AI in finance final report

Banruo Zhang
Department of Mathematics
The Hong Kong University of Science
and Technology
bzhangaz@connect.ust.hk

Chuchu Qiu
Department of Mathematics
The Hong Kong University of Science
and Technology
cqiuac@connect.ust.hk

Lu Li
Department of Mathematics
The Hong Kong University of Science
and Technology
llibp@connect.ust.hk

Mandi Tang
Department of Mathematics
The Hong Kong University of Science
and Technology
mtangae@connect.ust.hk

Yawen Zhou
Department of Mathematics
The Hong Kong University of Science
and Technology
yzhoucm@connect.ust.hk

Abstract

This report applies three methods in AI aspect to monitor the performance of two AI related firms, including sensetime and Yewno. The result shows that Yewno has expanded its business into financial industry recently, while sensetime has shifted its business from IT or computer industry toward education and high technology industry. The paper also includes reflections on two articles, which are in finance and AI perspectives respectively.

1 Background introduction

1.1 Company choice

In the project, we choose two companies to monitor their developments by time which are Sensetime and Yewno, in addiction to that, we also choose the counterpart Yitu to make cross-comparison with Sensetime.

We choose Sensetime for the following reasons:
a. Sensetime largely participated on Chinese market, thus we can collect abundant information and data revealed officially for analysis.
b. Sensetime is the largest "Unicorn" AI company, it is reasonable to analyze the representative of AI industry.
c. Sensetime launched a series of advanced artificial intelligence technologies serving multiple industries.

We chose Yewno for following reasons:
a. Yewno has significant partnerships with notable institutions, thus we can monitor it from value chain.
b. Yewno have active media activities, it updates company news on twitter in a timely manner, convenient to track company big events.

1.2 Company introduction

Sensetime:

As the world's leading artificial intelligence platform company, SenseTime Technology is a new generation of artificial intelligence open innovation platform of "Intelligent Vision" by the Ministry of Science and Technology of China. At the same time, SenseTime Technology is also a world-renowned AI innovation company, with a total leading amount and valuation in the industry.

SenseTime Technology has a mission of "holding on originality and let AI lead human progress". The company independently developed and established the world's leading deep learning platform and supercomputer center, and launched a series of advanced artificial intelligence technologies, including face recognition, image recognition, text recognition, medical image recognition, video analysis, driverless and remote sensing. Wait. SenseTime Technology has become the provider of Chinese head AI algorithms.

SenseTime Technology is a leading market share in a number of vertical sectors, covering smart cities, smart phones, Internet entertainment, and advertising, automotive, finance, retail, education, real estate and many other industries. At present, Shangtang Technology has established cooperation with more than 700 world-renowned enterprises and institutions at home and abroad.

Yewno:

Yewno's mission is that of extracting knowledge from an overwhelming quantity of unstructured and structured data. Its technology helps to overcome the "Information Overload" problem and to research and to understand the world in a more natural manner.It is inspired by the way humans process information from multiple sensorial channels and it leverages state-of-the-art Computational Linguistics, Network Theory, Machine Learning, as well as methods from the classical Artificial Intelligence.

2   Methodology

2.1   Value chain

Analyze the value chain is collecting information and data from upstream companies and downstream companies as well as the cooperation or competition between them, hence to monitor the business activities of the company.

2.2   Keywords generation

For text mining we use RAKE algorithm to extract keywords.

The RAKE algorithm is used to extract keywords (keywords), actually it extracts key phrases (phrases) and tends to longer phrases. In English, keywords usually include multiple words, but rarely contain punctuation and deactivated words. For example, and so on, as well as other words which do not contain semantic information. The RAKE algorithm first use punctuation (such as half-width, question mark, exclamation point, comma, etc.) to divide documents into clauses. Then, for each clause, use the word as a separator to break the clause into several phrases. These phrases are used as candidates for the final extracted keywords. We noticed that each phrase can be further divided into several words by a space, and each score can be scored by assigning a score to each word. The key point is to consider the co-occurrence relationship of each word in the phrase. The final definition is:

- **wordScore = wordDegree(w) / wordFrequency(w)**

That is, the score of the word w is the degree of the word (a concept in a network, each time a word is shared with a phrase, the degree is added by 1, considering the word itself) divided by the word frequency of the word (the word is in the document) The total number of occurrences in the document).

Then for each candidate key phrase, the scores of each of the words are accumulated and sorted, and RAKE considers the first third of the total number of candidate phrases to be the extracted keywords.

In addition, for the part of the score calculation, wordDegree(w) is actually equal to the number of words and the number of words in each phrase plus the frequency of the word.

2.3 Data visualization

For Data visualization, we use wordcloud package in Python to visualize keywords.

Word cloud is a visual representation of the "keywords" that appear frequently in text data. The formation of keywords forms a cloud-like color picture, so that you can see at a glance. In python, there is a package named wordcloud. It is easy for us to do the visualization with the help of this package.

3 Workflow

3.1 Sensetime

3.1.1 Data source

We crawled the news in the official website, we crawled the news between sensetime and YiTu from 2017-2019.

3.1.2 Crawling

We use request package in python to crawl news in sensetime. Firstly, we analyse the information of sensetime by typing F12. We find that there are many pages of news, so we obtain the link of each page and the link of each piece of news in one page. Then, we access the link of news via python automatically and obtain the content, time and title of news by searching the tag of them in webpage. Finally, all the information we need would be downloaded as csv file. But for news webpage in yitu, it is dynamic webpage. Firstly, we should observe the change of URL by clicking the network button. After finding the change, then do it as we can crawl to sensetime.

3.1.3 Keywords generation and Data visualization

For Keyword extraction, we extract the word according to Rake, an open source algorithm from github. We extract keyword for every company each year and use wordcloud in Python visualize the result.

3.2 Yewno

Yewno deploys its business mainly on three aspect, including education, publishing and financial services. We will look at Yewno's partnership with other entities in each industry and attempt to track Yewno's business development in each field. Currently, we only take advantage of the media and press release resources on the website of Yewno to find its partners.

4 Results

sensetime

2017	2018	2019

Vertical comparison:

We can see that from 2017-2019, the main keywords of Sensetime are constant, which are AI technology and Sensetime, for Sensetime is a company that provides AI technology service to multiple industries, while we can also see the development by time through the change of keywords.

In 2017 the main keywords of Sensetime contain qualcomm, computer, vision and capital, during 2017, Sensetime have lots of cooperation with Qualcomm, and mainly provide products and service to computer and vision related companies, we can also see that sense time have frequent financing activities and capital absorption in 2017.

In 2018 the main keywords of Sensetime contain 3D, camera and smartphone. During 2018, Sensetime set up cooperation mainly with smartphone and camera related companies such as oppo, xiaomi and so on, because the service that Sensetime provided including face recognition, image recognition, medical image recognition, video analysis are highly related and applicable to such companies.

In 2019 the main keywords of Sensetime contain passenger, city, education and innovation. During 2019, Sensetime set up cooperation with education and government institution instead of business institution, to provide service for better education and better city construction. It also put much attention to innovation and provides service to automobile companies.
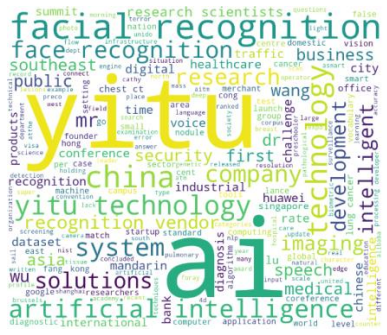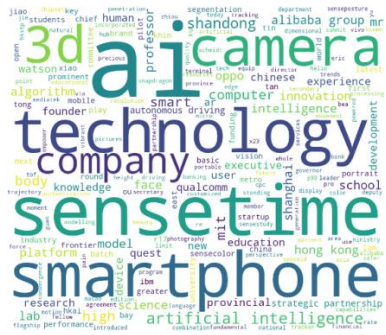
Cross comparison:

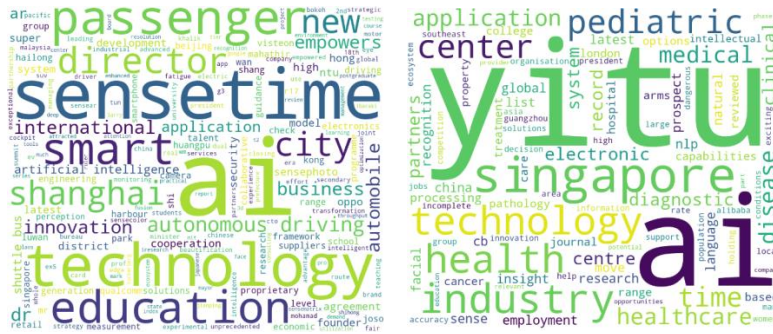| **sensetime**	**Yitu**

2017
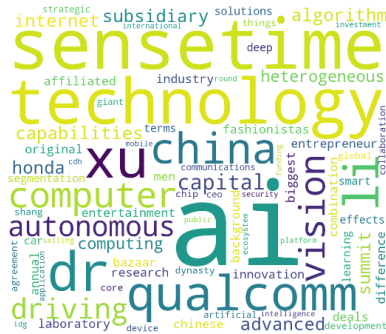
2018

2019

We can see that the commonalities of Sensetime and Yitu is that they both provide AI technology and they both put much attention to innovation, but the differences are that

1. Yitu concentrated more on facial recognition service while Sensetime on facial recognition, text recognition, image recognition as well as video analysis.

2. Yitu set up cooperation with multiple industries but mainly health industry, while Sensetime mainly on smartphone industry.

Yewno:

Data	Partner	Industry
19-Jul-17	**Bryan College (TN), Hillsdale College (MI), Roger Williams University (RI), and Stonehill College (MA)**	Edu
15-Aug-17	**The University of Texas Rio Grande Valley (UTRGV)**	Edu
24-Aug-17	**Hamish McDonald(a respected library technology expert)**	Edu
31-Aug-17	**Project MUSE, Annual Reviews and IOP Publishing.**	Pub
26-Sep-17	**Cambridge University Press**	Pub
7-Nov-17	**Silverchair Information Systems (Silverchair)**	Pub
16-Nov-17	**Publishers Communication Group (PCG)**	Pub
15-Jan-18	**De Gruyter, Stanford University Press, Brill, and Wiley**	Pub
22-Jan-18	**STOXX(Deutsche Boerse)**	Fin
2018/2/25	**Georgetown College**	Edu
19-Mar-18	**US Federal Government**	Gov
22-Mar-18	**Bavarian State Library**	Edu
10-Apr-18	**Vita-Salute San Raffaele University**	Edu
21-May-18	**FactSet(Finanical Research)**	Fin
30-Jul-18	**Coinsquare**	Fin
20-Aug-18	**Inderscience(journal publisher)**	Pub
9-Jan-19	**Santa Clara County Library District (SCCLD)**	Edu
14-Jan-19	**Institute of Civil Engineers (ICE)**	Pub
18-Jan-19	**Benzinga**	Fin
28-Jan-19	**Warren High School**	Edu
7-Feb-19	**DWS/Deuysche Boerse**	Fin
11-Feb-19	**Peter Lang(leading academic publisher)**	Pub

Number of partners
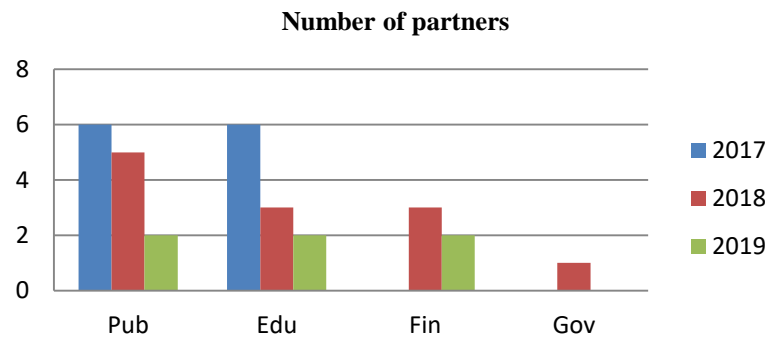
We can see from tables above that Yewno's deployments in publishing and education have decreased during these 3 years. And the deployment in finance has decreased in 2018 and 2019.

5 Literature review

5.1 AI

Nowadays while machine learning is a popular field, data used for training is not limited to the numerical one. The rich information behind text data has been drawing lots of attention. People tried to implement various tools in order to dig the treasure on the text to support their decision making process. Among all, using tools like machine learning, deep learning has growing to be a big subject on artificial intelligence, which is natural language processing.

Based on this background, the paper selected this time is Sentence-Level Sentiment Analysis of Financial News (Lutz, Pröllochs, Neumann, 2019). The objective is to analyze semantic textual content in financial news on a fine-grained basis. While the bag-of-words as one of the most popular direction on performing text categorization, which is to count the frequency of a given set of terms, has received remarkable outcome, on the other hand though, the shortcomings are also very obvious at the same time. Simply counting the frequency cannot produce negation context and may causes information loss since such method failed to identify the meaning neither sentiment under different word under. For example, "reduced cost and increased profit" and "reduced profit and increased cost" points to totally different situation. Hence, a deeper insight failed to be retrieved under such approach. This paper tackles on this issue through statistical learning.

There are mainly two tools used on the paper. First the authors use distributed text representation to transform the text data into numerical vectors for later model training. Then they implemented multi-instance learning to train a classifier in order to transfer the information from document-level to the sentence-level. Interestingly enough, the label they chose is the stock return. Hence, they are trying to identify the abnormality in financial news and further make prediction on the effect that will cause on stock market of that share. In fact, as shown on the paper, the authors successfully identify all the negative and positive sentiment layng behind the data chose. And final predictive performance reached more than 3.8% accuracy compared to the traditional method.

On the other hand, though, several problems exposed after the reading. First during the data cleaning process, news length out of a certain range is filtered out, which raised reasonable doubt that the final model in fact failed to capture the whole picture. And it also leads to the second question that the result of resulting dataset is "a majority of 53.16% of all sentences are assigned a positive polarity, whereas the remaining 46.84% are assigned a negative polarity… Specifically, we see that positive news contain 57.70% positive sentences and 42.30% negative sentences. In contrast, news with a negative market reaction contains only 47.62% positive sentences and 52.38% negative sentences."(Lutz, Pröllochs, Neumann, 2019), where we can easily see here the distribution of positiveness and negativeness on good or bad news does not display a significant difference, which could post challenge on the later stage on prediction. Thirdly, the model under

the essay is built upon the availability of stock price, as well as a decent quantity as well as quality of financial news, which might not be the case for companies that are not yet being public, such as start-ups. Hence, such method may not be quite replicable on all scenarios.

Regardless to the concerns raised above, it is undeniable that the outcome achieved by the authors is very encouraging. And it is believed that there is definitely tremendous potential on this topic.

5.2 Finance

Finance is an information business that is susceptible to information technology. The financial industry is huge and profitable, but it lacks efficiency in some parts. In developed countries, strict supervision may slow down the pace of development of the financial industry, creating a vision and disconnecting from social needs. Flexible fintech companies can take advantage of these visions to bypass traditional operators who still use large amounts of physical or paper. As for developing countries, a large part of the society still has no access to financial services, and financial technology is creating new opportunities.

The main components of financial technology are generally considered to be: financing (including P2P lending), crowdfunding, wealth technology/investment technology (including investment advice and trading activities for robots consulting) and insurance technology; payment and settlement; evidence (including analysis, Monetization and cyber security); customer interfaces (such as smart phones, social media and Internet programs); blockchains and etc.

At present, there are 21 financial technology companies in the world, "unicorns" (that is, billion-dollar enterprises).[1] Among developed countries, the UK is the world's leading financial technology center.[2]

Among developing countries, there's a huge market. Four-hundred million people, or five times the population of Germany - the most populous country that lies entirely within Europe. That is the user base of Alipay, the online payments service offered by Ant Financial, the financial affiliate of Chinese e-commerce platform Alibaba.

Mainland China has the world's largest financial technology industry. Alibaba's Ant Financial and Tencent's WeChat are growing rapidly, attracting hundreds of millions of users. This is inconsistent with the mainland's original financial system, preferential treatment of state-owned enterprises, and a large part of the population and most of the private sector does not use banking services. 40% of Chinese consumers are using financial technology to pay, while Singapore's plaque is 4%. 35% of Chinese consumers are using financial technology-based insurance products, which are 1-2% in many Southeast Asian countries.

Among Chinese financial technology companies, especially third-party payment platforms are developing optimally. The scale of third-party payment in China expanded more than 74 times in the six years from 2010 to 2016. 16% came from consumer-related businesses and 56% from P2P (person-to-person) transfers, overtaking the size of the American counterpart and leaving it far behind.

95% of Chinese netizens use the mobile device to access the Internet, and 68% of them use mobile payments in their daily lives. Third-party payments made through mobile devices account for 75% of total online payments. In contrast, 80% of e-commerce transactions in the United States still require payment through computer terminals. [3] In China, traditional bank cards have been retired as the payment route for large-scale transactions, with an average single transaction amount of 6424 RMB; while the average transaction amount of third-party payment is 606 RMB, the transaction volume of the former is nearly 10 times.

Chinese fintech firms jumped on this opportunity and are leaving traditional banks behind both in terms of size - in some cases they have more end customers than banks - and in terms of innovation.

Alibaba now has a separate finance arm called Ant Financial. It includes online payments service Alipay, online money market fund Yu'eBao, which had Rmb621bn ($93.24bn) in assets as of the

end of 2015, and online bank MYbank.[4] Tencent established the first web-based lender to launch in China – WeBank, based on 1.5 billion monthly active users of QQ & WeChat.

In this way, the bank is not only weakened into a functional role to provide an account, but in the process, the bank loses not only the payment of fees, but more importantly, the massive transaction data will be used by third-party payment platforms. Interception", cannot reach the hands of the bank. It is impossible for banks and central banks to grasp the specific transaction information and to grasp the accurate flow of funds.

At the same time, this model has brought great difficulties to the financial work of central banks such as anti-money laundering, financial supervision, monetary policy adjustment, and financial data analysis. This is a key issue related to data privacy and even data ownership. It has already attracted the attention of the banking industry and even the government regulatory authorities.

At present, in order to protect the customers, information channels and business scale of the bank, the government has introduced relevant regulations in 2018, and the network payment services involving the bank accounts accepted by the payment institution are all processed through the network connection platform. Any third-party payment institution wants to access the bank, and users can make cross-bank transfers.

There are only two ways in the future. One is to go through the clearing channel of UnionPay, and the other is through the network-connected platform. As a fourth-party financial service platform, the networked platform provides services to the market and consumers indirectly through coordinating banks and third-party financial institutions without directly providing financial services to consumers and without touching funds. [5] This provides the Chinese government with a regulatory approach to effectively grasp the specific commodity trading information and capital flows, prevent money laundering, misappropriation of provisions, etc., and its establishment also controls the risks of the third-party payment industry.

6   Conclusion

In brief, according to the analysis we have done previously, we notice that the keyword of sensetime from 2017 to 2019 has shifted from quancomm to education. The most frequent keyword, however, is still AI. As for the main competitor of sensetime, which is Yitu, the main direction of its business is facial recognition, while the business direction for sensetime is more diverse. The strategy of Yewno is studied by collecting and analyzing news from social media, counting the number of partners, and grouping the partners according to the industry. The result shows that the deployment of Yewno has shifted from publishing and education toward finance.

Articles we selected from fintech and AI give us a clear view on the development and current status of fintech. The traditional finance suffers from many drawbacks and challenges, for instance, massive data volume, complicated process, etc. The development of fintech has fixed these obstacles. The representative applications including P2P, AI investor, and blockchain, have significantly simplified the process of many businesses. However, the development of fintech also influences the traditional business of banks, especially payment channel. Also, it is getting harder for regulators to supervise the market since there are more channels for people to conduct financial activities. Currently, fintech is more mature in developed countries like USA and UK. However, fintech in China has a very positive prospective, because it has powerful technology industry and abundant application scenarios.

7   Suggestions for further study

In future, we will continue monitoring the performance of companies by analyzing abnormalities of the company using clustering. If the event or news does not contain the keyword that related to the business of the company, then this kind of event should be regarded as abnormalities. Also, we will further analyze the partners of Yewno. Previously, we only divided them by industry. Next step, we will try to quantify them by their scales, ranks in the industry, etc. So we can further study the business strategy or performance of Yewno.

References

[1] Bernhard Lutz, Nicolas Pröllochs, and Dirk Neumann. "Sentence-Level Sentiment Analysis of Financial News Using Distributed Text Representations and Multi-Instance Learning." Proceedings of the 52nd Hawaii International Conference on System Sciences, 2019. doi:10.24251/hicss.2019.137.

[2] CB Insights，https://www.cbinsights.com/research-unicorn-companies ，14th May 2019

[3] E&Y Fintech, UK FinTech – On the cutting edge, Feb. 2016

[4] https://wallstreetcn.com/articles/3024196

[5] http://money.people.com.cn/bank/n1/2017/0807/c202331-29452876.html

[6] Sibos: Fintech - China - The irresistible rise of Chinese fintech - Expansion plans. Palma, Stefania. The Banker; London (Sep 2016): n/a.

Individual contribution:

Banruo Zhang:
1. Analyze Yewno's performance using value chain.
2. Gather Yewno's partners' information and detected trend of Yewno's business deployment.
3. Consolidate the final report, and collaborate the whole team to make sure the quality of the report.
4. Do presentation.

Mandi Tang
1. Business view about Yewno, its upstream and downstream.
2. Fintech literature review.
3. PPT slides

Chuchu Qiu:
1. Data crawling and data visualization of sensetime and Yitu, Big events and abnormality of sensetime
2. Write technical report
3. Prepare part of ppt

Yawen Zhou:
1. keyword generation, writing technical report, vertical and cross analysis of sensetime and yitu
2. Prepare part of ppt
3. Take research of sensetime partnership

Lu Li:
1. Participated in the research for Yewno such as encoding program for crawling data from the company's twitter
2. performed literature review and wrote reflection on selected essay focused on AI aspect
3. Collaborated the whole team to make sure the distribution and the quality of the work assignment as well as guarantee meeting the deadline