Group HAWK
AI in Finance Final Report

Ruiting CHEN **Di CHENG** **Yang XUE** **Yifan YE** **Mengkai ZHANG**
Department of Mathematics, School of Science
The Hong Kong University of Science and Technology
Clear Water Bay, Hong Kong
{rchenax, dchengab, yxueai, yyeam, mzhangbh}@connect.ust.hk

Abstract

We aim to monitor and track high-tech AI start-ups like Sensetime and, Megvii, one of its major competitors and present our results for the comparison. Many kinds of AI techniques have been adopted and finally we come to conclusion that Sensetime is developing smoothly and worth of investment. Besides, we present our reflections on some papers and give some future directions for the project.

1 Brief Introduction to Technical Parts

In part 2, we use facial recognition, voice recognition, web crawling, etc. to collect materials from different information carriers and sources.

In part 3, we perform NLP on materials collected in part 2. Through statistics of high frequency words, auto-generated topics and sentiment analysis and comparing these results with Megvii's.

In part 4 we review an AI paper and reproduce the algorithm in it, i.e. apply it to some different datasets to see the performance and further explore next step work.

In part 5 we review a finance paper and strongly agree with that this essay shows that VCs' on-site involvement with their portfolio companies is an vital element of innovation and success.

2 Information Collection

Our major goal is to track updated trends of some AI companies, SenseTime is the part that we care about most. For comparisons among the same industry, we choose our target companies to be:

2.1 Web Crawler

Web crawler acts as an automated script which browses through the Internet in a systematic way. The web crawler looks at the keywords in the pages, the kind of content each page has and the links, before returning the information to the search engine. We use XML Path Language (XPath) to conduct the web crawling in the following steps (set SenseTime as an example):

- *Obtain the information of target website.* For example we choose one piece of news to illustrate: `https://www.sensetime.com/news/view/id/21.html`.

* Final Report – MAFS 6010U Artificial Intelligence in Finance (2018-19 Spring), HKUST.

- *Collect the information of target website.* We need to extract the Title, Date and Contexts.
  1. Locate the Title & Date labels and corresponding codes:

```
<h1 class="new-title">商汤科技与上海市黄浦区教育局联合打造人工智能标杆校</h1>
title = html.xpath("//h1[@class='new-title']/text()")[0]
```

```
<p class="new-time border-bottom">2019-05-07 22:40</p>
time = html.xpath("//p[@class='new-time border-bottom']/text()")[0]
```

  2. We locate the label of Content `<div class="new-nr-con">` with following:

```
contents = ''
neir = html.xpath("//div[@class='new-nr-con']//text()")
```

```
for content in neir:
    contents += contents.strip()
```

- *Save and export the data.* After we extract the three parts of each news, we save them as an excel file for further researching. We give an example below:

```
dic = {
    'title':title,
    'time':time,
    'content':contents
    }
ls.append(dic)
```

```
with open('sensetime.csv', 'w', encoding='utf-8-sig') as f:
    fieldnames = ['title', 'time', 'content']
    writer = csv.DictWriter(f, fieldnames=fieldnames)
    for information in ls:
        writer.writerow(information)
```

2.2 Facial Recognition

- **Introduction:** We use face recognition technology to process video files, and find the key characters of SenseTime from the video to determine whether the video contains content related to SenseTime.
- **Face Detection:** The algorithm needs to slide in the image with a window of different sizes and positions, and then determine whether there is a face in the window.
- **Face Recognition:** when the video captures the face, let the machine determine who the person is. First, the image captured by the video is processed by the ResNet model to return a 128-dimensional face feature vector. After acquiring the feature vector, the Euclidean distance can be matched with the local face feature vector, and the nearest neighbor classifier is used to return the tag of the sample.

2.3 Voice Recognition

- Introduction: Videos from social medias often directly reflect current status of target companies, so research and analysis of video is a very important step. A large number of video files obtained by Web Crawler (via official account, WeChat, twitter, YouTube) and that obtained by the Facial Recognition require speech recognition to convert into text, so as to facilitate the subsequent analysis and research.
- Determine the speech recognition module: Today's speech recognition technology is relatively mature, IBM, Google, Baidu have corresponding speech recognition module which can be called through python. Through comparison of accuracy and usage, we choose to use IBM module whose speech recognition technology is close to humans Horizontal. The SpeechRecognition module is installed to recognize the audio, and the speech can be recognized as text. The code is used for SpeechRecognition of simple audio material (not the material required in the project, only for accuracy test). The accuracy rate is about 90%.
- Convert online video to audio: The SpeechRecognition module can only process the audio that has been downloaded to the local. So how to convert the online video to wav format audio? There are two main methods:
  1. Save the recording with the Microphone through the pyaudio module;
  2. After downloading the video, extract the audio through the ffmpeg package to get the audio in wav format.

  The 1^{st} takes too much time and there maybe external noise interference, so we choose the 2^{nd} with higher technical requirements and get several audios in wav format.

- We need to remove the noise, because most of the videos are speech, etc., which will be interfered by noise. "r.adjust_for_ambient_noise(source)" can remove noise interference.
- Solve problems in coding:
  1. I have experienced a network interruption in the middle when test the code, and the previously recognized content is gone. So, I try to solve this question and find that it is best to adopt a segmentation method to split the audio into shorter segments, which is equivalent to real-time return and reduce the loss. In the project, I divide the audio into 30s segments and use looping statements for speech recognition.
  2. After testing, there are new problems. If the segmentation produces more than 10 segments, the remote server will interrupt the service, that is to say: the interface cannot be adjusted frequently, so I refer to the prompt on IBM bluemix to add 5s waiting time before proceeding to the next cycle.
- Speech recognition: After determining the module, cycle time, and solving the existing problems, we use python to perform speech recognition on the audio. The following is one of the results. It can be seen that there are very few grammar errors (most of the red marks are punctuation errors). A small number of conjunctions are wrong with little effect on the results of the study.

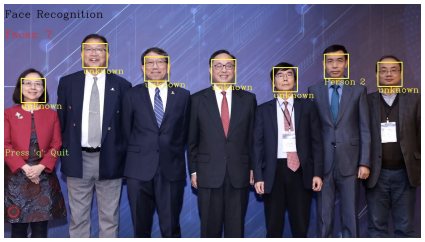


Figure 1: Facial Recognition

Throughout the morning we've heard about the incredible promise of exploring intelligence.The potential to deeply understand how humans learn.To use that understanding together with novel computational methods to create new algorithms.and to apply those methods to virtually every scientific discipline, into every aspectof everyday life.Finance, transportation, health care, social interactions. One might be inclined to infer from thisthat the impact of intelligent algorithms are mostly still to come.And that impact is high.But in many cases, the impact is already here.Tang Xiao'ou and his company sense time are a wonderful example of that.An MIT PhD graduate from 1996, Xiao'ou has decades of experience in computer vision and machinelearning.I had the pleasure of serving on his thesis committee.His supervisor was one of my first students.and I'd been delighted to stay in contact with Xiao'ou ever since.Currently professor at the Chinese university of Hong Kong and a former lead researcher at Microsoft Asia, Xiao'ou is one of the most visible and influential leaders in artificial intelligence, both within China and internationally. Since time, Hong Kong's first

Figure 2: Audio to Text

3 Information Analysis

3.1 Word Clouds

In the generating procedure of these wordclouds, we use the package "jieba" and "wordcloud" to process the materials. In order to get rid of the impact of high-frequency words, we expand the default stopwords list and implement tf-idf to give a inverse weight to high frequency words. Two different language source of news have been considered.

- English Comparison: We present the comparison between the English wordclouds of the two companies. (The wordclouds are based on all time period news we collected.)

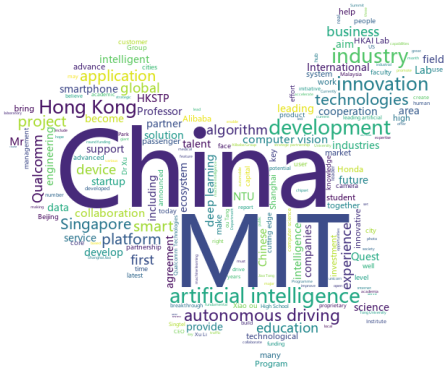


Figure 3: SenseTime English (All-time)

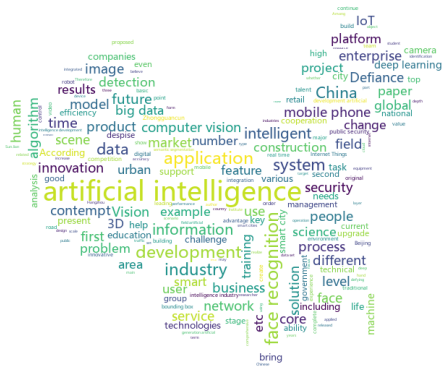


Figure 4: Megvii English (All-time)

- Chinese Comparison: We present the comparison between the Chinese wordclouds of the two companies. (The wordclouds are based on all time period news we collected.)

Figure 5: SenseTime Chinese (All-time)

Figure 6: Megvii Chinese (All-time)

3.2 High-frequency Words and Topic Model Sentences

The wordclouds above are overall statistics, then we will perform some more delicate analysis to our collected information. Specifically, we will perform word frequency analysis to bi-monthly data and display them by order to see the trend. Besides, we use Latent Dirichlet Analysis (LDA) to find the central topic of every bi-monthly collection of news. Further, we use these topics to auto-generate some new sentences, which can be viewed as predictions for future news trend (Those sentences are generated by algorithms so sometimes some sentences don't care about grammar mistakes).

- SenseTime Word Frequency and Topic Model (Bi-Monthly)

	SenseTime	
	Top High-frequency Words	**Topic Model Sentences (Auto-generate by Algorithm)**
May, 2018 and before	China / MIT / Industry	complete image monitor raw
June & July, 2018	Singapore / NSCC / Singtel (telecommunications company)	complementary city establish opportunity
Aug & Sept, 2018	MIT / Project / Science / Quest / Student	experience entity nuclear incorporate / cognitive new executive science
Oct & Nov, 2018	Hong Kong / HKAI Lab / Startup	commit machine mobility equip
Dec, 2018 & Jan, 2019	Autonomous Driving / Singapore / NTU / User / 3D / Smartphone	sensetime ai technology autonomous / also mobile beautification algorithm
Feb & Mar, 2019	Education / Qualcomm (semiconductor and telecommunications equipment company) / CUHK	AI sensetime education school
Apr & May, 2019	Malysia / G3 Global / Visteon (automotive electronics supplier) / Bureau	closely real accompany cockpit / ai educaiton sensetime shanghai

- Megvii Word Frequency and Topic Model (Bi-Monthly)

	Magvii	
	Top High-frequency Words	**Topic Model Sentences (Auto-generate by Algorithm)**
May, 2018 and before	Retail / Fresh Life / Convenience Store	important solve display capability
June & July, 2018	Unicorn / Fu Yingbo / Structured Light / hangzhou	hardware rapid step afternoon / three equipment password android
Aug & Sept, 2018	Face Recognition / 3D / Zhongguancun / Bank Beijing / Algorithm	expensive constantly peking computer / technology AI recognition
Oct & Nov, 2018	Security / Data / Mobile Phone / Bounding Box / Detection / Innovation	RNN foreground mathematical early / coefficient mainland role workshop
Dec, 2018 & Jan, 2019	Smart City / Information / System / Transportation / Urban	intelligence artificial technology city / senior dog quantityt segmentation
Feb & Mar, 2019	Scene / Core / IoT (Internet of things) / Enterprise / Innovative / Enthusiasm	super essential city ecosystem / conform 000 chain robot
Apr & May, 2019	Image / Categories / Objects 365 (a data set) / Financing / Ant Financial / Security	index objects365 food mark / fund expand main competitor

3.3 Sentiment Analysis

This sentiment analysis part is an important and interesting part in the procedure of analyzing our collected information. The package "textblob" is used. Some elements in the sentiment trend figure 7 need to be clarified and the trend should be considered together with news numbers in figure 8.

- we have drawn two lines, the blue line represents the sentiment trend for SenseTime and the orange line represents the sentiment trend for Megvii.
- The range of sentiment score is from -1 (most negative) to 1 (most positive) and the data in news number figure has been standardized into the range between zero and one.

- The x-axis coordinate is orderd by monthly dates.

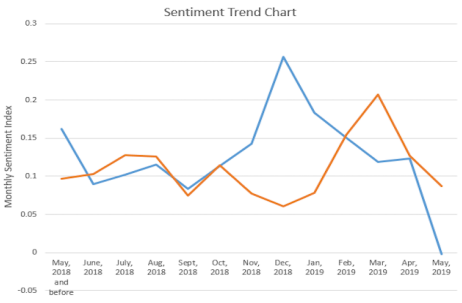


Figure 7: Sentiment Trend Analysis

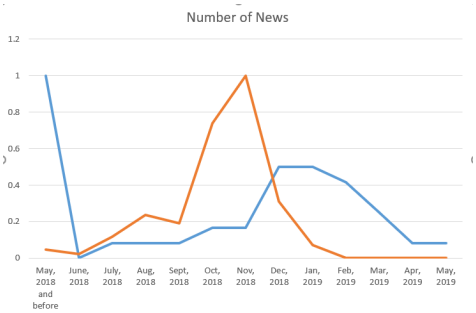


Figure 8: Comparison of News Number

From figures above we know that:

- During most of time, sentiment index of SenseTime is higher than that of Megvii, which means investors and observers have more optimistic anticipation for SenseTime. Index of SenseTime peaks in December of 2018.
- Indexes of both two companies are on the decline in this month (May of 2019), which we believe is highly relevant to the recent breaking news as figure 14 shows. SenseTime & Megvii are both on the black list and negative impacts on SenseTime are even more serious, which is in accordance with the character figure 7 shows, SenseTime drops quicker.
- Delicate analysis about the relationship between index and numbers of reports would be an interesting topic in further study.

3.4 Patent

- **Summary.** As of May 2019, SenseTime applied for 943 domestic patents and 67 foreign patent applications, including some brand new fields other than computer vision.
- **Medical Field.** A patent "Glaucoma Diagnostic Methods, Devices, and Electronic Devices" reflects the company's ambitions in medical imaging diagnosis.
- **Computational Photography.** An original AI super-division technology was applied by mobile phone manufacturer Vivo, which enabled the application of super-resolution technology based on deep learning algorithm to land on domestic mobile phones.
- **Autonomous Driving.** A long-term cooperation agreement with Honda to decide on Honda's vehicle control technology system (L4-level automatic driving scheme).

3.5 Researchers' Background

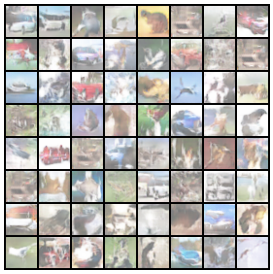
- **Example 1.** Tang Xiaoou, founder of SenseTime Technology, the "blazer" and "pathfinder" of global face recognition technology, began to focus on computer vision technology when he was pursuing his Ph.D. at the Massachusetts Institute of Technology in 1992. In 2001, the Multimedia Laboratory of the Chinese University of Hong Kong was established, which ranks among the top ten artificial intelligence pioneer laboratories in the world.
- **Example 2.** Xu Li, co-founder of SenseTime Technology, has more than ten years of research and product development experience in the fields of computer vision, pattern recognition and image processing. He has published more than 40 papers in international top conferences and journals in the field of vision. He has worked in computer vision research institutions such as Motorola Research Institute, Omron Institute, Microsoft Research Institute, and Lenovo Research Institute.
- Further investigation can be made by looking at more professional websites. We will mention that on section 6 and give an example.

4 Selected Literature in AI

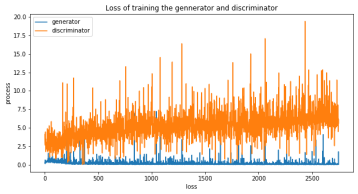
We present two models here, which are originally be considered as our group's option A project. Here we present our reproduction results for the two models. Besides, we will make this part short, in order to adapt the length limit of the report.

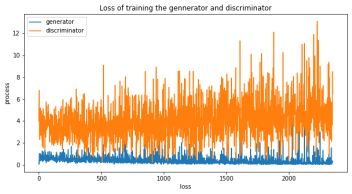
Topic: DCGAN Radford, A., et al. (2015)

- Our Goal: We reproduce the algorithm in this paper, i.e. to generate images through game-based training with a efficient way, apply it to some different datasets to see the performance and further explore next step work.
- Model Construction: The Deep Convolutional GAN in this paper has made several important changes based on other related work, which ensures that we can get a stable training result.
  1. Replace any pooling layers with strided (namely, skip some units), which means, convolutions (discriminator) and fractional-strided convolutions (generator).
  2. Use batchnorm in both the generator and the discriminator.
  3. Remove fully connected hidden layers for deeper architectures.
  4. Use ReLU activation in generator for all layers except for the output, which uses Tanh.
  5. Use LeakyReLU activation in the discriminator for all layers.
- Datasets:
  1. Fashion-MNIST: It contains a training set of 60,000 examples and a test set of 10,000 examples. Each example is a 28x28 grayscale image, labeled from 10 classes.
  2. CIFAR-10: It contains 60,000 32x32 color images in 10 different classes (6,000 images of each class). There are 50000 training images and 10000 test images.
- Final Results:
  - Left: Fashion-MNIST & Right: CIFAR-10



- Loss Function:
  - Although loss function is not important in training GANs (except WGAN), we still display the loss function below for reference. The left loss figure is for the first 50 epochs of training on Fashion-MNIST. And the right loss figure is for the first 50 epochs training on CIFAR-10. Typical patterns of training GAN structure show up.

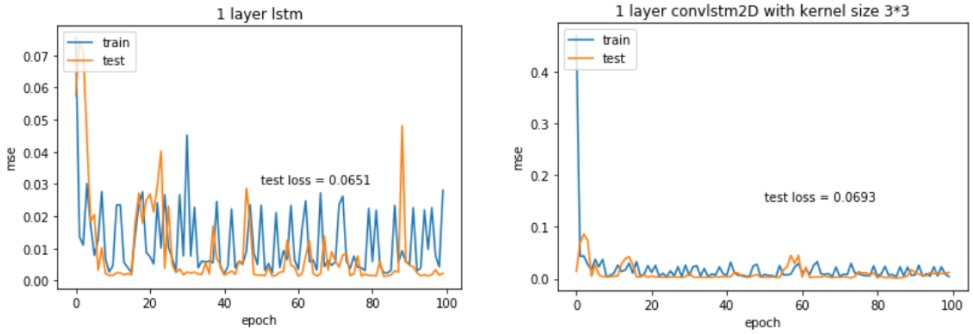




- Future Directions: We may use ResNet to replace the current convolution structure.

Topic: ConvLSTM Xingjian, S. H. I., et al. (2015)

- Our Goal: We use the algorithm in this paper to do some prediction task. Here, past data records from the vegetation index are used to predict future index.
- Model Construction: We adopt the encoding-forecasting structure with ConvLSTM as building blocks. Specifically, the encoding LSTM compresses the whole input sequence into a hidden state tensor and the forecasting LSTM unfolds this hidden state to give the final prediction.
- Datasets: Tiff image files of vegetation index data. The size of each Tiff file is 1200x1200 pixels, and each pixel represents the vegetation index on an area of 1 square kilometer on Earth. 20% data is used as test set.
- Evaluation: Interestingly, under our experiment, the ConvLSTM is not as good as simple LSTM (notice the scale of vertical coordinates or y-axis).



- Future Directions: Some other data sets may be used, to show the actual advantage of ConvLSTM over normal LSTM.

5 Selected Literature in Finance

We chose a famous paper from the No.1 finance academic journal *The Journal of Finance*.

Topic: Venture Capital Bernstein, S., et al. (2015)
Do VCs contribute to the innovation and success of their portfolio companies, or do they simply identify and invest in companies that are already poised to innovate and succeed even absent their involvement? This essay explains the importance of VCs quantitative and qualitatively by reducing monitoring costs:

- **Control Variable (quantitative method)**:
  1. They exploit exogenous reductions in monitoring costs stemming from the introduction of new airline routes that reduce the travel time between VCs and their existing portfolio companies, thereby holding company selection fixed.
  2. The reductions in travel time lead to an increase in the number of patents and the number of citations per patent of the portfolio company, as well as an increase in the likelihood of an IPO or acquisition. These results are robust to controlling for local shocks that could potentially drive the introduction of the new airline routes. The effect is concentrated in routes that connect lead VCs (as opposed to other investors) with portfolio companies. Overall, the results indicate that VCs' on-site involvement with their portfolio companies is an important determinant of innovation and success.
- **Conduct A large-scale survey of VC investors (qualitative method):**
  1. The key assumption underlying the empirical strategy is that VCs are responsive to the treatment, that is, VC involvement increases following a reduction in travel time. Since VC involvement is not observable, this assumption cannot be directly tested. Thus, to assess the plausibility of this assumption, they conduct a large-scale survey of VCs.

2. The authors confirm the importance of this channel by conducting a large-scale survey of VC investors. Almost 90% of the respondents agreed that they would visit a portfolio company more frequently if an indirect flight were replaced by a direct flight. Moreover, survey participants agreed that the introduction of a direct flight would help them establish better relationships with management teams, improve their understanding of the state of their companies, and generally add more value.

- **In conclusion,** this essay shows that VCs' on-site involvement with their portfolio companies is an important determinant of innovation and success.
- **In our opinion,** the introduction of short-haul direct-travel routes and greatly shortened direct flight time will have a significant impact on the company's innovation capability and patent filing situation within 1-2 years. At the same time, the company on the direct route also has an advantage in field supervision costs. When conducting investigation evaluation, you can put whether there is a direct flight in destination into consideration.

6 Synthesis and Suggestion for Further Study

By learning this course and completing the group project, we have a basic understanding of the theory and application of artificial intelligence. In this project, we use python to perform face recognition, voice recognition and text crawling on different information carriers and sources related to Sensetime and Megvii, one of his major competitors.

Then we perform natural language processing on all the materials collected. Through statistics of high frequency words, auto-generated topics and sentiment analysis and comparing these results with Megvii's, we come to conclusion that Sensetime is developing smoothly and worth of investment. Although it is not as promising as before according to the result of sentiment analysis, we are still optimistic about SenseTime's future.

We have four suggestions for future study.

1. Further learning of programming languages, especially Python is essential to solve some technical problems, such as maximum duration limit in voice recognition. We also need to explore, learn and apply some more powerful and advanced NLP packages such as StanfordNLP in the future.
2. We need to have a deeper understanding of AI technology especially latest technology such as AGI to form an intuition about the future trend of the AI industry.
3. Some creative sources of information, such as the research team's background and patents of the company, deserve more attention and analysis. (See figure 13 from MathSciNet).
4. Some updated news may be considered as a supplement to the report and the subsequent influence is worthy of attention (See figure 14). US weighs blacklisting up to five Chinese surveillance firms on May 23. Sensetime and Megvii are both on the list, which shows their challenge to US's technology hegemony. Although the short-term negative impact of the Sino-US trade war is inevitable, the US sanctions against Chinese high-tech companies such as Sensetime will further stimulate or even force Chinese companies to continue to develop new technologies independently, which is helpful in enhancing their core competitiveness. We wish them all the best to have a bright future.

Figure 13: Collaborators of Prof. Tang

Figure 14: Latest News of SenseTime

Acknowledgments

Thanks Anthony and Professor Yao for their fully devotion to the course this semester.

References

[1] Radford, A., Metz, L., & Chintala, S. (2015) Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434.*

[2] Xingjian, S. H. I., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C. (2015) Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems* (pp. 802-810).

[3] Bernstein, S., Giroud, X., & Townsend, R. R. (2016) The impact of venture capital monitoring. *The Journal of Finance*, **71**(4): 1591-1622.

Appendix A: Individual Efforts on the Project

Table 1: Individual Contribution

Name	Student ID	Contribution
Ruiting CHEN	20568400	Section 2.1, News Crawling & Presentation
Di CHENG	20454037	Section 2.3, 5, Video Editing & Presentation
Yang XUE	20565862	Section 1, 3.1, 3.2, 3.3, 6 & Presentation
Yifan YE	20551471	Section 3.1, 3.2, 3.3, 4 & Latex Typesetting
Mengkai ZHANG	20563711	Section 2.2, 3.4, 3.5, Video Recording & Presentation

Appendix B: GitHub URL

`https://github.com/gigondor/HKUST-MAFS-6010U-HAWK`

Appendix C: YouTube URL

`https://youtu.be/dnuZeVYy988`