Big Data Analytics for Financial Sentiment Analysis in FinTech
(金融科技財務大數據情感分析)

Time: 2016/11/15 (Tue) (10:25-12:10)
Place: 朝陽科技大學資訊工程研究所 <教室：E520 演講廳>
Host: 吳世弘 教授 (Professor Shih-Hung Wu)

Min-Yuh Day
戴敏育
Assistant Professor
Dept. of Information Management, Tamkang University

http://mail.tku.edu.tw/myday/
2016-11-15
戴敏育博士
(Min-Yuh Day, Ph.D.)
淡江大學資管系專任助理教授
中央研究院資訊科學研究所訪問學人
國立台灣大學資訊管理博士

Publications Co-Chairs, IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013- )
Program Co-Chair, IEEE International Workshop on Empirical Methods for Recognizing Inference in TExt (IEEE EM-RITE 2012- )
Workshop Chair, The IEEE International Conference on Information Reuse and Integration (IEEE IRI)
Outline

• Big Data Analytics

• FinTech

• Financial Sentiment Analysis
Big Data Analytics
Big Data 4 V

Volume

40 ZETTABYTES
(43 TRILLION GIGABYTES)
of data will be created by 2020, an increase of 300 times from 2005

6 BILLION PEOPLE
have cell phones

WORLD POPULATION: 7 BILLION

2020

It’s estimated that
2.5 QUINTILLION BYTES
(23 TRILLION GIGABYTES)
of data are created each day

Most companies in the U.S. have at least
100 TERABYTES
(100,000 GIGABYTES)
of data stored

Velocity

The New York Stock Exchange
1 TB OF TRADE
INFORMATION
during each trading session

Modern cars have close to
100 SENSORS
that monitor items such as fuel level and tire pressure

By 2015

4.4 MILLION IT JOBS
will be created globally to support big data, with 1.9 million in the United States

Variety

DIFERENT FORMS OF DATA

As of 2011, the global size of data in healthcare was estimated to be
150 EXABYTES
(160 BILLION GIGABYTES)

By 2014, it’s anticipated there will be
420 MILLION WEARABLE, WIRELESS
HEALTH MONITORS

4 BILLION+ HOURS OF VIDEO
are watched on YouTube each month

30 BILLION PIECES OF CONTENT
are shared on Facebook every month

400 MILLION TWEETS
are sent per day by about 200 million monthly active users

By 2014, it’s anticipated there will be
27% OF RESPONDENTS
in one survey were unsure of how much of their data was inaccurate

Veracity

UNCERTAINTY
OF DATA

1 IN 3 BUSINESS LEADERS
don’t trust the information they use to make decisions

Poor data quality costs the US economy around
$3.1 TRILLION A YEAR

Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTEC, QAS

Source: https://www-01.ibm.com/software/data/bigdata/
Value
Big Data Technologies are Enabling a New Approach

Big Data Analytics and Data Mining
Stephan Kudyba (2014),
**Big Data, Mining, and Analytics: Components of Strategic Decision Making**, Auerbach Publications

Architecture of Big Data Analytics

Big Data Sources
- * Internal
- * External
- * Multiple formats
- * Multiple locations
- * Multiple applications

Big Data Transformation
- Raw Data
- Middleware
- Extract Transform Load
  - Data Warehouse
  - Traditional Format CSV, Tables

Big Data Platforms & Tools
- Hadoop
- MapReduce
- Pig
- Hive
- Jaql
- Zookeeper
- Hbase
- Cassandra
- Oozie
- Avro
- Mahout
- Others

Big Data Analytics Applications
- Queries
- Reports
- OLAP
- Data Mining

Source: Stephan Kudyba (2014), Big Data, Mining, and Analytics: Components of Strategic Decision Making, Auerbach Publications
Architecture of Big Data Analytics

Big Data Sources

- Internal
- External
- Multiple formats
- Multiple locations
- Multiple applications

Big Data Transformation

Big Data Platforms & Tools

Big Data Analytics Applications

- Queries
- Reports
- OLAP
- Data Mining

Data Mining
Big Data Analytics Applications

Source: Stephan Kudyba (2014), Big Data, Mining, and Analytics: Components of Strategic Decision Making, Auerbach Publications
Social Big Data Mining
(Hiroshi Ishikawa, 2015)

Architecture for Social Big Data Mining
(Hiroshi Ishikawa, 2015)

Enabling Technologies
- Integrated analysis model
- Natural Language Processing
- Information Extraction
- Anomaly Detection
- Discovery of relationships among heterogeneous data
- Large-scale visualization
- Parallel distrusted processing

Analysts
- Model Construction
- Explanation by Model
- Construction and confirmation of individual hypothesis
- Description and execution of application-specific task

Conceptual Layer
- Integrated analysis
- Construction and confirmation of individual hypothesis
- Description and execution of application-specific task

Logical Layer
- Multivariate analysis
- Application specific task
- Discovery of relationships among heterogeneous data
- Large-scale visualization
- Parallel distrusted processing

Software

Hardware

Social Data

Physical Layer

Source: Hiroshi Ishikawa (2015), Social Big Data Mining, CRC Press
Business Intelligence (BI) Infrastructure

Operational Data
Historical Data
Machine Data
Web Data
Audio/Video Data
External Data

Extract, transform, load

Data Mart

Data Warehouse

Hadoop Cluster

Analytic Platform

Casual users
- Queries
- Reports
- Dashboards

Power users
- Queries
- Reports
- OLAP
- Data mining

References

• Jiawei Han and Micheline Kamber (2011), Data Mining: Concepts and Techniques, Third Edition, Elsevier
• Jennifer Golbeck (2013), Analyzing the Social Web, Morgan Kaufmann
• Stephan Kudyba (2014), Big Data, Mining, and Analytics: Components of Strategic Decision Making, Auerbach Publications
• Hiroshi Ishikawa (2015), Social Big Data Mining, CRC Press
LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton.

"Deep learning."

Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics. Deep learning discovers intricate structure in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer. Deep convolutional nets have brought about breakthroughs in processing images, video, speech and audio, whereas recurrent nets have shone light on sequential data such as text and speech.

Machine-learning technology powers many aspects of modern society: from web searches to content filtering on social networks to recommendations on e-commerce websites, and it is increasingly present in consumer products such as cameras and smartphones. Machine-learning systems are used to identify objects in images, transcribe speech into text, match news items, posts or products with users’ interests, and select relevant results of search. Increasingly, these applications make use of a class of techniques called deep learning.

Conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, conventional machine-learning techniques were limited in their ability to process natural data in their raw form.
Sebastian Raschka (2015),
Python Machine Learning,
Packt Publishing

Sunila Gollapudi (2016),
Practical Machine Learning,
Packt Publishing

Machine Learning Models

- Deep Learning
- Association rules
- Decision tree
- Clustering
- Bayesian

- Kernel
- Ensemble
- Dimensionality reduction
- Regression Analysis
- Instance based

Source: Sunila Gollapudi (2016), Practical Machine Learning, Packt Publishing
What makes a data scientist?

The big data phenomenon trained a bright spotlight on those who perform deep information analysis and can combine quantitative and statistical modeling expertise with business acumen and a talent for finding hidden patterns. Here’s a closer look.

Data scientists rely on analytics, predictive models, statistical analysis and modeling, data mining, sentiment analysis and more to do their jobs. Cleaning raw data and building models is detailed work, and the right tools make the process much more efficient.

Good data scientists select and address the problems that have the most value to the organization. Armed with data and analytical results, they may present their informed conclusions and recommendations to technical and nontechnical stakeholders.

Deep Learning
Intelligence from Big Data

Source: https://www.vlab.org/events/deep-learning/
Big Data

Mobile Sensors
Social Media
Video Surveillance
Video Rendering
Smart Grids
Geophysical Exploration
Medical Imaging
Gene Sequencing

Source: EMC Education Services, Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, Wiley, 2015
Data Scientist: The Sexiest Job of the 21st Century

(Davenport & Patil, 2012)(HBR)
Data Scientist: The Sexiest Job of the 21st Century

Meet the people who can coax treasure out of messy, unstructured data.

by Thomas H. Davenport and D.J. Patil

When Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren’t seeking out connections with the people who were already on the site at the rate executives had expected. Something was apparently missing in the social experience. As one LinkedIn manager put it, “It was like arriving at a conference reception and realizing you don’t know anyone. So you just stand in the corner sipping your drink—and you probably leave early.”

Data Scientist Profile

- Technical
- Quantitative
- Curious and Creative
- Skeptical
- Communicative and Collaborative

Source: EMC Education Services, Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, Wiley, 2015
Key Roles for a Successful Analytics Project

Source: EMC Education Services, Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, Wiley, 2015
Key Outputs from a Successful Analytics Project

Source: EMC Education Services, Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, Wiley, 2015
Data Science vs. 
Big Data vs. Data Analytics

DATA IS GROWING FASTER THAN EVER BEFORE.

Each person-
1.7 megabytes
created

Data Science vs. Big Data vs. Data Analytics

**WHAT ARE THEY?**

**Data Science** is a field that comprises of everything that related to data cleansing, preparation, and analysis.

**Big Data** is something that can be used to analyze insights which can lead to better decision and strategic business moves.

**Data Analytics** involves automating insights into a certain dataset as well as supposes the usage of queries and data aggregation procedures.

What are they used?

Data Science algorithms are used in industries like:
- Internet searches
- Search
- Recommenders
- Digital advertisements

Big Data is used in industries like:
- Financial Services
- Retail
- Communication

Data Analytics is used in industries like:
- Healthcare
- Travel
- Gaming
- Energy Management

Data Science
What are the Skills Required?

DATA SCIENTIST
- In-depth knowledge in SAS and/or R
- Python coding
- Hadoop platform
- SQL database/coding
- Working with unstructured data

BIG DATA SPECIALIST
- Analytical skills
- Creativity
- Mathematics and
- Statistical skills
- Computer science
- Business skills

DATA ANALYST
- Programming skills
- Statistical skills
- Mathematics
- Machine learning skills
- Data wrangling skills
- Communication and Data Visualization skills
- Data Intuition

FinTech
FinTech
Financial Technology

FinTech

“providing financial services by making use of software and modern technology”

Source: https://www.fintechweekly.com/fintech-definition
Financial Revolution with Fintech

A financial services revolution
Consumer Trends

1. Simplification
2. Transparency
3. Analytics
4. Reduced Friction

Source: http://www.hedgethink.com/fintech/european-fintech-top-100/
FinTech: Financial Services Innovation

Source: http://www3.weforum.org/docs/WEF_The_future__of_financial_services.pdf
FinTech: Investment Management Market Provisioning
FinTech: Market Provisioning

Source: http://www3.weforum.org/docs/WEF_The_future__of_financial_services.pdf
FinTech: Investment Management

Source: http://www3.weforum.org/docs/WEF_The_future__of_financial_services.pdf
FinTech

- **功能**
  - 支付 (Payments)
  - 保險 (Insurance)
  - 存貸 (Deposit & Lending)
  - 稱資 (Capital Raising)
  - 投資管理 (Investment Management)
  - 市場資訊供應 (Market Provisioning)

- **創新項目**
  - 無現金世界 (Cashless World)
  - 新興支付 (Emerging Payment Rails)
  - 價值鏈裂解 (Insurance Disaggregation)
  - 保險串接裝置 (Connected Insurance)
  - 替代管道 (Alternative Lending)
  - 通路偏好移轉 (Shifting Customer Preferences)
  - 群眾募資 (Crowdfunding)
  - 賦權投資者 (Empowered Investors)
  - 流程外部化 (Process Externalisation)
  - 機器革命 (Smarter, Faster Machines)
  - 新興平台 (New Market Platforms)

Source: https://www.stockfeel.com.tw/2015年世界經濟論壇－未來的金融服務/
FinTech: Market Provisioning Smarter, Faster Machines

- 機器革命 Smarter, Faster Machines
- 機器易用數據 (Machine Accessible Data)、人工智慧 / 機器學習、大數據
- 新興平台 New Market Platforms
- 固定收益商品平台 ALGOMI、基金 / 組合型基金平台 NOVUS、私募 / 創投平台 BISON、未公發股權平台 LIQUITY、原物料商品與衍生性合約平台 ClauseMatch

图表来源：Fugle团队整理

Source: https://www.stockfeel.com.tw/2015年世界經濟論壇－未來的金融服務/
FinTech: Investment Management
FinTech for Financial Services

- Retail Banking
- Lending and Financing
- Payments and Transfers
- Wealth and Asset Management
- Markets and Exchanges
- Insurance
- Blockchain Transactions

Source: https://fintechweekly.com/fintech-companies
Fintech Companies

Source: https://fintechweekly.com/fintech-companies
Major Participants in the FinTech Ecosystem

FinTech Ecosystem Development Framework

1. Business environment/access to markets
   - Cost advantages
   - Degree of clustering & integration
   - Labor availability and know-how
   - Infrastructure quality & access

2. Government/regulatory support
   - Policy setting
   - Land ownership & development
   - Ease of doing business
   - Taxes & work permits

3. Access to capital
   - Governmental funding
   - Bank funding
   - PE/VC funding
   - Incubators / accelerators

4. Financial expertise
   - Funding strategies
   - "Know your customer" & regulatory requirements
   - Deals structuring
   - Due diligence

The FinTech Innovation Ecosystem

The U.S. FinTech landscape

**Entrepreneurs**
- New York is the fastest-growing FinTech ecosystem in the U.S.
- Talent feed from world’s biggest financial center
- New York is a lifestyle choice for talented young entrepreneurs

**Support structures**
- Tax credits for business R&D and patents
- Incubators & accelerators (e.g., Partnership Fund for New York City)

**Financiers**
- International banks
- Global and local PE shops
- Venture capital funds
- University funds

**Investment**
- U.S. received 83 percent of global FinTech investments in 2013
- The financial services industry globally spent more than US$485 billion on ICT in 2014

**Payment platforms**
- Crowd funding
- E-commerce
- Others

**Investment advisories**

**Customers**
- Business to business: high density of financial services firms seeking support for digitalization
- Business to consumer: widespread mobile & e-commerce usage, low bank client “stickiness”

## Fintech Startups

### USA Fintech Ecosystem

**Crowdfunding**

<table>
<thead>
<tr>
<th>STARTUP</th>
<th>DESCRIPTION</th>
<th>LINKS</th>
<th>STATUS</th>
<th>MONEY RAISED</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crowdera</strong></td>
<td>Community of start-ups &amp; investors who make fundraising efficient</td>
<td>cb</td>
<td>alive</td>
<td>$24.1M</td>
</tr>
<tr>
<td><strong>AngelList</strong></td>
<td>Online marketplace that links accredited investors with consumer product and retail companies.</td>
<td>cb</td>
<td>alive</td>
<td>$53M</td>
</tr>
<tr>
<td><strong>CircleUp</strong></td>
<td>Real estate crowdfunding platform</td>
<td>cb</td>
<td>alive</td>
<td></td>
</tr>
<tr>
<td><strong>Indiegogo</strong></td>
<td>Crowdfunding platform</td>
<td>cb</td>
<td>alive</td>
<td>$56.5M</td>
</tr>
<tr>
<td><strong>Kickstarter</strong></td>
<td>Crowdfunding platform</td>
<td>cb</td>
<td>alive</td>
<td>$10M</td>
</tr>
<tr>
<td><strong>Local lift</strong></td>
<td>Brings crowdfunding to your local area.</td>
<td>cb</td>
<td>alive</td>
<td>$160k</td>
</tr>
<tr>
<td><strong>Onevest (Rock The Post)</strong></td>
<td>Equity crowdfunding platform</td>
<td>cb</td>
<td>alive</td>
<td>$2M</td>
</tr>
<tr>
<td><strong>Quirky</strong></td>
<td>Community-led invention platform</td>
<td>cb</td>
<td>alive</td>
<td></td>
</tr>
</tbody>
</table>

Source: https://fintechstartups.co
Financial Technology (Fintech) Categories

1. Banking Infrastructure
2. Business Lending
3. Consumer and Commercial Banking
4. Consumer Lending
5. Consumer Payments
6. Crowdfunding
7. Equity Financing
8. Financial Research and Data
9. Financial Transaction Security
10. Institutional Investing
11. International Money Transfer
12. Payments Backend and Infrastructure
13. Personal Finance
14. Point of Sale Payments
15. Retail Investing
16. Small and Medium Business Tools

Source: http://www.venturescanner.com/financial-technology
FinTech Ecosystem (April 2015)

FinTech
1,072 Companies

Contact info@venturescanner.com to see all companies

Financial Technology (FinTech)

THE FINTECH ECOSYSTEM

<table>
<thead>
<tr>
<th>ROBO-ADVISORS &amp; PERSONAL FINANCE</th>
<th>BLOCKCHAIN &amp; BITCOIN</th>
<th>INSURTECHS</th>
</tr>
</thead>
</table>
| BILIGUARD | moneyfarm | nutmeg | Robinhood | Kraken | Coinbase | Slice | Shift. | SafeShare | Cuvva | | | Dr. Bought By Many | WeSavvy | Knip | Risk | Lemonade | Clover | Hoven | 2606 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |...
Financial Technology (FinTech)

THE FINTECH ECOSYSTEM

Payments & Transfers
- Dwolla
- Stripe
- PayPal
- Square
- Klarna
- Venmo
- Braintree
- iZettle
- Adyen
- Paydiant
- VISA Checkout
- Amex Express Checkout
- Retail Banking
  - Xoom
  - Remitly
  - Azimo
  - WorldRemit
  - TransferWise
  - flywire
- Samsung Pay
- Apple Pay
- Android Pay

Lending & Financing
- Lending Club
- Prosper
- OnDeck
- Funding Circle
- RateSetter
- Funding Circle
- AVANT
- Zopa
- Avant
- Patch of Land
- Orchard
- Lendio
- Fundrise
- Asset Avenue
- Lufax.com
- Credit
- Cardlytics
- Kabbage
- Kred
- Jitnika
- Bondora
- Octanet
- Capital One
- Capital One
- Atom
- Moven
- WeBank
- Beacon
- Upstart
- Earnest
- Even
- Zest Finance
- Biz2Credit
- Even
- CommonBond

Insurance
- Metromile
- Oscar
- Friendsurance
- Nutmeg
- MyDrive
- Bizinsure
- Bill Guard
- Mint

Financial Management
- LendingRobot
- Betterment
- Wealthfront
- Robinhood

Markets & Exchanges
- WeSwap
- Ripple
- Coinbase
- Bitstamp
- Kraken
- Bitfinex
- BTC

Fintech Landscape

Banking & Payment (256)
- Banking (54)
- Payments (59)
- Wallets (18)
- Mobile Payments (6)
- Peer-to-Peer Payments (5)
- Blockchain Payments (5)
- APIS/Connectivity (19)
- Platforms (23)

Investments (108)
- Financial Planning (3)
- Fundraising (9)
- Investment Advice (7)
- Insurance (2)
- Artificial Intelligence (2)

Financing (196)
- P2P Lending (49)
- Direct Lending (88)
- Real Estate Lending (15)
- Crowdfunding (44)

Insurance (82)
- Life Insurance (5)
- Health Insurance (2)
- Home Insurance (2)
- Travel Insurance (3)

Infrastructure & Enabling Technologies (226)
- API/Connectivity (26)
- Platforms (23)
- Crypto Currencies (23)
- Security (43)
- Compliance & Scoring (19)
- Business Tools (99)
- Blockchain (12)
- Data & Analytics (81)

FinTech Landscape Enabling Technologies

Data & Analytics

Source: https://www.vbprofiles.com/l/fintech
FinTech
Startups
Worldwide
Fintech Startups WorldWide

Source: https://startups.watch/fintech-ecosystems-worldwide/
Fintech Startups WorldWide

German FinTech Overview - Unbundling Banks
powered by www.paymentandbanking.com, March 2016

Source: https://startups.watch/fintech-ecosystems-worldwide/
Fintech Startups WorldWide

Source: https://startups.watch/fintech-ecosystems-worldwide/
Fintech Startups WorldWide

FinTech Map Switzerland

Juni 2016

Source: https://startups.watch/fintech-ecosystems-worldwide/
Fintech Startups WorldWide

Source: https://startups.watch/fintech-ecosystems-worldwide/
Fintech Startups WorldWide

Source: https://startups.watch/fintech-ecosystems-worldwide/
Fintech Startups Worldwide

Singapore FinTech Landscape

Source: https://startups.watch/fintech-ecosystems-worldwide/
Fintech Startups WorldWide

Source: https://startups.watch/fintech-ecosystems-worldwide/
Fintech Startups WorldWide

Source: https://startups.watch/fintech-ecosystems-worldwide/
FinTech Trends

More than 50% of respondents predict that by 2030, most payments will be cashless and non-paper based.

Respondents cited big data analytics and alternative payment forms as the two innovations that are likely to have the greatest impact on the financial services space in the next 3-5 years.

Source: http://mashable.com/2016/01/27/financial-tech-brandspeak/
FinTech

Big Data Analytics

Which innovation will have the greatest impact on the financial services space in the next 3-5 years?

- 27% Big data analytics
- 26% Alternative forms of payments/lending
- 19% Blockchain
- 17% IoT
- 9% AI
- 2% Other

Source: http://mashable.com/2016/01/27/financial-tech-brandspeak/
Blockchain Technology

The blockchain is a decentralized ledger of all transactions across a peer-to-peer network.

Blockchain Technology

How it works:

Someone requests a transaction.

The transaction is complete.

The requested transaction is broadcast to a P2P network consisting of computers, known as nodes.

The network of nodes validates the transaction and the user's status using known algorithms.

Validation

A verified transaction can involve cryptocurrency, contracts, records, or other information.

Once verified, the transaction is combined with other transactions to create a new block of data for the ledger.

Blockchain Technology

**Benefits**
- Increased transparency
- Accurate tracking
- Permanent ledger
- Cost reduction

**Unknowns**
- Complex technology
- Regulatory implications
- Implementation challenges
- Competing platforms

Blockchain Technology

Potential Applications

**Automotive**
Consumers could use the blockchain to manage fractional ownership in autonomous cars.

**Financial services**
Faster, cheaper settlements could shave billions of dollars from transaction costs while improving transparency.

Blockchain Technology

Potential Applications

**Voting**
Using a blockchain code, constituents could cast votes via smartphone, tablet or computer, resulting in immediately verifiable results.

**Healthcare**
Patients' encrypted health information could be shared with multiple providers without the risk of privacy breaches.

Financial Sentiment Analysis
Big Data Approach to Combining Internal and External Data

Ultra-Fast Text Analytics in Trading Strategies

Twitter stock prices affected by news, Source: econob

Source: Susanne Chishti and Janos Barberis,
Internet Evolution

Internet of People (IoP): Social Media
Internet of Things (IoT): Machine to Machine

Social Media

Source: http://hungrywolfmarketing.com/2013/09/09/what-are-your-social-marketing-goals/
Emotions

Love
Joy
Surprise
Anger
Sadness
Fear

Example of Opinion: review segment on iPhone

“I bought an iPhone a few days ago.
It was such a nice phone.
The touch screen was really cool.
The voice quality was clear too.
However, my mother was mad with me as I did not tell her before I bought it.
She also thought the phone was too expensive, and wanted me to return it to the shop. ...”
Example of Opinion: review segment on iPhone

“(1) I bought an **iPhone** a few days ago.

(2) It was such a **nice** phone.

(3) The **touch screen** was really **cool**.

(4) The **voice quality** was **clear** too.

(5) However, my mother was mad with me as I did not tell her before I bought it.

(6) She also thought the phone was too **expensive**, and wanted me to return it to the shop. ...”

How consumers think, feel, and act

Emotions

- Love
- Joy
- Surprise
- Anger
- Sadness
- Fear

Maslow’s Hierarchy of Needs

1. **Physiological Needs**
   - (food, water, shelter)

2. **Safety Needs**
   - (security, protection)

3. **Social Needs**
   - (sense of belonging, love)

4. **Esteem Needs**
   - (self-esteem, recognition, status)

5. **Self-actualization Needs**
   - (self-development and realization)

Maslow’s hierarchy of human needs
(Maslow, 1943)

Maslow’s Hierarchy of Needs

- **Physiological needs:**
  - food, water, warmth, rest

- **Safety needs:**
  - security, safety

- **Belongingness and love needs:**
  - intimate relationships, friends

- **Esteem needs:**
  - prestige and feeling of accomplishment

- **Self-actualization:**
  - achieving one’s full potential, including creative activities

Source: http://sixstoriesup.com/social-psyche-what-makes-us-go-social/
Social Media Hierarchy of Needs

Maslow’s Hierarchy of Needs

1. Physiological
   - air, sleep, food, hunger, thirst, warmth

2. Safety & Security
   - shelter, protection, safety & stability

3. Love & Belonging
   - love, affection, family, & relationships

4. Esteem
   - self-esteem, status, reputation

5. Self Actualization
   - self-actualization

6. Personal Fulfillment
   - building a solid image. Engaging in conversations. Voicing your expertise.

7. Optimization & Monetization
   - creating personal branding

8. Community Building

9. Structure
   - choosing the platforms you are comfortable with to voice your opinion. A human tendency to follow what the masses are using – the safe choice!

10. Existence (Presence)
    - create your social network profile. The need to exist and have a voice!

Source: http://2.bp.blogspot.com/_Rta1VZItiMk/TPavcanFtfI/AAAAAAAAACo/OBGnRL5arSU/s1600/social-media-heirarchy-of-needs1.jpg
Social Media Hierarchy of Needs

1. Physiological
   - Basic human needs
2. Safety
   - Employment
3. Love/Belonging
   - Friendship and family
4. Esteem
   - Respect of and by others
5. Self actualization
   - Creativity & sense-making
The Social Feedback Cycle
Consumer Behavior on Social Media

Marketer-Generated

Awareness
Consideration

User-Generated

Purchase
Use
Form Opinion
Talk

Source: Evans et al. (2010), Social Media Marketing: The Next Generation of Business Engagement
The New Customer Influence Path

Source: Evans et al. (2010), Social Media Marketing: The Next Generation of Business Engagement
Architectures of Sentiment Analytics
Bing Liu (2015),
Sentiment Analysis:
Mining Opinions, Sentiments, and Emotions,
Cambridge University Press

http://www.amazon.com/Sentiment-Analysis-Opinions-Sentiments-Emotions/dp/1107017890
Sentiment Analysis and Opinion Mining

- Computational study of opinions, sentiments, subjectivity, evaluations, attitudes, appraisal, affects, views, emotions, etc., expressed in text.
  - Reviews, blogs, discussions, news, comments, feedback, or any other documents

Research Area of Opinion Mining

• Many names and tasks with difference objective and models
  – Sentiment analysis
  – Opinion mining
  – Sentiment mining
  – Subjectivity analysis
  – Affect analysis
  – Emotion detection
  – Opinion spam detection

Sentiment Analysis

• Sentiment
  – A thought, view, or attitude, especially one based mainly on emotion instead of reason

• Sentiment Analysis
  – opinion mining
  – use of natural language processing (NLP) and computational techniques to automate the extraction or classification of sentiment from typically unstructured text
Applications of Sentiment Analysis

• Consumer information
  – Product reviews

• Marketing
  – Consumer attitudes
  – Trends

• Politics
  – Politicians want to know voters’ views
  – Voters want to know politicians’ stances and who else supports them

• Social
  – Find like-minded individuals or communities
Sentiment detection

• How to interpret features for sentiment detection?
  – Bag of words (IR)
  – Annotated lexicons (WordNet, SentiWordNet)
  – Syntactic patterns

• Which features to use?
  – Words (unigrams)
  – Phrases/n-grams
  – Sentences
Problem statement of Opinion Mining

- Two aspects of abstraction
  - Opinion definition
    - What is an opinion?
    - What is the structured definition of opinion?
  - Opinion summarization
    - Opinion are subjective
      - An opinion from a single person (unless a VIP) is often not sufficient for action
    - We need opinions from many people, and thus opinion summarization.

What is an opinion?

- **Id: Abc123 on 5-1-2008** “I bought an *iPhone* a few days ago. It is such a *nice phone*. The *touch screen* is really *cool*. The *voice quality* is *clear* too. It is much *better* than my old *Blackberry*, which was a *terrible phone* and so *difficult to type* with its *tiny keys*. However, *my mother was mad* with me as I did not tell her before I bought the phone. *She also thought the phone was too expensive*, …”

- One can look at this review/blog at the
  - Document level
    - Is this review + or -?
  - Sentence level
    - Is each sentence + or -?
  - Entity and feature/aspect level
Entity and aspect/feature level

• Id: **Abc123** on **5-1-2008** “I bought an *iPhone* a few days ago. It is such a *nice phone*. The *touch screen* is really *cool*. The *voice quality* is *clear* too. It is much *better* than my old *Blackberry*, which was a *terrible phone* and so *difficult to type* with its *tiny keys*. However, *my mother* was *mad* with me as I did not tell her before I bought the phone. *She also thought the phone* was too *expensive*, …”

• **What do we see?**
  – **Opinion targets**: entities and their features/aspects
  – **Sentiments**: positive and negative
  – **Opinion holders**: persons who hold the opinions
  – **Time**: when opinion are expressed
Two main types of opinions

• **Regular opinions**: Sentiment/Opinion expressions on some target entities
  
  – **Direct opinions**: sentiment expressions on one object:
    
    • “The touch screen is really cool.”
    
    • “The picture quality of this camera is great”
  
  – **Indirect opinions**: comparisons, relations expressing similarities or differences (objective or subjective) of more than one object
    
    • “phone X is cheaper than phone Y.” (objective)
    
    • “phone X is better than phone Y.” (subjective)

• **Comparative opinions**: comparisons of more than one entity.
  
  – “iPhone is better than Blackberry.”

Subjective and Objective

• Objective
  – An objective sentence expresses some factual information about the world.
  – “I returned the phone yesterday.”
  – Objective sentences can implicitly indicate opinions
    • “The earphone broke in two days.”

• Subjective
  – A subjective sentence expresses some personal feelings or beliefs.
  – “The voice on my phone was not so clear”
  – Not every subjective sentence contains an opinion
    • “I wanted a phone with good voice quality”

• ➔ Subjective analysis

# Sentiment Analysis vs. Subjectivity Analysis

<table>
<thead>
<tr>
<th>Sentiment Analysis</th>
<th>Subjectivity Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>Subjective</td>
</tr>
<tr>
<td>Negative</td>
<td>Objective</td>
</tr>
<tr>
<td>Neutral</td>
<td></td>
</tr>
</tbody>
</table>
A (regular) opinion

• Opinion (a restricted definition)
  – An opinion (regular opinion) is simply a positive or negative sentiment, view, attitude, emotion, or appraisal about an entity or an aspect of the entity from an opinion holder.

• Sentiment orientation of an opinion
  – Positive, negative, or neutral (no opinion)
  – Also called:
    • Opinion orientation
    • Semantic orientation
    • Sentiment polarity

Entity and aspect

• Definition of Entity:
  – An *entity* $e$ is a product, person, event, organization, or topic.
  – $e$ is represented as
    • A hierarchy of components, sub-components.
    • Each node represents a components and is associated with a set of attributes of the components

• An opinion can be expressed on any node or attribute of the node

• Aspects(features)
  – represent both components and attribute

Opinion Definition

• An opinion is a quintuple
  \((e_j, a_{jk}, s_{ijkl}, h_i, t_l)\)
  where
  – \(e_j\) is a target entity.
  – \(a_{jk}\) is an aspect/feature of the entity \(e_j\).
  – \(s_{ijkl}\) is the sentiment value of the opinion from the opinion holder on feature of entity at time.
    \(s_{ijkl}\) is +ve, -ve, or neu, or more granular ratings
  – \(h_i\) is an opinion holder.
  – \(t_l\) is the time when the opinion is expressed.

• \((e_j, a_{jk})\) is also called opinion target

Terminologies

• **Entity**: object
• **Aspect**: feature, attribute, facet
• **Opinion holder**: opinion source

• **Topic**: entity, aspect

• **Product features, political issues**

Subjectivity and Emotion

• Sentence subjectivity
  – An objective sentence presents some factual information, while a subjective sentence expresses some personal feelings, views, emotions, or beliefs.

• Emotion
  – Emotions are people’s subjective feelings and thoughts.

Classification Based on Supervised Learning

• Sentiment classification
  – Supervised learning Problem
  – Three classes
    • *Positive*
    • *Negative*
    • *Neutral*

Opinion words in Sentiment classification

• topic-based classification
  – topic-related words are important
    • e.g., politics, sciences, sports

• Sentiment classification
  – topic-related words are unimportant
  – opinion words (also called sentiment words)
    • that indicate positive or negative opinions are important,
      e.g., great, excellent, amazing, horrible, bad, worst

Features in Opinion Mining

• Terms and their frequency
  – TF-IDF

• Part of speech (POS)
  – Adjectives

• Opinion words and phrases
  – beautiful, wonderful, good, and amazing are positive opinion words
  – bad, poor, and terrible are negative opinion words.
  – opinion phrases and idioms, e.g., cost someone an arm and a leg

• Rules of opinions

• Negations

• Syntactic dependency

Sentiment Analysis Architecture

Sentiment Classification Based on Emoticons

Lexicon-Based Model

1. Preassembled Word Lists
2. Generic Word Lists
3. Merged Lexicon
4. Tokenized Document Collection
5. Sentiment Scoring and Classification: Polarity
6. Sentiment Polarity

Sentiment Analysis Tasks

- Opinionated Document
- Subjectivity Classification
- Sentiment Classification
- Opinion holder extraction
- Object/Feature extraction
## Sentiment Analysis vs. Subjectivity Analysis

<table>
<thead>
<tr>
<th>Sentiment Analysis</th>
<th>Subjectivity Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>Subjective</td>
</tr>
<tr>
<td>Negative</td>
<td>Objective</td>
</tr>
<tr>
<td>Neutral</td>
<td></td>
</tr>
</tbody>
</table>
Levels of Sentiment Analysis

- Word level Sentiment Analysis
- Sentence level Sentiment Analysis
- Document level Sentiment Analysis
- Feature level Sentiment Analysis
Sentiment Analysis

Sentiment Classification Techniques

- Sentiment Analysis
  - Machine Learning Approach
    - Supervised Learning
      - Decision Tree Classifiers
      - Linear Classifiers
      - Rule-based Classifiers
      - Probabilistic Classifiers
    - Unsupervised Learning
      - Support Vector Machine (SVM)
      - Neural Network (NN)
      - Deep Learning (DL)
      - Naïve Bayes (NB)
      - Bayesian Network (BN)
      - Maximum Entropy (ME)
  - Lexicon-based Approach
    - Dictionary-based Approach
      - Statistical
    - Corpus-based Approach
      - Semantic
  - Corpus-based Approach

# A Brief Summary of Sentiment Analysis Methods

<table>
<thead>
<tr>
<th>Study</th>
<th>Analysis Task</th>
<th>Sentiment Identification Method</th>
<th>Level</th>
<th>Sentiment Aggregation Method</th>
<th>Level</th>
<th>Nature of Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hu and Li, 2011</td>
<td>Polarity</td>
<td>ML (Probabilistic model)</td>
<td>Snippet</td>
<td>Sum</td>
<td>Snippet</td>
<td>Valence</td>
</tr>
<tr>
<td>Li and Wu, 2010</td>
<td>Polarity</td>
<td>Lexicon/Rule</td>
<td>Phrase</td>
<td>Sum</td>
<td>Snippet</td>
<td>Valence</td>
</tr>
<tr>
<td>Thelwall et al., 2010</td>
<td>Polarity</td>
<td>Lexicon/Rule</td>
<td>Sentence</td>
<td>Max &amp; Min</td>
<td>Snippet</td>
<td>Valence</td>
</tr>
<tr>
<td>Boiy and Moens, 2009</td>
<td>Both</td>
<td>ML (Cascade ensemble)</td>
<td>Sentence</td>
<td></td>
<td></td>
<td>Valence</td>
</tr>
<tr>
<td>Chung 2009</td>
<td>Polarity</td>
<td>Lexicon</td>
<td>Phrase</td>
<td>Average</td>
<td>Sentence</td>
<td>Valence</td>
</tr>
<tr>
<td>Wilson, Wiebe, and Hoffmann, 2009</td>
<td>Both</td>
<td>ML (SVM, AdaBoost, Rule, etc.)</td>
<td>Phrase</td>
<td></td>
<td></td>
<td>Valence</td>
</tr>
<tr>
<td>Zhang et al., 2009</td>
<td>Polarity</td>
<td>Lexicon/Rule</td>
<td>Sentence</td>
<td>Weighted average</td>
<td>Snippet</td>
<td>Valence</td>
</tr>
<tr>
<td>Abbasi, Chen, and Salem, 2008</td>
<td>Polarity</td>
<td>ML (GA + feature selection)</td>
<td>Snippet</td>
<td></td>
<td></td>
<td>Valence</td>
</tr>
<tr>
<td>Subrahmanian and Reforgiato, 2008</td>
<td>Polarity</td>
<td>Lexicon/Rule</td>
<td>Phrase</td>
<td>Rule</td>
<td>Snippet</td>
<td>Valence</td>
</tr>
<tr>
<td>Tan and Zhang 2008</td>
<td>Polarity</td>
<td>ML (SVM, Winnow, NB, etc.)</td>
<td>Snippet</td>
<td></td>
<td></td>
<td>Valence</td>
</tr>
<tr>
<td>Airoldi, Bai, and Padman, 2007</td>
<td>Polarity</td>
<td>ML (Markov Blanket)</td>
<td>Snippet</td>
<td></td>
<td></td>
<td>Valence</td>
</tr>
<tr>
<td>Das and Chen, 2007</td>
<td>Polarity</td>
<td>ML (Bayesian, Discriminate, etc.)</td>
<td>Snippet</td>
<td>Average</td>
<td>Daily</td>
<td>Valence</td>
</tr>
<tr>
<td>Liu et al., 2007</td>
<td>Polarity</td>
<td>ML (PLSA)</td>
<td>Snippet</td>
<td></td>
<td></td>
<td>Valence</td>
</tr>
<tr>
<td>Kennedy and Inkpen, 2006</td>
<td>Polarity</td>
<td>Lexicon/Rule, ML (SVM)</td>
<td>Phrase</td>
<td>Count</td>
<td>Snippet</td>
<td>Valence</td>
</tr>
<tr>
<td>Mishne 2006</td>
<td>Polarity</td>
<td>Lexicon</td>
<td>Phrase</td>
<td>Average</td>
<td>Snippet</td>
<td>Valence</td>
</tr>
<tr>
<td>Liu et al., 2005</td>
<td>Polarity</td>
<td>Lexicon/Rule</td>
<td>Phrase</td>
<td>Distribution</td>
<td>Object</td>
<td>Range</td>
</tr>
<tr>
<td>Mishne 2005</td>
<td>Polarity</td>
<td>ML (SVM)</td>
<td>Snippet</td>
<td></td>
<td></td>
<td>Valence</td>
</tr>
<tr>
<td>Popescu and Etzioni 2005</td>
<td>Polarity</td>
<td>Lexicon/Rule</td>
<td>Phrase</td>
<td></td>
<td></td>
<td>Valence</td>
</tr>
<tr>
<td>Efron 2004</td>
<td>Polarity</td>
<td>ML (SVM, NB)</td>
<td>Snippet</td>
<td></td>
<td></td>
<td>Valence</td>
</tr>
<tr>
<td>Wilson, Wiebe, and Hwa, 2004</td>
<td>Both</td>
<td>ML (SVM, AdaBoost, Rule, etc.)</td>
<td>Sentence</td>
<td></td>
<td></td>
<td>Valence</td>
</tr>
<tr>
<td>Nigam and Hurst 2004</td>
<td>Polarity</td>
<td>Lexicon/Rule</td>
<td>Chunk</td>
<td>Rule</td>
<td>Sentence</td>
<td>Valence</td>
</tr>
<tr>
<td>Dave, Lawrence, and Pennock, 2003</td>
<td>Polarity</td>
<td>ML (SVM, Rainbow, etc.)</td>
<td>Snippet</td>
<td></td>
<td></td>
<td>Valence</td>
</tr>
<tr>
<td>Nasukawa and Yi 2003</td>
<td>Polarity</td>
<td>Lexicon/Rule</td>
<td>Phrase</td>
<td>Rule</td>
<td>Sentence</td>
<td>Valence</td>
</tr>
<tr>
<td>Yi et al., 2003</td>
<td>Polarity</td>
<td>Lexicon/Rule</td>
<td>Phrase</td>
<td>Rule</td>
<td>Sentence</td>
<td>Valence</td>
</tr>
<tr>
<td>Yu and Hatzivassiloglou 2003</td>
<td>Both</td>
<td>ML (NB) + Lexicon/Rule</td>
<td>Phrase</td>
<td>Average</td>
<td>Sentence</td>
<td>Valence</td>
</tr>
<tr>
<td>Pang, Lee, and Vaithyanathan 2002</td>
<td>Polarity</td>
<td>ML (SVM, MaxEnt, NB)</td>
<td>Snippet</td>
<td></td>
<td></td>
<td>Valence</td>
</tr>
<tr>
<td>Subasic and Huettner 2001</td>
<td>Polarity</td>
<td>Lexicon/Fuzzy logic</td>
<td>Phrase</td>
<td>Average</td>
<td>Snippet</td>
<td>Valence</td>
</tr>
<tr>
<td>Turney 2001</td>
<td>Polarity</td>
<td>Lexicon/Rule</td>
<td>Phrase</td>
<td>Average</td>
<td>Snippet</td>
<td>Valence</td>
</tr>
</tbody>
</table>

(Both = Subjectivity and Polarity; ML = Machine Learning; Lexicon/Rule = Lexicon enhanced by linguistic rules)

Word-of-Mouth (WOM)

• “This book is the best written documentary thus far, yet sadly, there is no soft cover edition.”

• “This book is the best written documentary thus far, yet sadly, there is no soft cover edition.”

This book is the best written documentary thus far, yet sadly, there is no soft cover edition.

<table>
<thead>
<tr>
<th>Word</th>
<th>POS</th>
</tr>
</thead>
<tbody>
<tr>
<td>This</td>
<td>DT</td>
</tr>
<tr>
<td>book</td>
<td>NN</td>
</tr>
<tr>
<td>is</td>
<td>VBZ</td>
</tr>
<tr>
<td>the</td>
<td>DT</td>
</tr>
<tr>
<td>best</td>
<td>JJS</td>
</tr>
<tr>
<td>written</td>
<td>VBN</td>
</tr>
<tr>
<td>documentary</td>
<td>NN</td>
</tr>
<tr>
<td>thus</td>
<td>RB</td>
</tr>
<tr>
<td>far</td>
<td>RB</td>
</tr>
<tr>
<td>,</td>
<td></td>
</tr>
<tr>
<td>yet</td>
<td>RB</td>
</tr>
<tr>
<td>sadly</td>
<td>RB</td>
</tr>
<tr>
<td>,</td>
<td></td>
</tr>
<tr>
<td>there</td>
<td>EX</td>
</tr>
<tr>
<td>is</td>
<td>VBZ</td>
</tr>
<tr>
<td>no</td>
<td>DT</td>
</tr>
<tr>
<td>soft</td>
<td>JJ</td>
</tr>
<tr>
<td>cover</td>
<td>NN</td>
</tr>
<tr>
<td>edition</td>
<td>NN</td>
</tr>
</tbody>
</table>

## Conversion of text representation

### Word Vector (WV)

<table>
<thead>
<tr>
<th>Word</th>
<th>Vector Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>This</td>
<td></td>
</tr>
<tr>
<td>book</td>
<td></td>
</tr>
<tr>
<td>is</td>
<td></td>
</tr>
<tr>
<td>the</td>
<td></td>
</tr>
<tr>
<td>best</td>
<td></td>
</tr>
<tr>
<td>written</td>
<td></td>
</tr>
<tr>
<td>documentary</td>
<td></td>
</tr>
<tr>
<td>thus</td>
<td></td>
</tr>
<tr>
<td>far</td>
<td></td>
</tr>
<tr>
<td>,</td>
<td></td>
</tr>
<tr>
<td>yet</td>
<td></td>
</tr>
<tr>
<td>sadly</td>
<td></td>
</tr>
<tr>
<td>,</td>
<td></td>
</tr>
<tr>
<td>there</td>
<td></td>
</tr>
<tr>
<td>is</td>
<td></td>
</tr>
<tr>
<td>no</td>
<td></td>
</tr>
<tr>
<td>soft</td>
<td></td>
</tr>
<tr>
<td>cover</td>
<td></td>
</tr>
<tr>
<td>edition</td>
<td></td>
</tr>
</tbody>
</table>

### Polarity Score Vector (PSV)

<table>
<thead>
<tr>
<th>pscore</th>
<th>nscore</th>
<th>Polarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>Neutral (0)</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>Neutral (0)</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>Neutral (0)</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>Neutral (0)</td>
</tr>
<tr>
<td>0.75</td>
<td>0</td>
<td>Positive (0.75)</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>Neutral (0)</td>
</tr>
<tr>
<td>0.375</td>
<td>0</td>
<td>Neutral (0)</td>
</tr>
<tr>
<td>0.375</td>
<td>0</td>
<td>Positive (0.375)</td>
</tr>
<tr>
<td>0.125</td>
<td>0</td>
<td>Neutral (0)</td>
</tr>
<tr>
<td>0.25</td>
<td>0.5</td>
<td>Negative (0.25)</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>Neutral (0)</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>Neutral (0)</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>Neutral (0)</td>
</tr>
<tr>
<td>0.75</td>
<td>0</td>
<td>Negative (0.75)</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>Neutral (0)</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>Neutral (0)</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>Neutral (0)</td>
</tr>
</tbody>
</table>

### Microstate Sequence (MS)

<table>
<thead>
<tr>
<th>Microstate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>-1</td>
</tr>
<tr>
<td>-1</td>
</tr>
</tbody>
</table>

### Probability Distribution (P)

- $P(1) = 3/17$
- $P(-1) = 3/17$
- $P(0) = 11/17$

---

## Example of SentiWordNet

<table>
<thead>
<tr>
<th>POS</th>
<th>ID</th>
<th>PosScore</th>
<th>NegScore</th>
<th>SynsetTerms</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>00217728</td>
<td>0.75</td>
<td>0</td>
<td>beautiful#1</td>
<td>delighting the senses or exciting intellectual or emotional admiration; &quot;a beautiful child&quot;; &quot;beautiful country&quot;; &quot;a beautiful painting&quot;; &quot;a beautiful theory&quot;; &quot;a beautiful party“</td>
</tr>
<tr>
<td>a</td>
<td>00227507</td>
<td>0.75</td>
<td>0</td>
<td>best#1</td>
<td>(superlative of <code>good</code>) having the most positive qualities; &quot;the best film of the year&quot;; &quot;the best solution&quot;; &quot;the best time for planting&quot;; &quot;wore his best suit“</td>
</tr>
<tr>
<td>r</td>
<td>00042614</td>
<td>0</td>
<td>0.625</td>
<td>unhappily#2 sadly#1</td>
<td>in an unfortunate way; &quot;sadly he died before he could see his grandchild“</td>
</tr>
<tr>
<td>r</td>
<td>00093270</td>
<td>0</td>
<td>0.875</td>
<td>woefully#1 sadly#3 lamentably#1 deplorably#1</td>
<td>in an unfortunate or deplorable manner; &quot;he was sadly neglected&quot;; &quot;it was woefully inadequate“</td>
</tr>
<tr>
<td>r</td>
<td>00404501</td>
<td>0</td>
<td>0.25</td>
<td>sadly#2</td>
<td>with sadness; in a sad manner; &quot;`She died last night,' he said sadly&quot;</td>
</tr>
</tbody>
</table>
The car is very old but it is rather not expensive.

The car is very old but it is rather not expensive.

The car is very old but it is rather not expensive.

Polarity Detection with SenticNet

The car is very old but it is rather not expensive.
The car is very old but it is rather not expensive.

Polarity Detection with SenticNet

Polarity Detection with SenticNet

Polarity Detection with SenticNet

Polarity Detection with SenticNet

Evaluation of Text Mining and Sentiment Analysis

• Evaluation of Information Retrieval
• Evaluation of Classification Model (Prediction)
  — Accuracy
  — Precision
  — Recall
  — F-score
Deep Learning for Sentiment Analytics
Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank

Richard Socher, Alex Perelygin, Jean Y. Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng and Christopher Potts
Stanford University, Stanford, CA 94305, USA
richard@socher.org, {aperelygin, jchuang, ang}@cs.stanford.edu
{jeaneis, manning, cgpotts}@stanford.edu

Abstract

Semantic word spaces have been very useful but cannot express the meaning of longer phrases in a principled way. Further progress towards understanding compositionality in tasks such as sentiment detection requires richer supervised training and evaluation resources and more powerful models of composition. To remedy this, we introduce a Sentiment Treebank. It includes fine grained sentiment labels for 215,154 phrases in the parse trees of 11,855 sentences and presents new challenges for sentiment compositionality. To address them, we introduce the Recursive Neural Tensor Network. When trained on the new treebank, this model out-

Figure 1: Example of the Recursive Neural Tensor Network accurately predicting 5 sentiment classes, very negative to very positive (−−, −, 0, +, + +), at every node of a parse tree and capturing the negation and its scope in this sentence.

Recursive Neural Tensor Network (RNTN)

Recursive Neural Network (RNN) models for sentiment

Recursive Neural Tensor Network (RNTN)

Roger Dodger is one of the most compelling variations on this theme.

Roger Dodger is one of the least compelling variations on this theme.

Roger Dodger is one of the most compelling variations on this theme.

Roger Dodger is one of the least compelling variations on this theme.

Accuracy for fine grained (5-class) and binary predictions at the sentence level (root) and for all nodes

<table>
<thead>
<tr>
<th>Model</th>
<th>Fine-grained</th>
<th>Positive/Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>Root</td>
</tr>
<tr>
<td>NB</td>
<td>67.2</td>
<td>41.0</td>
</tr>
<tr>
<td>SVM</td>
<td>64.3</td>
<td>40.7</td>
</tr>
<tr>
<td>BiNB</td>
<td>71.0</td>
<td>41.9</td>
</tr>
<tr>
<td>VecAvg</td>
<td>73.3</td>
<td>32.7</td>
</tr>
<tr>
<td>RNN</td>
<td>79.0</td>
<td>43.2</td>
</tr>
<tr>
<td>MV-RNN</td>
<td>78.7</td>
<td>44.4</td>
</tr>
<tr>
<td>RNTN</td>
<td>80.7</td>
<td>45.7</td>
</tr>
</tbody>
</table>

Accuracy of negation detection

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Negated Positive</td>
<td>Negated Negative</td>
</tr>
<tr>
<td>biNB</td>
<td>19.0</td>
<td>27.3</td>
</tr>
<tr>
<td>RNN</td>
<td>33.3</td>
<td>45.5</td>
</tr>
<tr>
<td>MV-RNN</td>
<td>52.4</td>
<td>54.6</td>
</tr>
<tr>
<td>RNTN</td>
<td>71.4</td>
<td>81.8</td>
</tr>
</tbody>
</table>

Long Short-Term Memory (LSTM)

Source: https://cs224d.stanford.edu/reports/HongJames.pdf
# Deep Learning for Sentiment Analysis

## CNN RNTN LSTM

<table>
<thead>
<tr>
<th>Model</th>
<th>Fine (5-class)</th>
<th>Binary</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCNN (Blunsom, et al. 2014)</td>
<td>0.485</td>
<td>0.868</td>
</tr>
<tr>
<td>RNTN (Socher, et al. 2013)</td>
<td>0.457</td>
<td>0.854</td>
</tr>
<tr>
<td>CNN-non-static (Kim, 2014)</td>
<td>0.480</td>
<td>0.872</td>
</tr>
<tr>
<td>CNN-multi-channel (Kim, 2014)</td>
<td>0.474</td>
<td>0.881</td>
</tr>
<tr>
<td>DRNN w. pretrained word-embeddings (Irsoy and Cardie, 2014)</td>
<td>0.498</td>
<td>0.866</td>
</tr>
<tr>
<td>Paragraph Vector (Le and Mikolov, 2014)</td>
<td>0.487</td>
<td>0.878</td>
</tr>
<tr>
<td>Dependency Tree-LSTM (Tai, et al, 2015)</td>
<td>0.484</td>
<td>0.857</td>
</tr>
<tr>
<td>Constituency Tree-LSTM (Tai, et al, 2015)</td>
<td>0.439</td>
<td>0.820</td>
</tr>
<tr>
<td>Constituency Tree-LSTM (Glove vectors) (Tai, et al, 2015)</td>
<td>0.510</td>
<td>0.880</td>
</tr>
<tr>
<td>Paragraph Vector</td>
<td>0.391</td>
<td>0.798</td>
</tr>
<tr>
<td>LSTM</td>
<td>0.456</td>
<td>0.843</td>
</tr>
<tr>
<td>Deep Recursive-NN</td>
<td>0.469</td>
<td>0.847</td>
</tr>
</tbody>
</table>

Source: [https://cs224d.stanford.edu/reports/HongJames.pdf](https://cs224d.stanford.edu/reports/HongJames.pdf)
## Performance Comparison of Sentiment Analysis Methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Data Set</th>
<th>Acc.</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>Machine Learning</td>
<td>SVM</td>
<td>86.40%</td>
<td>Pang, Lee[23]</td>
</tr>
<tr>
<td></td>
<td>CoTraining SVM</td>
<td>82.52%</td>
<td>Liu[14]</td>
</tr>
<tr>
<td></td>
<td>Deep learning</td>
<td>80.70%</td>
<td>Richard[18]</td>
</tr>
<tr>
<td>Lexical based</td>
<td>Corpus</td>
<td>74.00%</td>
<td>Turkey</td>
</tr>
<tr>
<td></td>
<td>Dictionary</td>
<td>---</td>
<td>Taboada[20]</td>
</tr>
<tr>
<td>Cross-lingual</td>
<td>Ensemble</td>
<td>81.00%</td>
<td>Wan.X[16]</td>
</tr>
<tr>
<td></td>
<td>Co-Train</td>
<td>81.30%</td>
<td>Wan.X.[16]</td>
</tr>
<tr>
<td></td>
<td>EWGA</td>
<td>&gt;90%</td>
<td>Abbasi,A.</td>
</tr>
<tr>
<td></td>
<td>CLMM</td>
<td>83.02%</td>
<td>Mengi</td>
</tr>
<tr>
<td>Cross-domain</td>
<td>Active Learning</td>
<td>80% (avg)</td>
<td>Li, S</td>
</tr>
<tr>
<td></td>
<td>Thesaurus</td>
<td></td>
<td>Bollegala[22]</td>
</tr>
<tr>
<td></td>
<td>SFA</td>
<td></td>
<td>Pan S J[15]</td>
</tr>
<tr>
<td></td>
<td>Book, DVD, Electronics,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kitchen</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Resources of Opinion Mining
Datasets of Opinion Mining

- Blog06
  - 25GB TREC test collection
  - [http://ir.dcs.gla.ac.uk/test collections/access to data.html](http://ir.dcs.gla.ac.uk/test collections/access to data.html)
- Cornell movie-review datasets
- Customer review datasets
- Multiple-aspect restaurant reviews
  - [http://people.csail.mit.edu/bsnyder/naacl07](http://people.csail.mit.edu/bsnyder/naacl07)
- NTCIR multilingual corpus
  - NTCIR Multilingual Opinion-Analysis Task (MOAT)

Lexical Resources of Opinion Mining

- SentiWordnet
  - http://sentiwordnet.isti.cnr.it/
- General Inquirer
  - http://www.wjh.harvard.edu/~inquirer/
- OpinionFinder’s Subjectivity Lexicon
  - http://www.cs.pitt.edu/mpqa/
- NTU Sentiment Dictionary (NTUSD)
  - http://nlg18.csie.ntu.edu.tw:8080/opinion/
- Hownet Sentiment
### Example of SentiWordNet

<table>
<thead>
<tr>
<th>POS</th>
<th>ID</th>
<th>PosScore</th>
<th>NegScore</th>
<th>SynsetTerms</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>00217728</td>
<td>0.75</td>
<td>0</td>
<td>beautiful#1</td>
<td>delighting the senses or exciting intellectual or emotional admiration; &quot;a beautiful child&quot;; &quot;beautiful country&quot;; &quot;a beautiful painting&quot;; &quot;a beautiful theory&quot;; &quot;a beautiful party&quot;</td>
</tr>
<tr>
<td>a</td>
<td>00227507</td>
<td>0.75</td>
<td>0</td>
<td>best#1</td>
<td>(superlative of `good') having the most positive qualities; &quot;the best film of the year&quot;; &quot;the best solution&quot;; &quot;the best time for planting&quot;; &quot;wore his best suit&quot;</td>
</tr>
<tr>
<td>r</td>
<td>00042614</td>
<td>0</td>
<td>0.625</td>
<td>unhappily#2 sadly#1</td>
<td>in an unfortunate way; &quot;sadly he died before he could see his grandchild&quot;</td>
</tr>
<tr>
<td>r</td>
<td>00093270</td>
<td>0</td>
<td>0.875</td>
<td>woefully#1 sadly#3 lamentably#1 deplorably#1</td>
<td>in an unfortunate or deplorable manner; &quot;he was sadly neglected&quot;; &quot;it was woefully inadequate&quot;</td>
</tr>
<tr>
<td>r</td>
<td>00404501</td>
<td>0</td>
<td>0.25</td>
<td>sadly#2</td>
<td>with sadness; in a sad manner; &quot;`She died last night,' he said sadly&quot;</td>
</tr>
</tbody>
</table>
《知網》情感分析用詞語集（beta版）

- “中英文情感分析用詞語集”
  - 包含詞語約 17887
- “中文情感分析用詞語集”
  - 包含詞語約 9193
- “英文情感分析用詞語集”
  - 包含詞語 8945

中文情感分析用詞語集

<table>
<thead>
<tr>
<th>詞語類型</th>
<th>數量</th>
</tr>
</thead>
<tbody>
<tr>
<td>中文正面情感詞語</td>
<td>836</td>
</tr>
<tr>
<td>中文負面情感詞語</td>
<td>1254</td>
</tr>
<tr>
<td>中文正面評價詞語</td>
<td>3730</td>
</tr>
<tr>
<td>中文負面評價詞語</td>
<td>3116</td>
</tr>
<tr>
<td>中文程度級別詞語</td>
<td>219</td>
</tr>
<tr>
<td>中文主張詞語</td>
<td>38</td>
</tr>
<tr>
<td>Total</td>
<td>9193</td>
</tr>
</tbody>
</table>
中文情感分析用詞語集

• “正面情感” 詞語
  - 如：
    愛，讚賞，快樂，感同身受，好奇，喝彩，魂牽夢縈，嘉許 ...

• “負面情感” 詞語
  - 如：
    哀傷，半信半疑，鄙視，不滿意，不是滋味兒，後悔，大失所望 ...

中文情感分析用詞語集

• “正面評價”詞語
  - 如：
    不可或缺，部優，才高八斗，沉魚落雁，
    催人奮進，動聽，對勁兒 ...

• “負面評價”詞語
  - 如：
    醜，苦，超標，華而不實，荒涼，混濁，
    畸輕畸重，價高，空洞無物 ...

中文情感分析用詞語集

“程度級別”詞語

1. “極其 | extreme / 最 | most”
   • 非常，極，極度，無以倫比，最為
2. “很 | very”
   • 多麼，分外，格外，著實

“主張”詞語

1. {perception | 感知}
   • 感覺，覺得，預感
2. {regard | 認為}
   • 認為，以為，主張

References

  http://www.cs.uic.edu/~liub/WebMiningBook.html
• Bing Liu (2013), Opinion Spam Detection: Detecting Fake Reviews and Reviewers,  
  http://www.cs.uic.edu/~liub/FBS/fake-reviews.html
• Wiltrud Kessler (2012), Introduction to Sentiment Analysis,  
  http://www.ims.uni-stuttgart.de/~kesslewd/lehre/sentimentanalysis12s/introduction_sentimentanalysis.pdf
• Guandong Xu, Yanchun Zhang, Lin Li (2011), Web Mining and Social Networking: Techniques and Applications, 2011, Springer
References

• Cambria, Erik, Soujanya Poria, Rajiv Bajpai, and Björn Schuller. "SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives." In the 26th International Conference on Computational Linguistics (COLING), Osaka. 2016.


• Steven Struhl (2015), Practical Text Analytics: Interpreting Text and Unstructured Data for Business Intelligence (Marketing Science), Kogan Page

• Bing Liu (2015), Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Cambridge University Press
Deep Learning for Financial Sentiment Analysis on Finance News Providers

Min-Yuh Day

Chia-Chou Lee

Department of Information Management
Tamkang University, Taiwan

myday@mail.tku.edu.tw

The 7th International Workshop on Mining and Analyzing Social Networks for Decision Support (MSNDS 2016)
Outline

• Introduction
• System Architecture
• Experimental Results and Discussion
• Conclusion
Introduction
Financial Sentiment Analysis

or  ?
Motivation

• Investors have always been interested in stock price forecasting.

• The rapid development of electronic media, the big data of financial news are released on different media every day.
Research Gap

• Few research involved the discussion on whether using different media could affect forecasting results.

• Financial sentiment analysis is an important research area of financial technology (FinTech).
Highlights

• This research focuses on investigating the influence of using different financial resources to investment and how to improve the accuracy of forecasting through deep learning.

• The experimental result shows various financial resources have significantly different effects to investors and their investments, while the accuracy of news categorization could be improved through deep learning.
The relationship between Events, News and Markets (price) through Information.
Financial Sentiment Analysis

- Machine Learning Approach
- Lexicon-based Approach
System Architecture
System Architecture

Step 1: Internet
- Web Crawler
  - HTML Document

Step 2: HTML Parser
- Corpus

Step 3: Sentiment Word Expansion
- NTUSD
- HowNet-VSA
- NTUFSD
- iMFinanceSD

Integrate SD

Step 4: Text Segment
- Future Design
- Feature Selection

Step 5: News Classification Tagger

Step 6: Training Data
- Deep Learning

Step 7: Testing Data

Step 8: Result & Evaluation

Tick Data Processing Module

Step 1: Tick Data
- Stock Price Calculator
## Feature used for financial sentiment analysis

<table>
<thead>
<tr>
<th>ID</th>
<th>Future Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>F01</td>
<td>NewsCharacters</td>
<td>Total word number of news text</td>
</tr>
<tr>
<td>F02</td>
<td>NewsTokens</td>
<td>Number of news words</td>
</tr>
<tr>
<td>F03</td>
<td>NTUSD_Positive</td>
<td>NTUSD positive word</td>
</tr>
<tr>
<td>F04</td>
<td>NTUSD&gt;Negative</td>
<td>NTUSD negative word</td>
</tr>
<tr>
<td>F05</td>
<td>NTUSD_PNDiff</td>
<td>NTUSD difference of positive and negative word</td>
</tr>
<tr>
<td>F06</td>
<td>HowNet_Positive</td>
<td>HowNet positive sentiment word</td>
</tr>
<tr>
<td>F07</td>
<td>HowNet_Negative</td>
<td>HowNet negative sentiment word</td>
</tr>
<tr>
<td>F08</td>
<td>HowNet_PNDiff</td>
<td>HowNet difference of positive and negative word</td>
</tr>
<tr>
<td>F09</td>
<td>FinanceSD_Positive</td>
<td>NTUFSD+iMFinanceSD positive word</td>
</tr>
<tr>
<td>F10</td>
<td>FinanceSD&gt;Negative</td>
<td>NTUFSD+iMFinanceSD negative word</td>
</tr>
<tr>
<td>F11</td>
<td>FinanceSD_PNDiff</td>
<td>NTUFSD+iMFinanceSD difference of positive and negative word</td>
</tr>
</tbody>
</table>
## Example of feature used for Financial Sentiment Analysis

<table>
<thead>
<tr>
<th>Code</th>
<th>Value1</th>
<th>Value2</th>
<th>Value3</th>
<th>Value4</th>
<th>Value5</th>
<th>Value6</th>
<th>Value7</th>
<th>Value8</th>
<th>Value9</th>
<th>Value10</th>
<th>Value11</th>
<th>Value12</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>D00001</td>
<td>516</td>
<td>185</td>
<td>3</td>
<td>8</td>
<td>-5</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td>14</td>
<td>8</td>
<td>6</td>
<td>3008</td>
<td>2013-01-04</td>
</tr>
<tr>
<td>D00002</td>
<td>534</td>
<td>185</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>2</td>
<td>11</td>
<td>3008</td>
<td>2013-01-07</td>
</tr>
<tr>
<td>D00003</td>
<td>846</td>
<td>296</td>
<td>6</td>
<td>9</td>
<td>-3</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td>25</td>
<td>15</td>
<td>10</td>
<td>3008</td>
<td>2013-01-09</td>
</tr>
<tr>
<td>D00004</td>
<td>1495</td>
<td>489</td>
<td>55</td>
<td>19</td>
<td>36</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>23</td>
<td>16</td>
<td>7</td>
<td>3008</td>
<td>2013-01-10</td>
</tr>
<tr>
<td>D00005</td>
<td>872</td>
<td>282</td>
<td>3</td>
<td>5</td>
<td>-2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>17</td>
<td>-5</td>
<td>3008</td>
<td>2013-01-14</td>
</tr>
<tr>
<td>D00006</td>
<td>573</td>
<td>183</td>
<td>2</td>
<td>5</td>
<td>-3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>9</td>
<td>-1</td>
<td>3008</td>
<td>2013-01-21</td>
</tr>
</tbody>
</table>
Deep Learning and Neural Network
Deep Learning and Neural Networks

Input Layer (X)  Hidden Layer (H)  Output Layer (Y)

X1  X2  Y
Deep Learning and Neural Networks

Input Layer (X)

Hidden Layer (H)

Output Layer (Y)
Deep Learning and Neural Networks

Input Layer (X)

Hidden Layers (H)

Output Layer (Y)

Deep Neural Networks

Deep Learning
Deep Neural Networks

Input Layer

\( X \)

\( X_1 \)

\ldots

\( X_n \)

Hidden Layer 1

\( H_1 \)

Hidden Layer N

\( H_n \)

Output Layer

\( Y \)

\( Y_1 \)

\( Y_2 \)

\( Y_3 \)
Construction flow chart of deep learning prediction model
## Comparison of editorial team and contents of news providers

<table>
<thead>
<tr>
<th>News providers</th>
<th>Platform</th>
<th>Content</th>
<th>Editorial team</th>
</tr>
</thead>
<tbody>
<tr>
<td>NowsNews</td>
<td>Electronic media</td>
<td>Comprehensive</td>
<td>Owned</td>
</tr>
<tr>
<td>AppleDaily</td>
<td>Electronic media/newspaper</td>
<td>Comprehensive</td>
<td>Owned</td>
</tr>
<tr>
<td>LTN</td>
<td>Electronic media/newspaper</td>
<td>Comprehensive</td>
<td>Owned</td>
</tr>
<tr>
<td>MoneyDJ</td>
<td>Electronic media</td>
<td>Finance</td>
<td>Owned</td>
</tr>
</tbody>
</table>
Developing Finance Sentiment Dictionary

- NTUSD
- HowNet-VSA
- NTUFSD
- iMFinanceSD
# Sample iMFinanceSD opinion words

<table>
<thead>
<tr>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No</strong></td>
<td><strong>No</strong></td>
</tr>
<tr>
<td><strong>1</strong> 衝破 (break through)</td>
<td><strong>1</strong> 震盪 (shock)</td>
</tr>
<tr>
<td><strong>2</strong> 穩住 (stabilize)</td>
<td><strong>2</strong> 低於 (less then)</td>
</tr>
<tr>
<td><strong>3</strong> 持注 (inject)</td>
<td><strong>3</strong> 放緩 (slowdown)</td>
</tr>
<tr>
<td><strong>4</strong> 歷史新高 (Historical high)</td>
<td><strong>4</strong> 走緩 (slowly)</td>
</tr>
<tr>
<td><strong>5</strong> 不遑多讓 (be no slouch)</td>
<td><strong>5</strong> 大幅砍殺 (Stabbed sharply)</td>
</tr>
</tbody>
</table>
Experimental Results and Discussion
Total news of each news provider for Financial Sentiment Analysis (Text Data)

<table>
<thead>
<tr>
<th>Provider</th>
<th>NowNews</th>
<th>AppleDaily</th>
<th>LTN</th>
<th>MoneyDJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>5,499</td>
<td>456</td>
<td>1,147</td>
<td>1,370</td>
</tr>
</tbody>
</table>
ROI of 5 days trading with Deep Learning predicting stock price trend
ROI of 20 days trading with Deep Learning predicting stock price trend
ROI of 60 days trading with Deep Learning predicting stock price trend
ROI of 5 days with Lexicon-based Trading

NowNews | AppleDaily | LTN | MoneyDJ

Returns (000'NTD):

-2,000 -1,500 -1,000 -500 0 500 1,000 1,500 2,000

1.50% 1.00% 0.50% 0.00% -0.50% -1.00% -1.50%
ROI of 20 days with Lexicon-based Trading

Returns (000'NTD)

<table>
<thead>
<tr>
<th>NowNews</th>
<th>AppleDaily</th>
<th>LTN</th>
<th>MoneyDJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>12,000</td>
<td>2,000</td>
<td>6,000</td>
<td>10,000</td>
</tr>
</tbody>
</table>

ROI over 20 days with Lexicon-based Trading.
ROI of 60 days with Lexicon-based Trading

Returns (000'NTD)

NowNews | AppleDaily | LTN | MoneyDJ

Returns: 25,000 | 5,000 | 20,000 | 30,000

ROI: 14.00% | 0.00% | 12.00% | 10.00%
ROI Heatmap of Lexicon-base Trading

![Heatmap Image](image-url)
ROI Heatmap of Trading with Deep Learning Approach

NowNews | AppleDaily | LTN | MoneyDJ
---|---|---|---
0.80% | 0.79% | 1.02% | 1.27%
1.54% | -0.64% | -0.46% | 5.70%
12.74% | 1.50% | 14.50% | 22.43%
Comparison of ROI Heatmap from Various Finance News Media

NowNews  AppleDaily  LTN  MoneyDJ
Conclusions
Findings

• We proposed **analytical methods with deep learning in financial news sources** on the stock price trend forecasts.

• The results showed that the source of financial news media for the exclusive domain of **Finance and Economics**, revealed its investment information representing a reference value.
Contributions

• The different news media release financial news different from its reference value level messages containing investment, as investors choose finance message referenced sources.

• The prediction accuracy will be improved via a prediction model of the deep learning.
Managerial Implications

• Different news media with their own characteristics and specializations.
• The values of the financial information may be different due to the following reasons:
  – Company’s business principles.
  – Edition team’s specializations and their knowledge of industry.
  – Journalist’s habits and preferences in wording.
  – Sensitivity of financial market trends of the Media.
Q&A

Deep Learning for Financial Sentiment Analysis on Finance News Providers

Min-Yuh Day

Department of Information Management
Tamkang University, Taiwan

myday@mail.tku.edu.tw

Chia-Chou Lee
Summary

• Big Data Analytics
• FinTech
• Financial Sentiment Analysis
Big Data Analytics for Financial Sentiment Analysis in FinTech
(金融科技財務大數據情感分析)

Time: 2016/11/15 (Tue) (10:25-12:10)
Place: 朝陽科技大學資訊工程研究所 <教室：E520 演講廳>
Host: 吳世弘 教授 (Professor Shih-Hung Wu)

Min-Yuh Day
戴敏育
Assistant Professor
專任助理教授
Dept. of Information Management, Tamkang University
淡江大學 資訊管理學系

http://mail.tku.edu.tw/myday/
2016-11-15