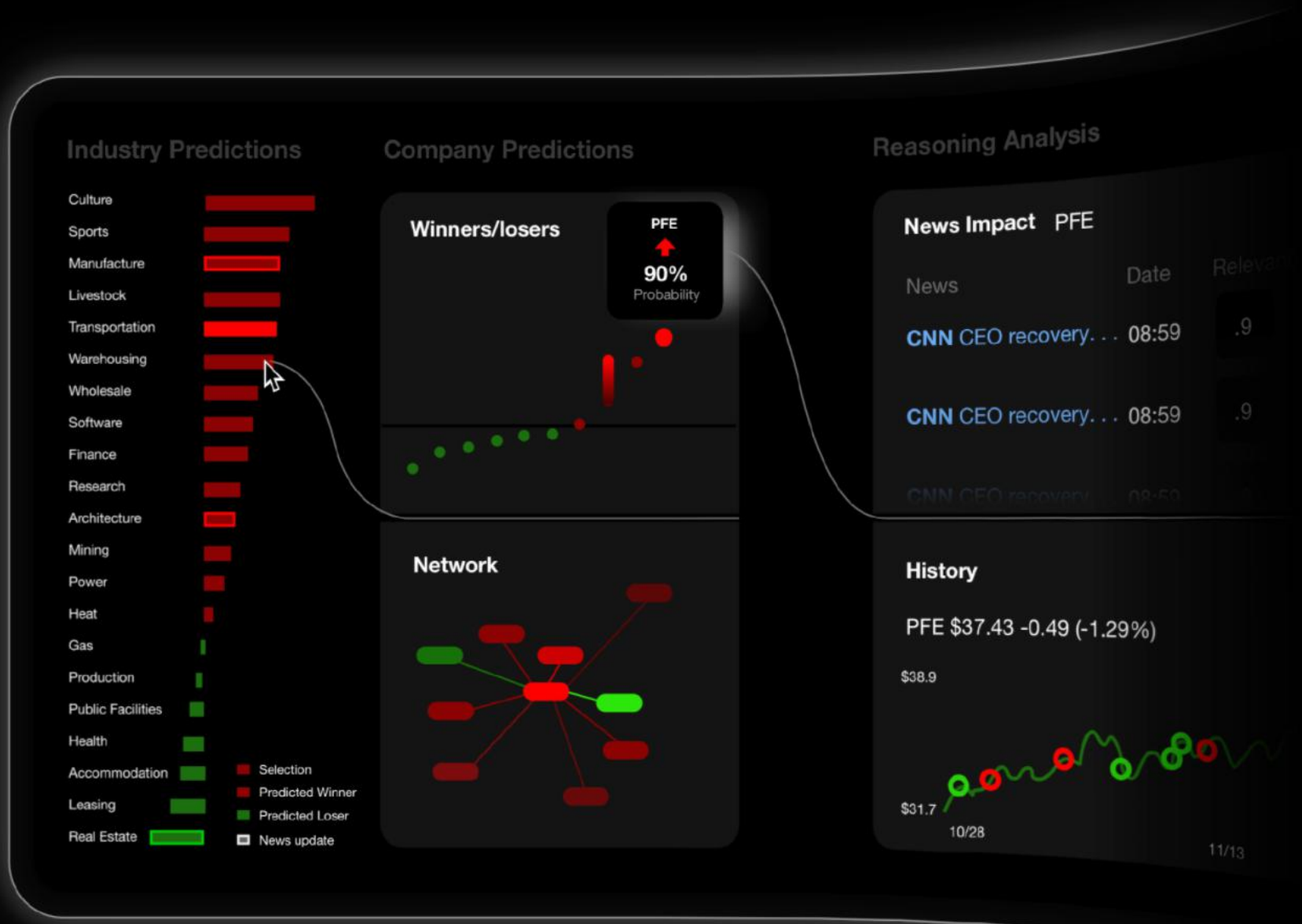
Annual return: 34.09%

Excess return: 9.09%

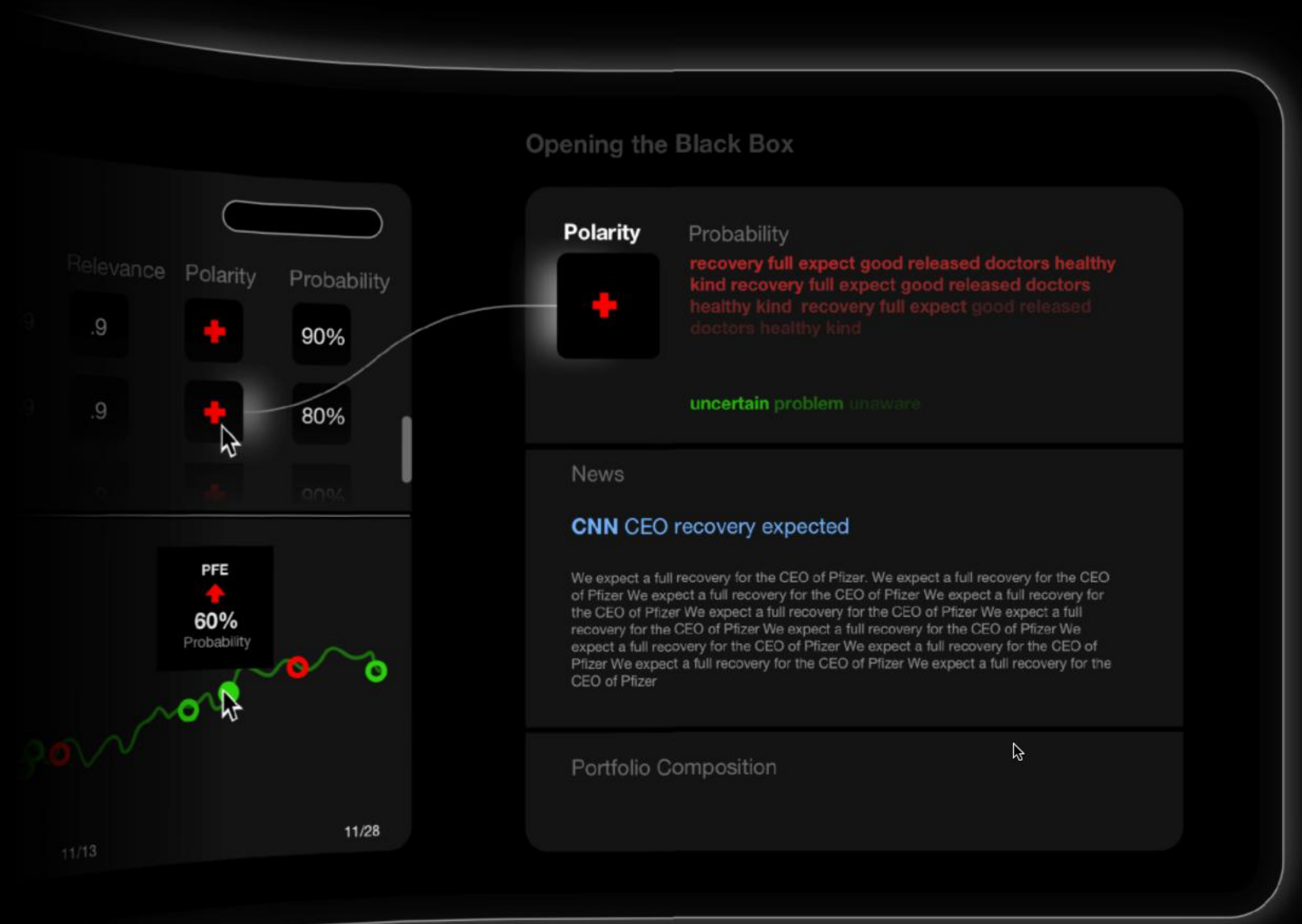
Global Accuracy: 55%

Demo With Real World Data




Industry Predictions
Culture
Sports
Manufacture
Livestock
Transportation
Warehousing
Wholesale
Software
Finance
Research
Architecture
Mining
Power
Heat
Gas
Production
Public Facilities
Health
Accommodation
Leasing
Real Estate
Selection
Predicted Winner
Predicted Loser
News update
Company Predictions
Winners/losers
PFE
90%
Probability
Network
Reasoning Analysis
News Impact PFE
News
Date
CNN CEO recovery. . .
08:59
CNN CEO recovery. . .
08:59
History
PFE $37.43 -0.49 (-1.29%)
$38.9
$31.7
10/28
11/13


Reasoning Analysis
PFE
90%
Probability
News Impact PFE
News
Date
Relevance
Polarity
Probability
CNN CEO recovery. . .
08:59
.9
90%
CNN CEO recovery. . .
08:59
.9
80%
History
PFE $37.43 -0.49 (-1.29%)
PFE
60%
Probability
$38.9
$31.7
10/28
11/13
11/28
Polarity
News
CNN CEO
Portfolio


Opening the Black Box
Relevance
Polarity
Probability
.9
90%
.9
80%
PFE
60%
Probability
11/13
11/28
Polarity
Probability
recovery full expect good released doctors healthy kind recovery full expect good released doctors healthy kind recovery full expect good released doctors healthy kind
uncertain problem unaware
News
CNN CEO recovery expected
We expect a full recovery for the CEO of Pfizer. We expect a full recovery for the CEO of Pfizer We expect a full recovery for the CEO of Pfizer We expect a full recovery for the CEO of Pfizer We expect a full recovery for the CEO of Pfizer We expect a full recovery for the CEO of Pfizer We expect a full recovery for the CEO of Pfizer We expect a full recovery for the CEO of Pfizer We expect a full recovery for the CEO of Pfizer We expect a full recovery for the CEO of Pfizer We expect a full recovery for the CEO of Pfizer
Portfolio Composition