Business Case

As the advance of machine learning technology became a growing trend in computer science. The implementation of sentiment analysis is easy to achieve nowadays. As thousands of information is generated on social media every second, the sentiment analysis made it possible for businesses to understand their customers' opinions and feelings more than ever. Therefore, a business could improve products and services more effectively based on the analysis results. Moreover, the algorithm of sentiment analysis will be more and more accurate and efficient along with each execution.

What

Getting to know people's feeling is important for businesses, politicians, and some other potentials since people could express themselves more openly than ever through the Internet. Thereby, we aim to develop a sentiment analysis application named 'News tone', which could identify the polarity of given text material. It could distinguish whether an expressed opinion or position in a given content is positive or negative. These expressions not only include customer feedback, but also could be social media conversations, news reports, and many other types of content. Sentiment analysis could automatically classify information in different types of intent.

All processes mentioned above are not conducted by human but machine learning technique – a computer technology which has been improved considerably in recent years. Human beings can instinctively identify the polarity of a word, a clause, a sentence, or a paragraph but a machine cannot. A deep learning algorithm will firstly be trained with an infinite number of materials. The purpose of this is to ensure the machine could extract the subjective information and identify the polarity of a certain material correctly just as human do. Then, the analyzing algorithm can be applied again and again to analyze the information it received. By analyzing materials, the deep learning algorithm will constantly improve the performance of itself time by time to become more accurate and efficient. Afterwards, the outcome of sentiment analysis would be addressed clearly and explicitly in a visual form.
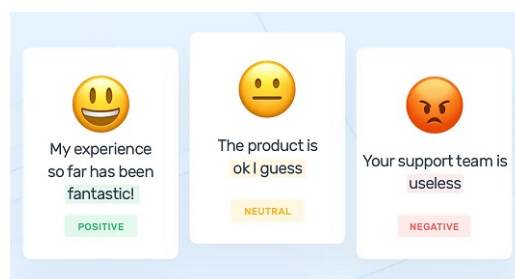

Figure 1 A sentiment analysis example

Why

With the rise of the Internet and the emergence of social media, people could easily express their attitude and opinion through the network at any time. A great deal of information is keeping generated for a certain topic with time elapsing. Although, it is a vital channel to connect to people's mind to understand their feelings and opinions. A huge amount of information is still a new breed of a challenge for the human to process and analysis. However, based on the advances of deep learning techniques improved considerably within recent years. Contextual mining becomes a possible solution to analyse sentiment among this information. Sentiment analysis could identify and extract subjective information from the source materials.

For instance, the figure shows below is a Twitter user who complained about the high price of Uber. It is easily identified by a human. However, there are thousands of these types of information generated and spread on the Internet. A human can't process and identify. However, these types of information are vital for Uber due to the information relates to their service. How could Uber extract all the relevant information on Twitter to understand their customer and improve their service? The sentiment analysis seems to be the possible solution to handle the situation like this.

Figure 2 Analyzing the intent of textual content

Goals

We aim to create a sentiment analyzing program which identifies the polarity of certain text material from 'news.com.au' and delivers the outcome precisely and clearly. The objective is to extract the subjective information from the input material and identifies the polarity automatically, whether it is positive or negative. The result of the sentiment analysis will be compared with human's identifications for the same information to assess the program.

Draft Plan

Iteration 1

The first iteration milestone is to develop a practicable web crawler. We will rely on this web crawler to grab past one-month front page titles of the news website – news.com.cn. To achieve this goal, 6 breakdown activities need to be conducted.

1. Searching for a suitable web crawler frameworkThere are many existing open-source web crawlers on the internet, but we need to pick one and adjust it to our project requirement.
2. Testing the web crawler framework
   After making some changes about the web crawler, we need to test the picked web crawler and make sure it works and suitable for our project requirement.
3. Crawling front page titles of the day on news.com.cn
   Our first step is just crawling only one front page of the day. We will look into the crawled data, and make sure the data is suitable for conducting sentiment analysis.
4. Crawling titles of the day on news.com.cn through Wayback Machine To access the past pages of the news website – news.com.cn, we have to access them through an archive website – Wayback Machine. Therefore, we will adjust our web crawler to obtain the news titles indirectly through the Wayback Machine.
5. Crawling past one-month titles through Wayback Machine
   We will extend our web crawler to obtain the past one-month news titles indirectly through Wayback Machine. At the next stage, we will again extend it to obtain the past 10-years news titles, which is the final goal of our web crawler.
6. Development testing – the web crawler
   We will test the web crawler and make sure that we can obtain the past one-month news titles correctly.

The schedule of the whole project plan (including the first iteration) is illustrated in Appendix 1.

Team Organization

Roles & Responsibility

For the first iteration, Ying Wang will act as the project manager, and also the developer. Yiru Li will act as the tester. Jin Zhou will act as the QA. Tasks assigning is shown below:

Tasks	Assignee
Search for a Suitable Web Crawler Framework	Ying Wang
Test the Web Crawler Framework	Ying Wang
Crawl Titles of the Day on news.com.cn	Ying Wang
Crawl Titles of the Day through WayBack Machine	Ying Wang
Create a Web Crawler which can crawl Titles in past 1 month	Ying Wang
Test the Crawler	Yiru Li
B. Case & Draft Plan	Ying Wang, Yiru Li
QA	Jin Zhou

Communication Plan - Group Meeting

In response to the COVI-19, we have updated our plan and we will hold the weekly group meeting online. As we are all familiar with WeChat, we will contact each via WeChat and hold an online video conference. The weekly group meeting will be held on each Monday morning, and the exact time will be negotiated in advance.

Each team member will take turns to prepare for the weekly group meeting. The preparation work includes preparing the agenda for the weekly group meeting, recording conference content during the meeting and recording the minutes. The responsibility schedule is shown below:

Week No	Assignee
W03	Ying Wang
W04	Ying Wang
W05	Yiru Li
W06	Yiru Li
W07	Jin Zhou
W08	Jin Zhou
W09	Yiru Li
W10	Yiru Li
W11	Jin Zhou
W12	Jin Zhou

Agendas will be uploaded to GitHub in advance, and the minutes (word file and pdf file) will be uploaded before the end of the week. Assigned tasks will be taken by corresponding members at the end of the meeting.

The project code will be organized on GitHub.

Communication Plan - Client Meeting

We plan to communicate with the client once a week, but we may not be able to do so. We know that the client may be busy. If we encounter some development-related difficulties or ambiguous product requirements, we will contact the supervisor/client according to the situation.

Appendix 1

MCI Project 2020 Team #23 - Timetable

Project Title: **News Tone**

No.		Activities	Week5 (M T W T F)	Week6 (M T W T F)	Week7 (M T W T F)	Week8 (M T W T F)	Week9 (M T W T F)	Week10 (M T W T F)	Week11 (M T W T F)	Week12 (M T W T F)	Week13 (M T W T F)
1	Milestone1	Search for a Suitable Web Crawler Framework	███ ·								
2		Test the Web Crawler Framework	· · ·█·								
3		Crawl Titles of the Day on news.com.cn	· · · ·█								
4		Crawl Titles of the Day through WayBack Machine		██ · · ·							
5		Create a Web Crawler which can crawl Titles in past 1 month		· ·██·							
6		Test the Crawler		· · · ·█							
7		B. Case & Draft Plan	· ·███								
8	Final Milestone	Milestone_1 Report			· · ·██						
9		Extend the Crawler to Crawl Titles in past 10 years			███ · ·						
10		Test the Enhanced Crawler			· · · ·█	█· · · ·					
11		Unit Test - the Title Crawler				· ·███					
12		Final Milstone Plan				█████ (orange)					
13		Testing Plan				█████ (orange)					
14		Search for a Suitable Sentiment Analysis (SA) Framework				· · · ·█ (orange)					
15		Test the SA Framework					██ · · · (orange)				
16		Process the Raw Data (Crawled Titles) -> Input Data					· ·██· (orange)				
17		Adjust the SA Framework to Fit Our Project					· · ·██ (orange)				
18		Test the Adjusted SA Framework					· · · ·█ (orange)	█· · · ·			
19		Apply Sentiment Analysis to the Input Data						██ · · · (orange)			
20		Check the Analysis Results						· ·██· (orange)			
21		Adjust the SA Framework -> SA Program						· · ·██ (orange)			
22		Analysis the Input Data with SA Program						· · · ·█ (orange)	█· · · ·		
23		Organise the Analysis Results							██ · · · (orange)		
24		Unit Test - the SA Program							· ·███ (orange)		
25		Create a Dashboard (Display the Analysis Results)						███ · · (green)			
26		Display Analysis Outcomes on the Dashboard							██ · · · (green)		
27		Improve the Dashboard Performance							· · ·██ (green)		
28		Unit Test - Dashboard								█· · · · (green)	
29		User Testing								· ██ · · (green)	
30		Release Testing								· · ███ (green)	
31	Delivery	Final Report							·████ (green)	█████ (green)	
32		Final Presentation Slides								█████ (green)	
33		Final Presentation and Demo Rehearsal									█████ (green)

Legend:

■ (blue) Crawl all homepage titles on news.com.cn in past 10 years

■ (orange) Apply sentiment analysis to the crawled titles

■ (green) Display analysis outcomes & Delivery