

# A Study of Company — Sense Time

**LI Huiwen**  
Team Leader  
20573106  
LinkedIn Scraping

**FENG Yuan**  
Team Member  
20566684  
YC Approach

**ZHOU Yunzhi**  
Team Member  
20550063  
Competitors Analysis

**ZHAO Jingru**  
Team Member  
20568204  
PPT

**LEI Qiuyue**  
Team Member  
20552152  
Academic Literature Reviews  
Synthesis

## Abstract

In this project research, we research an AI company that mainly focuses on computer version area. We evaluate it by devising several AI/ML methods and traditional methods.

## 1 Introduction

Sense Time (Chinese: 商汤科技) is an artificial intelligence (AI) company. It was founded in Hong Kong with additional offices across China, Singapore, Japan, and the United States. The company has one of the highest total financing and highest valuations in the industry. It is startup company valued at over \$1 billion, these are sometimes classified as unicorns.

## 2 Technical Analysis

### 2.1 TYC Approach

#### 2.1.1 Introduction

Tian Yan Cha is a 'commercial survey tool that can be used by everyone'. It can help realize over 40 kinds of data query, such as enterprise information, enterprise development, legal risk, management risk, operating conditions, the intellectual property rights (enterprise business information, legal proceedings, the court announced patent, trademark, outward investment, branches, change information, bond, web site record, copyright, bidding, broken, abnormal operation, corporate annual reports, recruitment and news, etc.). It can also implement the in-depth commercial relationship combing, professional credit report rendering, and some other functions. It is an online tool that is suitable for finance, investment, lawyer, consulting, journalist, business etc.

#### 2.1.2 Automatic update

We mainly tried two methods to update the information of Sense Time collected by Tian Yan Cha, TYC to be abbreviated. Because TYC has anti-webcrawler technology, we almost cannot directly derive report from TYC, so we will list the methods we have considered.

Web crawler:

```
#!/usr/bin/env python
# sample usage: checksites.py eriwen.com nixtutor.com yoursite.org

import pickle, os, sys, logging
from httplib import HTTPConnection, socket
from smtplib import SMTP

def email_alert(message, status):
    fromaddr = 'you@gmail.com'
    toaddrs = 'yourphone@txt.att.net'

    server = SMTP('smtp.gmail.com:587')
    server.starttls()
    server.login('you', 'password')
    server.sendmail(fromaddr, toaddrs, 'Subject: %s\r\n%s' % (status, message))
    server.quit()

def get_site_status(url):
    response = get_response(url)
    try:
        if getattr(response, 'status') == 200:
            return 'up'
    except AttributeError:
        pass
    return 'down'
```

43  
44

Figure 1: Part of web crawling code

-f is down  
/root/.local/share/jupyter/runtime/kernel-23a5a42b-0255-4b14-a567-1674aca3d34b.json is down

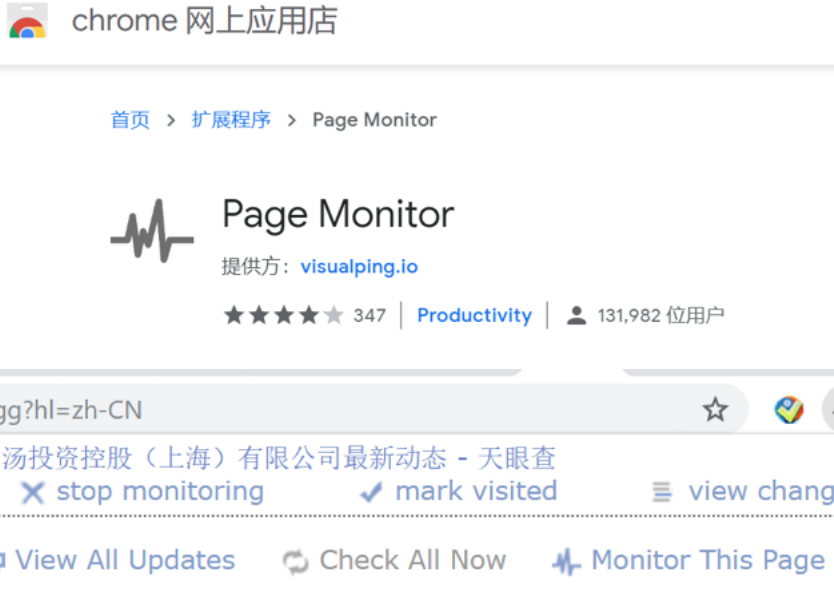
45  
46

Figure 2: Output

We can attempt webscrawler but the information in TYC needs accounts and verifying code, so we quit this approach.

47  
48  
49  
50  
51

Chrome web extension:



52  
53

Figure 3: Chrome page monitor

Using this chrome extension, we can easily get notifications of Sense Time. But the report will still be manually downloaded only.

54  
55  
56  
57  
58 **2.1.3 Main features**

Mainly providing professional enterprise information inquiry, enterprise relationship mining & digging services. Tian Yan Cha provides the users with a search query function, and the main information includes: industrial and commercial information, lawsuits, patent, trademark, dishonest information, enterprise change and enterprise annual reports, as well as

59  
60  
61  
62

63 the business associations query, etc. It has realized the integration of full volume of industrial  
 64 and commercial data, trademark data, public litigation data, and in-depth exploration of  
 65 corporate relations.  
 66

企业名称	商汤投资控股（上海）有限公司
二商注册号	310113000760618
统一社会信用代码	91310113685537782C
法定代表人	汤无杰
组织机构代码	685537782
企业类型	有限责任公司(自然人投资或控股)
所属行业	商务服务业
经营状态	存续
注册资本	5000万人民币
注册时间	23-03-09
注册地址	上海市宝山区牡丹江路1508号2217-G室
营业期限	2009-03-23 至 2029-03-22
经营范围	实业投资（除股权投资和股权投资管理）；企业投资管理（除股权投资和股权投资管理）；投资咨询；市场信息调查与咨询（不得从事社会调查、社会调研、民意调查、民意测验）；企业形象策划。（依法须经批准的项目，经相关部门批准后方可开展经营活动）

67  
68

Figure 4: Some examples



69  
70

Figure 5: Main relationship with Sense Time

71

## 72 2.2 LinkedIn Scraping

73

### 74 2.2.1 Objective

75

1) collect the number of employees in the interested company to track the change in the scale of the company.

76

77

2) From the background of employees in the company, track the potential trend of the company development.

78

79

80

### 2.2.2 Progress

81 We have developed 3 parts of the script, one for scraping the information from LinkedIn, the  
82 second to do data mining and analysis from the information been scraped from the web, and  
83 the 3rd for tracking the trend. The scripts can be scheduled to run monthly or quarterly to  
84 record and track the companies.

85 Data scraping:

86 selenium library is used for the scraping as LinkedIn has various anti-scraping methods built  
87 in the webpage while selenium is simulating human actions to browse through the website,  
88 where we don't see major issues scraping the data. By searching for the company name, we  
89 obtain a list of people who are connected to the company. Scanning through the list of people,  
90 we scrape the basic information (name, position, company, university) of each one.

91 Data mining & analysis:

92 1) tracking the number of employees by counting for people with company as the interested  
93 company

94 2) track employee's positions in the company to understand the structure of the company.  
95 Monitor the potential trend or focus of the company development by tracking the number of  
96 people under different positions (departments). NTLK library has been used to cluster the  
97 positions as there could be variance of wordings describing the same position since it's  
98 entered manually by each person. Using the natural language processing library, we are able  
99 to cluster them as the same category.

100 Trend Tracking:

101 Combined with the historical results, track the changes on the company scale as well as  
102 structure.

103

## 104 **2.2.2 Learning**

105 Through the process, we have experimented on different scraping technics and able to  
106 identify the most suitable method for our purpose. Also, we've studied and applied the basic  
107 natural language processing tools on text classification. The result from the LinkedIn  
108 scraping process is not very significant yet given the short time period we've been tracking.  
109 However, we believe that as time goes by and records get accumulated, the result would  
110 serve a meaningful part in our monitoring and tracking process.

111

## 112 **2.3 Competitors**

113

### 114 **2.3.1 Market Trends of AI**

115 A company rarely competes against just one competitor. In fact, in many cases, the biggest  
116 competition in our technology industries is coming from indirect competitors. These  
117 competitors hold a commanding position in their core market, allowing them to expand into  
118 different industries and verticals. Who would have thought that BAT would become die-hard  
119 competitors in the Artificial Intelligence market? It is almost impossible to distinguish direct  
120 and indirect competitors. In many cases everyone competes with everyone.

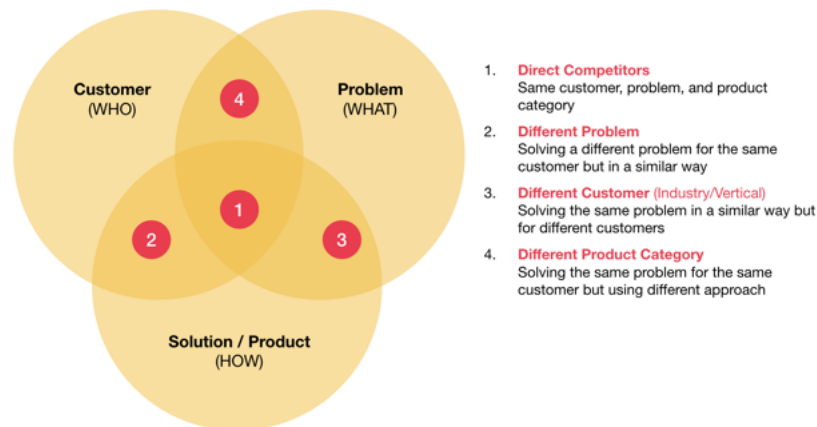
121

### 122 **2.3.2 Identify Competitors**

123 we all know that the competitors you select determines how you will perceive your company  
124 and the continue analysis. Firstly, we have to identify our competitors from two perspectives,  
125 direct competitors and indirect competitors.

126 Direct competitors are companies that sell to the same customers and solve the same problem  
127 using the same or similar solution (technology). The diagram below shows how Customer,  
128 Problem and Solution overlap into direct competition.

129



130

131

Figure 6: Methodology

132 As result, We identify direct competitors of Sense Time as Megvii, Cloudwalk and  
 133 Hikivision.

134 Limiting the organization to direct competitors only might lead us to a very narrow view of  
 135 the market. As we mentioned before, there are monopoly enterprise in technology also puts  
 136 their efforts on developing Artificial Intelligence. For instance, Baidu which is Chinese  
 137 biggest search engine established their AI platform in early 2017.

138 So we may identify our indirect competitors as the giant technology companies in China like  
 139 BAT(Baidu, Alibaba, Tencent)

140

141 **2.3.3 Evaluate opportunities and analyze threats**

142 After identify our competitors, we tend to contribute an auto-updating system to help us  
 143 refresh the data and comparison. We conducted a framework to compare Sense Time and its  
 144 competitors from the following aspects.

145

		Competitor 1	Competitor 2	Competitor 3
Company Specific	# of employees			
	Founded			
	Funding			
	Investors			
	Acquisitions			
	# of customers			
Target Customer / Message	Strengths / Weaknesses			
	Product			
	Primary Buyer / decision-maker			
	Secondary Buyer			
	Target Customer			
Product Specific	Messaging			
	Product Features			
	Pricing			
	Free Tier (?)			
	Customers			
	Product Strength			
	Product Weakness			
Positioning	Customer Reviews			
	How to Win			
	Why Customer should chose us			

146

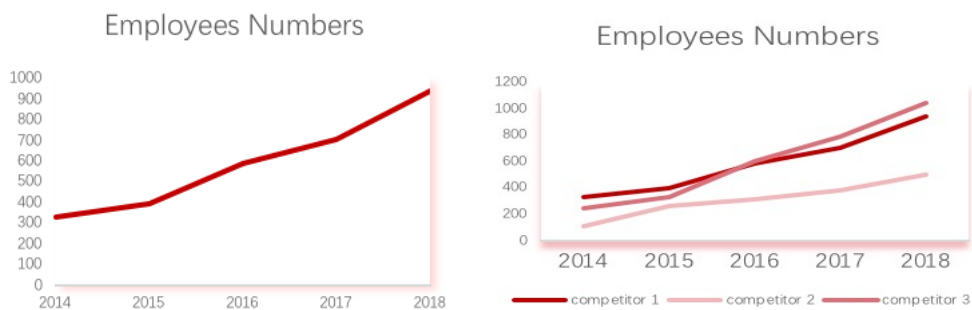
147

Figure 7: Framework

148 We manage to create an auto-updating framework in this form by comparing SenseTime with  
 149 its competitors from company specific, target customer, product specific and positioning.

150 Our data source are mainly from the data we collected through Tian Yancha or crawling  
 151 linkedin and databases such as Wind or Bloomberg. In the process, the main purpose is to  
 152 updating the data and visualize the data. In order to present the data more clearer and  
 153 visualize the data's degree of changing by plotting the dynamic time series. We try to specify  
 154 the framework by the timeline. For instance, when the manager would like to see the change  
 155 of employees number, he can just click the corresponding table and he will see a line chart as  
 156 followed. Further more, he can see the comparison situation between enterprises by click the  
 157 row ' of employees':

158



159

160

Figure 8: Employees number

161

162

### 163 **3 Article Reflection**

164

#### 165 **3.1 Universal Language Model Fine-tuning for Text Classification**

166 As we know, Inductive transfer learning (Pre-training on ImageNet) has greatly impacted on  
167 computer vision(CV)。 However, in NLP, transfer learning has not been as successful. In  
168 2018, one of the contenders for pre-trained natural language models is the Universal  
169 Language Model Fine-tuning for Text Classification, or ULMFiT, which was proposed by  
170 fast.ai’s Jeremy Howard and NUI Galway Insight Center’s Sebastian Ruder.

171 ULMFiT builds on similar approaches (CoVe, ELMo) and methods (Merity 2017). In CoVe  
172 and ELMo the encoder layers are frozen. ULMFiT instead describes a way to train all layers,  
173 and does so without overfitting or running into “catastrophic forgetting”, which has been more  
174 of a problem for NLP transfer learning in part because NLP models tend to be relatively  
175 shallow.

176 The network is trained in 3 phases: the first phase is unsupervised pre-training. The second  
177 phase is target task LM fine-tuning. The keys to making this work are a couple of simple  
178 ideas: discriminative fine-tuning and slanted triangular learning rates. The third phase is  
179 classifier fine-tuning.

180 This paper promises such a universal solution—universal in the sense that we’d like a single  
181 architecture and training method, minimal hyperparameter tuning, minimal pre-processing  
182 requirements. And We have found that the approach works well on different tasks from  
183 different researchers all around the world.

184 Besides text classification, there are many other important NLP problems, such as sequence  
185 tagging or natural language generation, that we hope ULMFiT will make easier to tackle in  
186 the future. In computer vision the success of transfer learning and availability of pre-trained  
187 Imagenet models has transformed the field. Many people including entrepreneurs, scientists,  
188 and engineers are now using fine-tuned Imagenet models to solve important problems  
189 involving computer vision—everything from improving crop yields in Africa to building  
190 robots that sort lego bricks. Now that the same tools are available for processing natural  
191 language, we hope to see the same explosion of applications in this field too.

192

#### 193 **3.2 FinTech Platforms and Strategy**

194 Fintech has taken the finance and banking sector by storm, especially as a response to 2008  
195 financial crisis. This research paper provides a taxonomy of platforms in finance and  
196 identifies the feasible strategies that are available to incumbents in the industry, innovators,  
197 and the major Internet giants.

198 In this article, it examines the potential trajectories and impacts of FinTech innovation on  
199 incumbent and new business models in the finance industry. And also provides a framework  
200 for understanding the value created through various types of platforms in financial services.  
201 This framework provides a natural mechanism for thinking about FinTech winners and losers  
202 and for predicting the trajectory of changes in the industry. It also provides a description of  
203 the possible strategies that innovators, incumbents, and the currently dominant Internet  
204 players can pursue.

205

### 206 **4 Suggestions for further study**

207 In addition to what we have devised, there are plenty of Artificial Intelligence & Machine  
208 Learning techniques which could be used to track and monitor the potential investing  
209 company in order to get a thoroughly understanding about its profile and a clear insight of  
210 the investment opportunity. For example, using NLP to process the interviews and speeches  
211 videos which are given by the key manger and sort the useful information which is related to  
212 the development and important decisions.

213 Also, data cleaning and data analysis are of vital important. There are many new techniques  
214 and researches in these areas and challenges as well. How to dig out key words, clustering,  
215 emotion analysis, etc.

216 All in all, there is a lot of space that we could improve and advance in the future.