Artificial Intelligence in Financial Service Industry

Liu Yuchen, Ma Huanyu, Sun Zhuo, Zheng Hanshen
MAFS 6010U Final Report
Hong Kong University of Science and technology
zsunao@connect.ust.hk

Abstract

Artificial Intelligence is rapidly developing and recently has been used in Financial Service Industry .In the first part we research on the Sencetime and discuss on its management background, law & regulation issues and industury conditions. In the second part we focus on the machine-learning techniques to construct nonlinear nonparamatic forecasting models of consumer credit risk.We have seen prospects of AI applying in other Fintech area.Finally there are the suggestions and the individual contribution of our group.

1 Track and Monitor Target Company: Sensetime

Our target company is Sensetime, a developer of deep learning technology-based computer vision solutions, aimed at a broad range of consumer and enterprise applications. The main activities of the company are specialized in security monitoring, smart phone, deep learning chip, financial service, mobile application, smart business.

1.1 Management Background

What we are concerned about is the current management of Sensetime and its relationship with other companies and individuals(See Figure 6 in Appendix). Based on the information collected by crawler, a relationship map can be shown above. Interestingly, two of four main founders, Tang Xiaoou and Wang Xiaogang have already taken a back seat. Xu Bing, Yang Fan, Xu Chiheng, Ma Kun, forming the current management group. Especially, Xu Bing is the chairman of the board and vice manager currently. It is obvious that Sensetime hold shares of a lot of other technology companies, namely Beijing Shu Zeyuan intelligent technology co. LTD,Shanghai Ju Tong Software development co. LTD and so no, further proving that it is not simply a bellwether of artificial intelligence in China.

1.2 Laws and regulations

1.2.1 Risk Information

This part is divided into the following aspects: information of the person subject to execution, dishonest information, legal proceedings, court announcement, administrative penalty, serious violation of law, equity pledge, chattel mortgage, tax arrears announcement, abnormal operation, equity freeze, liquidation information, court notice, judicial auction, judicial assistance, public notice.

It is convincing to conclude that the overall legal risk of Sensetime is relatively low for almost zero legal disputes. It is worth mentioning that all its 42 involved lawsuit cases are with the State Administration for Industry and Commerce. In October 2018, the Beijing court allowed the plaintiff, Beijing Sensetime technology development co., to withdraw its complaint.

32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montréal, Canada.

1.2.2 Intellectual Property Information

The amount of intellectual property can reflect the strength of a technology company efficiently. Below is relevant information.(See Figure 7 in Appendix)

Type	Number
Trademark	50
Software Copyright	98
Website records	14
Patent	100
the copyright of the works	11

Figure 1: Intellectual Property Information

1.2.3 Policy Support

In March 2017, human intelligence was first written into the national government report. In December 2017, the “Three-Year Action Plan for Promoting the Development of a New Generation of Artificial Intelligence Industry (2018-2020)” was released. In December 2018, following Baidu, Tencent, Ali, and IFLYTEC, Sensetime became the fifth national artificial intelligence innovation platform authorized by the Ministry of Science and Technology.

Adhering to the thinking of "Science and technology are primary productive forces”, the government’s support for the artificial intelligence industry continues to increase. Following is the relevant policy support.

Time	Name	Department	Main Content
2017.12.13	The Three-Year Action Plan for Promoting the Development of a New Generation of Artificial Intelligence Industry (2018-2020)	Ministry of Industry and Information Technology	The development of video image identification system, by 2020, the effective detection rate of face recognition in complex dynamic scenes exceeds 97%, the correct recognition rate exceeds 90%, and supports face recognition in different regions.
2017.11	A new generation of artificial intelligence development planning major science and technology project kick-off meeting	Ministry of Science	Four development and innovation platforms
2017.3.5	2017 Government Work Report	State Council	Accelerate the cultivation and growth, including artificial intelligence Emerging industry
2016.5.18	"Internet +" Artificial Intelligence Three-Year Action Implementation Plan	Development and Reform Commission, Ministry of Science and Technology, Ministry of Industry and Information Technology, Central Network Office	By 2018, the establishment of artificial intelligence basic resources and innovation platform, artificial intelligence industry system, innovative service system, and standardization system will be basically established. The release of this policy will spread human intelligence to the government and enterprises.
2015.6.15	“Technical requirements for security video surveillance face recognition system”	State Administration of Quality Inspection, Inspection and Quarantine, National Standardization Administration	It is applicable to the overall planning, scheme design, equipment production, quality control, etc. of video surveillance face recognition system for security protection. Other fields can be used for reference.
2015.4.14	"Opinions on Strengthening the Construction of Social Security Prevention and Control System"	Office of the Central Committee of the Communist Party of China, Office of the Civil Service	Propose network management requirements about resolving conflicts with accurate information, and accurately managing network in the future, which is the development direction of safe city and intelligent traffic management

Figure 2: Policy

1.3 Industry Report

Two major researches will be conducted in this part, namely Sensetime’s status in the industry, and the whole industry’s development situation.

Attributed to National strategic and policy support, AI industry has developed rapidly. Authorized by the Ministry of Science and Technology, Sensetime became the fifth national artificial intelligence innovation platform, following Baidu, Tencent, Ali, and IFLYTEC.

Looking back to the financing stage of Sensetime, it has won investment from many well-funded companies, and according to the data, valuation of Sensetime has reached 35billion, which contribute to its success. Sensetime and other three companies, Cloudwalk, Megvii and Yitutech called " Big 4" in Computer Vision.

1.3.1 Market Structure

As for the factors for industry growth, the principle factors our group concluded are industry demand, which will become the main driving force for revenue growth, and policy support, which will contribute to the improvement of profit.(See Figure 8 in Appendix)

Since Sensetime is known as a tycoon for computer vision, comparing the market structure of Computer Vision industry is necessary. Below are the plots, namely Global Computer Vision Industry and Sensetime. It can be concluded that and consumption accounts for most of Global Computer Vision Industry, while Sensetime focus mostly on its Video Surveillance and Internet Business.

1.3.2 Panorama

The whole industry has been separated into three parts. (See Figure 9 in Appendix)Basement part, which will provide the very basic support for further development, for example, hardware including chips and sensor and software like data resource and cloud computing platform. Technology part and Application part. The picture below shows the core companies which are in each procedure. Sensetime is a company in application.(See Figure 10 in Appendix)

Since AI is a brand-new industry in China, analyzing the whole industry trend by focusing on software is necessary. As the results shown here, PE-TTM most software companies high PE value than all the companies in A share.

1.3.3 Comparable Company Analysis

In terms of valuation of Sensetime, it is advisable to compare it with other companies. Results are shown here.

It can be seen from the above comparison that the current Price-to-Sales Ratio (P/S Ratio) of Sensetime is significantly higher than the industry level. High PS valuations are probably attribute to high revenue growth and a premium for industry visibility as algorithm vendor.

Company	Valuation/Market value	Revenue	PS	Revenue(YoY)	Net Profit	PE
Sensetime	**300~350**	**10**	**30~35**	**300%-400%**	**-**	**-**
Hikvision	3235	498	6	19%	113	27
Dahua	507	237	2	26%	25	20
Iflytek	762	80	10	48%	5	152
NavInfo	295	21	14	-0.3%	5	59
Ecovacs	244	58	4	28%	5	49
Pci-Suntek Technology	221	54	4	26%	3	74
Synthesis Electronic Technology	30	4	8	14%	0.1	300

Figure 3: Comparable company analysis

The P/S ratio is considered a particularly good metric for evaluating young, potential high-growth companies or that may not show an actual net profit every year.

High PS means Sensetime are likely to be overvalued, since investors are paying a big premium today for growth that is expected to happen in the future. However, with any deviation from future growth expectations, the premium paid today may no longer be deserved.

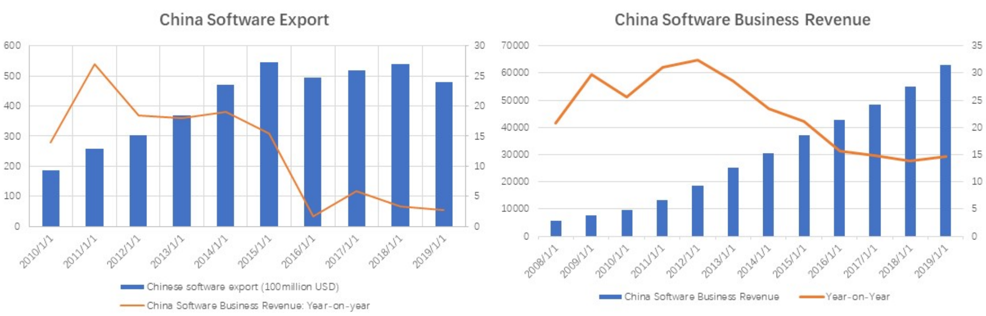

Figure 4: Business

Looking back to the global market, China's export on software and business revenue are growing. However, the year on year growth is decreasing. Turning back to the financial data of whole software industry to analyses the business conditions of all the listed companies, it is easier seek the reasons for the slow growth rate.(See Figure 11 in Appendix)

1.3.4 ROE and ROA

The industry's business climate is sluggish. The operating income is growing, and the profit is going down, especially in 2018.(See Figure 12 in Appendix)

Industry return on equity and return on asset (ROE and ROA fell for 5 consecutive years, hitting a record low ground. That is, the industry is suffering loss during the past few years. The gross profit margin was stable in recent year, which indicates the industry's profitability remained strong.(See Figure 13 in Appendix)

Affected by the rapid expansion of the industry in the past few years, the industry's operational efficiency has further declined, and the total asset turnover rate has been close to the lowest level in the past ten years, in spite of steady increase in total industry revenue.(See Figure 14 in Appendix)

1.3.5 Asset Impairment

The scale of asset impairment in 2018 is the highest in history.

Looking at two proportions: the proportion of asset impairment losses in owner's equity, and proportion of asset impairment losses in net profit. It is easier to find that both ratios are increasing rapidly.

According to the data provided by the institutions, the main contents of asset impairment losses, are goodwill and intangible assets. And that can be the main cause of huge losses in the software industry and explain why the ROE and ROA became relatively low year by year.

After analyzing the data of the industry, it can be concluded 4 major risks for the industry, as well as Sensetime.

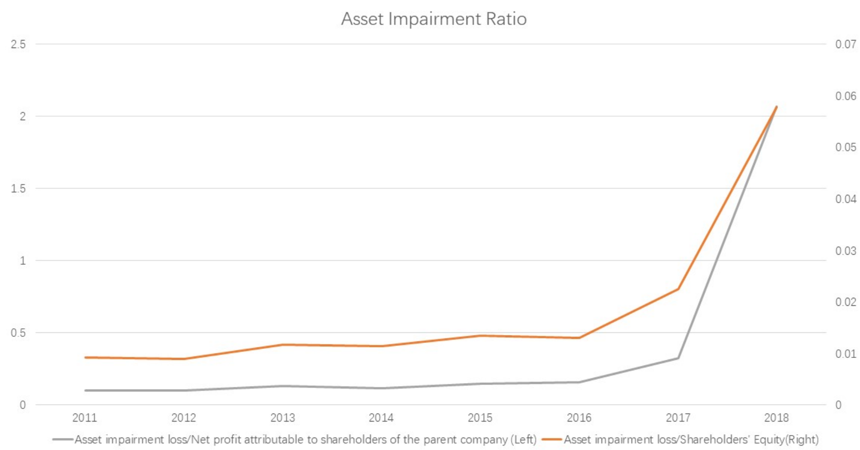

Figure 5: Impairment

2 Reflection on articles

2.1 Financial Services Industry is Spearheading AI Applications

As much as a 40% rise in productivity was reported by banks coming from diverse Artificial Intelligence (AI) applications including productive assistants. A survey of 424 senior executives from financial services and fintech companies released in March 2016 by law firm Baker McKenzie found that 49% of respondents expected their organization to use AI as part of its risk assessment process within the next three years. While 29% expected AI use in know-your-customer and anti-money laundering monitoring, those citing regulation, risk and compliance constituted 26% of the survey respondents.

The seemingly wide range of emerging AI solutions in the financial sector are now expected to be triggered by new capabilities to process high speed and high volume data aided by cloud computing with advanced deep learning techniques. AI leverages a set of technologies rather than one single product or system. The most popular proposition is that of teaching machines to learn and interact to solve cognitive tasks normally done by humans. The implications extend to computers resolving problems, reasoning, processing natural language and much more.

In the Baker and McKenzie survey carried out with Euromoney, three main problem areas for banks to apply AI include risk management, financial analytics and investment/portfolio management. It's all about decision making - internally by the banks and externally on the client side. These could range from decision on branch office location, loan optimizing, customer preferences and investment advice.

2.2 Using AI to slove Financial Problem

"Machine learning" is a branch of artificial intelligence, which is applied in wide field including finance. Such approach refers to build mathematical model based on big data, so called training data, so as to make prediction.

In Khandani's article, they use generalized classification and regression trees (CART) algorithms to build consumer credit-risk forecast model. Based on such machine-learning algorithms, while saving the predicted cost, the effect of model is greatly improved.

2.2.1 Classification And Regression Trees (CART) Algorithms

The CART is an important machine learning algorithm. It can easily be applied to problems with high dimensional feature spaces. The algorithm can either create a Classification Tree or to create a Regression Tree or a Model Tree.

In the process of creating a classification tree recursion, the CART algorithm adopts a technique of binary recursive segmentation. The algorithm always divides the current sample set into two sub-sample sets, so that each non-leaf node of the generated decision tree has only two branches. The dichotomy can simplify the size of the decision tree and improve the efficiency of generating decision trees. In addition, in order to avoid overfitting, the CART decision tree requires pruning. As for the pruning criterion, CART selects the feature with the smallest Gini information gain in the current data set as the node division decision tree.

We define $P_\tau(k)$ as the proportion of training data assigned to class k at leaf nodeτ, then, Gini information could be represented by:

$$G(\tau) \equiv \sum_{k=1}^{K} P_\tau(k)(1 - P_\tau(k))$$

The pruning criterion for CART model T is defined as:

$$C(T) \equiv \sum_{\tau=1}^{|T|} G(\tau) + \lambda |T|$$

Where |T| refers to a number of leaf nodes in CART model T, and λ refers to a regularization parameter chosen by cross validation. Once the pruning criterion reached the minimum, the CART algorithm will stop expending the tree.

The implementation of the algorithm can be represented as the following steps:

Input: Training data set D, stop calculation conditions

Output: CART Decision Tree

According to the training data set, starting from the root node, recursively perform the following operations on each node to construct a binary decision tree:

(1) Set the training data set of the node to D, and calculate the Gini index of the existing feature to the data set. At this time, for each feature A, for each value a that it may take, according to the test of the sample point pair A=a, "yes" or "no" divide D into two parts D1 and D2, and calculate A= The Gini index at a time.

(2) Among all possible features A and their possible segmentation points a, choose the feature with the smallest Gini index and its corresponding segmentation point as the optimal feature and the optimal segmentation point, according to the optimal feature and optimal Split points, generate two child nodes from the current node, and distribute the training data to the two child nodes according to the feature.

(3) Recursively call (1)(2) on the two child nodes until the stop condition is satisfied.

(4) Generate a CART decision tree.

The stop condition of the algorithm: the number of samples in the node is less than a predetermined threshold, or the Gini index of the sample set is less than a predetermined threshold (the samples are basically of the same class), or there are no more features.

2.2.2 Consumer Credit-Risk Model

When creating a regression tree, the observation value is continuous, there is no classification label, and only a value based on the observed data is used to create a prediction rule. In this case, the optimal partitioning rule of the Classification Tree could not be adopted. CART uses the Minimum Residuals Minimization to determine the optimal partition of the Regression Tree. The partitioning criterion is the minimum error variance of subtree e after the expectation partition.

In the consumer credit-risk model, they use a unique dataset consisting of transaction level, credit bureau, and account-balance data for individual consumers.

DATA 1 The bank provides us with account level transaction data for a subset of its customer base. In fact, individuals may have banking relationships with multiple institutions, so this data may only show part of their financial activities. However, in this limited data set, as long as such levels of incompleteness are stable over time, there is also considerable incremental value in consumer credit risk management

DATA 2 Credit documents provided by one of credit bureaus. This data contains information regarding all credit or loan facilities that each customer has across all financial institutions. The credit score (CScore) provided in this data is similar to the standard technical score for measuring consumer credit quality in terms of statistical characteristics and predictive power. This credit score can be used as a benchmark to compare it with the credit risk model established through machine learning to see the performance of the machine learning model.

DATA 3 Account balances from checking accounts and CDs that a customer has with the Bank. Taking these data sets as input, a decision tree could be established based on CART algorithm, which can be used to predict consumers' default behaviors. Under certain conditions, the output of the model can be interpreted as an estimate of the probability of an account becoming 90-days-or-more delinquent during the subsequent 3-month forecasting window.

2.2.3 Result and Model Checking

First, the author compared the result obtained from machine learning model with Cscore. Machine learning predictions differ greatly from CScore scores. In particular, according to the machine learning model, there are a small number of accounts with relatively high CScore scores (note that higher CScore scores indicate higher credit quality or lower credit risk) with higher predicted default and default risk.

To check the effect of model, the difference between the prediction of the accounts with or without actual default should be first considered. If the difference is not obvious, the model prediction has no value. The results showed significant differences in average forecasts for the two types of accounts.

In machine learning, another test is to identify " straight-rollers". the model has the ability of identify straight-rollers by comparing compare the model's average forecast among customers who were current on their accounts but became 90-days-or-more delinquent with the average forecast among customers who were current and did not become delinquent

By implementing the out-of-sample prediction, the result of this model is highly correlated with the actual default rate. In the monthly prediction of 6 months and 12 months, the linear regression R2 reached 85%. A rough estimate of the added value of these projections could save between 6 and 23 per cent of the total loss. In addition, compared to traditional methods, machine learning forecasts are more adaptable and can understand changing credit cycles and absolute levels of default rates.

2.3 Prospects

We can also apply this model to risk management and other aspects. Zhu Min and Philpotts David (2011) combined CART and Logistic Regression for Stock Ranking. Shalini Kalra Sahi and Nand Dhameja(2011) used CART analysis to predict the preference for financial investment products. Benjamin Yeo and Delvin Grant (2018) predicted service industry performance using decision tree analysis. Apart from decision tree, there are many other branches of artificial intelligence, such as natural language processing, biometric recognition, robotics and so on. These technologies have been widely used in various fields of the financial industry.Research shows that artificial intelligence technology can effectively control various risks.

3 Suggestions for Further Study

Artificial intelligence has changed the traditional model. The application of AI technique in finance is in the stage of rapid development. Traditional artificial decision-making is mainly based on business experience, summaries from the long-term practice, thus finding some rules. Such model tend to be hysteretic and subjective. However, artificial intelligence makes use of a large amount of data, algorithm models and machine language, which has more precise control over objective laws. AI technique can capture the intrinsic connection among things more accurately and keenly. In addition, when dealing with problems, especially a large number of problems, artificial intelligence can effectively reduce the time cost and labor cost.

The realization of artificial intelligence in the financial field includes enhancing risk control capabilities, activating sleep data, improving service levels and developing new products. For example, biometrics can improve system service efficiency and improve user experience; intelligent customer service technology improves efficiency and reduces the cost; financial anti-fraud technique improves security.

In the application process, we must pay attention to, first of all, for the data, we need to improve the quality, to ensure its accuracy and usability. Besides, under the guarantee of data security, technology must be extensible so as to interact with external technology. We also need to pay attention to the architecture can be implemented in different scenarios.

Technological progress often drives the innovation of business models. At present, AI technology changes rapidly and is constantly being innovated and revised. The world's major financial giants are actively applying AI technology, developing AI financial products and investing in fintech companies. Such development is based on a mature Internet environment and a stable and reliable information system. The involvement of artificial intelligence in the financial industry could not be underestimated, which has a profound impact on the future development of the financial industry.

4 Individual Contribution

Liu Yuchen Learned basic knowledge about Artificial Intelligence and some machine learning models by learning professional essay. Participated in searching lawsuits and other relevant risk of Sense Time in group project.

Ma Huanyu Extracted and summarized effective information from papers related to artificial intelligence . Investigated and analyzed management background, relevant laws and regulations in group project, and participated in presentation.

Sun Zhuo Searched for industry report and analysis technology industry in the group project. Researched articles on AI's application in Financial Industry and wrote report. Collated documents and formated final report.

Zheng Hanshen Conducted literature search and did literature review in Artificial Intelligence field. Investigated the industry's overall landscape, conducted financial data analysis and produced Youtube video.

References & Appendix

[1] Amir, E. Khandani & Adlar, J. Kim & Andrew, W. Lo (2010) Consumer credit-risk models via machine-learning algorithms *Journal of Banking & Finance***34**(7):2767-2787.

[2] Hong Kong Exchange (2018) *Research Report:Financial Technology applications and related regulatory framework*

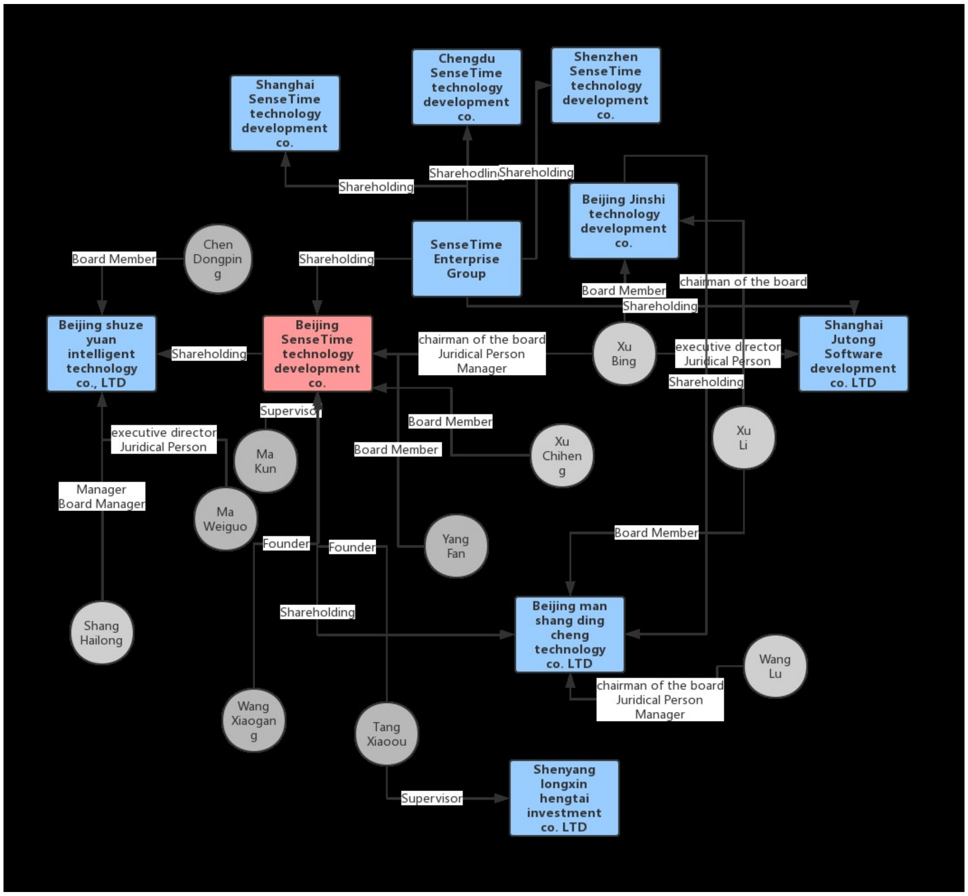

Figure 6: Relationship chart

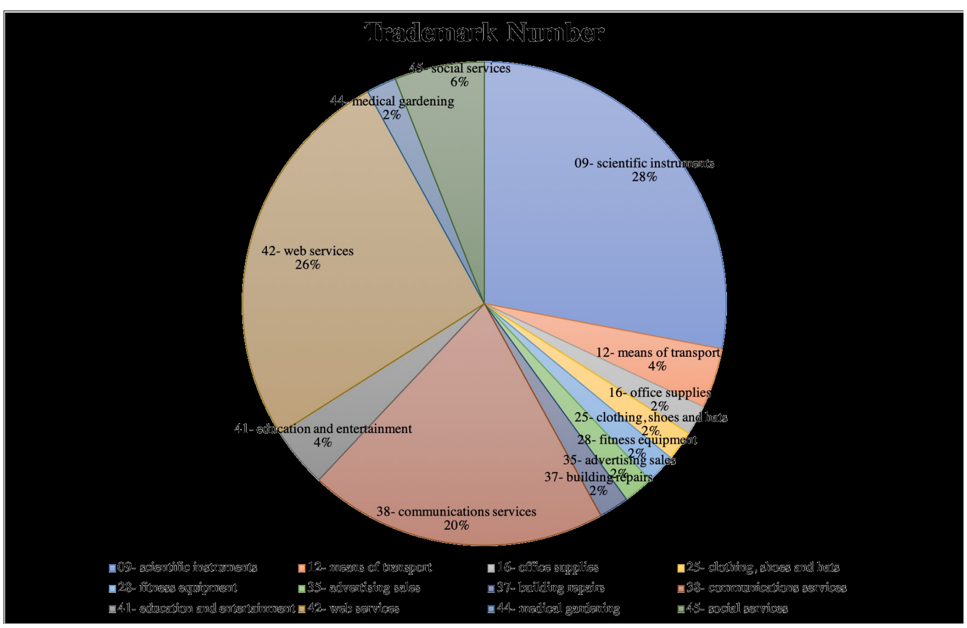

Figure 7: Trademark

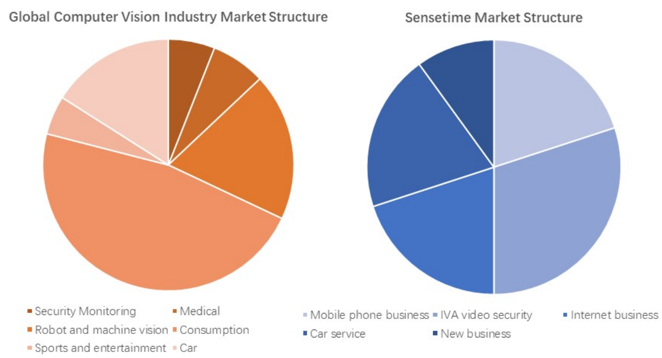

Figure 8: Market Structure

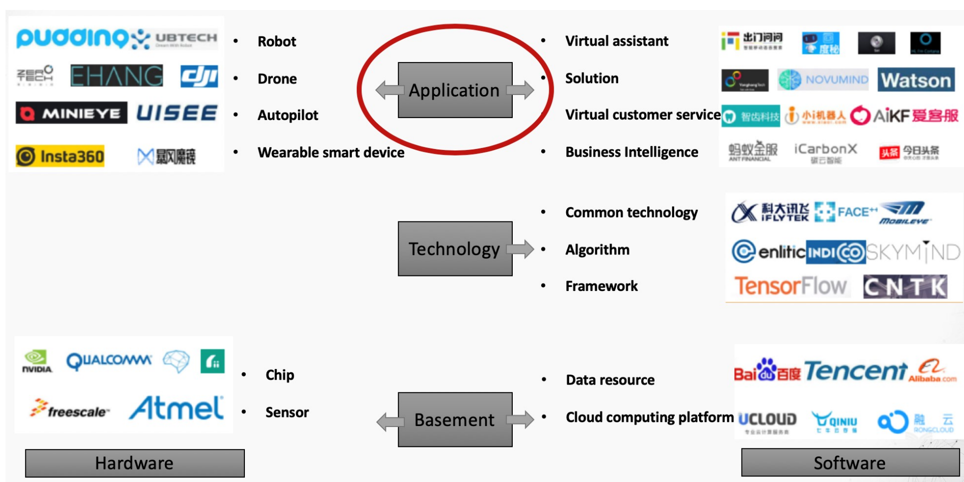

Figure 9: PARO

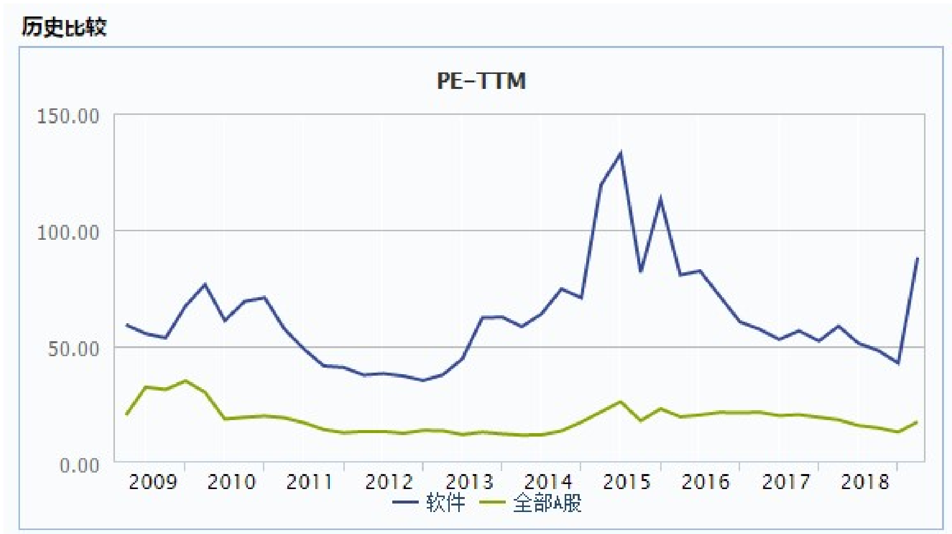

Figure 10: PE

	2018A	2017A	2016A	2015A	2014A
Operating income	3,420.76	3,135.67	2,609.46	1,900.29	1,314.040
Year-on-year (%)	12.23	17.63	21.3	18.72	13.22
Operating cost	2,185.57	1,975.77	1,636.66	1,181.83	837.31
Year-on-year (%)	14.01	17.86	21.28	16.89	10.56
Total profit	143.33	311.43	319.26	239.47	154.71
Year-on-year (%)	-53.4	-4.86	18.4	23.3	22.83
Net profit attributable to shareholders of the parent company	111.46	264.69	271.49	205.41	134.52
Year-on-year (%)	-57.14	-4.77	17.82	22.25	22.72
Total Assets	6,615.97	6,237.68	5,107.63	3,559.82	2,204.62
Total Liabilities	2,621.18	2,397.02	1,791.40	1,287.80	807.8
Shareholders' equity	3,994.79	3,837.29	3,316.23	2,272.02	1,396.82
Operating net cash flow	185.44	141.21	113.4	148.51	86.54
Net cash flow from investment	-249.49	-486.96	-499.68	-282.86	-161.16
Net financing	61	353.46	614.2	425.44	87.14
Net increase in cash and cash equivalents	1.32	3.65	231.9	293.81	11.97
Gross profit margin (%)	36.02	36.62	37.28	37.81	36.28
Sales margin(%)	3.26	8.59	10.63	11.03	10.49
ROE (%)	2.95	7.52	9.71	10.94	11.09
ROA (%)	1.75	4.68	6.18	6.85	6.91
Assets and liabilities(%)	39.62	38.43	35.07	36.18	36.64
Current ratio	1.84	1.94	2.15	2.1	2.13
Total asset turnover rate (times)	0.54	0.54	0.58	0.62	0.66
Accounts receivable turnover rate (times)	2.62	2.73	2.82	2.92	3.07
Inventory turnover rate (times)	3.26	3.25	3.35	3.55	3.77

Source: Wind Database

Figure 11: Financial Data

Figure 12: Profit

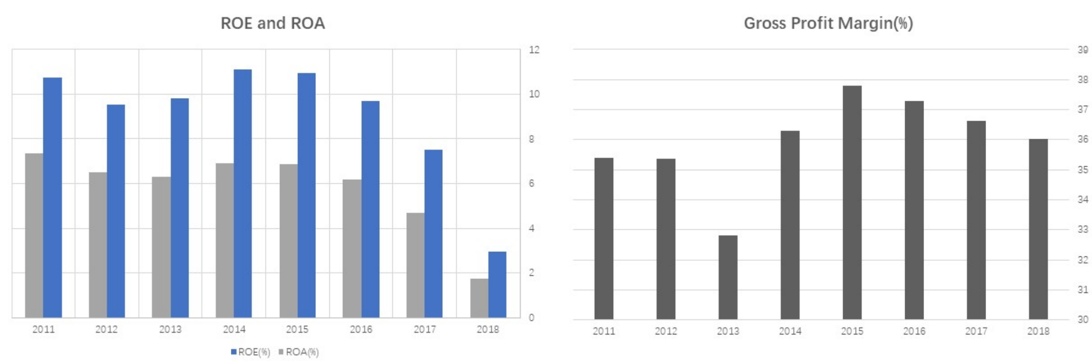

Figure 13: Profit

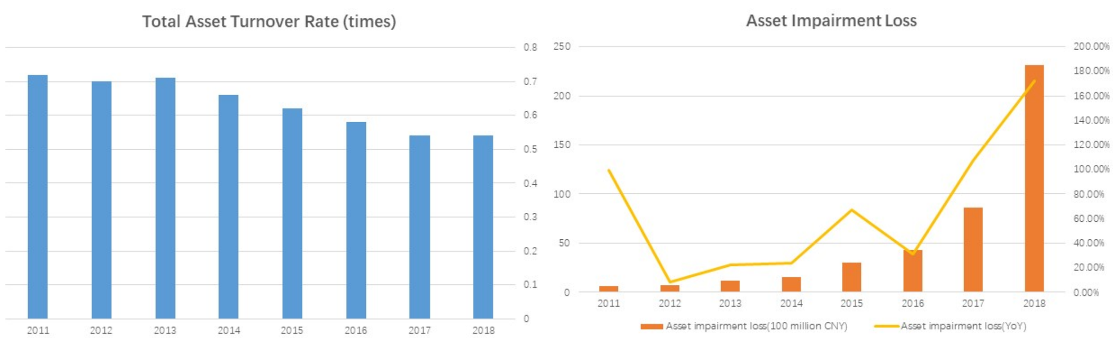

Figure 14: Asset