Machine Learning in FMCG: How Social Listening, Big Data, and AI Apply to Business Decision Process

Peter ChunYu YAU

Department of Computing Science, University of Glasgow

172 Ang Mo Kio Avenue 8, #04-01, SIT@NYP Building, Singapore 567739

ABSTRACT

Without a machine, human beings cannot process the massive data volume produced by us nowadays for business analysis purposes. Being part of the critical decision-making process in the commercial sector, the usage and role of Big Data, especially those collected from social listening are getting important: how a company listen to the true voices from a group of genuine customers and avoid those misleading, fake comments generated by the robot, to enhance the efficiency on measuring competitor's performance towards their marketing return compared to the investment (MROI), is critical and more challenging. In this paper, we will discuss various analytical methodologies from the angles of business and technological viewpoint: what decision-makers want from social media, what kind of information are they looking into, how visualization provides instant insights and what kind of system design provides enough interaction to the decision-maker. We will investigate the real-world challenges and difficulties: how the data inconsistency affects the computational analysis, what kind of obstacles was facing in the existing product design in the market. Finally, we will conclude the used machine learning techniques that can address the mentioned business problems, including the successful rate, accuracy and various efficiency level in the studied samples.

Keywords: Machine Learning, Artificial Intelligence, Big Data, Social Listening, FMCG, Business Process, Decision Making, Visualization

INTRODUCTION

The advancement of technology such as those fast-processing chips, and the new design of programming language (Rozman & Fellhoelter, 1995; Solomonoff, 2006), made certain kind of computer science topics largely moving forward in the last decade. These topics did not progressive much since their first date of academia setting discussion, due to the deficiency and limitation of the tools available for the research use (McKendrick, 2018). Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning (DL) are some of the good examples of these topics, they are now much advanced and greatly improved, benefited by the advanced design of hardware and software available in the market (Norrie et al., 2021).

In this paper, we will discuss how a business decision process is facilitated by using artificial intelligence, in the Fast-Moving Consumer Goods (FMCG) industry. The industry made their competitor analysis by understanding insights provided by the contextual analysis which those data were obtaining by social listening, and the subscribed data from the e-commence providers. Data volume is huge and cannot be processed by human beings in a reasonable amount of time (Simon, 2013). Commercial difficulty on how technology can be used to resolve the problems into the business environment will also be discussed, to give a useful business suggestion to the readers.

BUSINESS SCENARIO

In this case study, a conglomerate company is trying to use technology to enhance their daily analysis workflow. Under the big data era, huge data volume produced daily in an unstructured manner made the analysis job hard to perform (Tanwar, Duggal & Khatri, 2015), e.g., data source came from multiple providers, their input format is always different: some of them are structured file but not all the required data exists in one place, while the missing part can be found from other source but not in an easy readable format, such as the useful information is in an image format rather than a typed text.

Above-mentioned situation is very common to be found in workplace, which lots of human efforts are required to sort, filter, organize before useful information can be obtained: “This is the price of getting free information from the Internet”, “Yes, they are unstructured, but that’s what I need to do my analysis job”, said by the marketing man-ager who need to do their marketing plan base on the number and figures related to their products, and their competitors.

One million dollars, go or not go?

Quarterly, branding manager required to setup the goal and direction about how their managed brands should be promoted in different market; they need to decide how much resources (i.e., investment) should be placed in a particular market.

“You need to know how much your competitor invested before you make your judgment.”

“Why? If they already placed 1 million dollars in the market and received a thousand response, I am not going to invest in the same channel because the return is not justified”

“Yeah, and I don’t have this budget amount.”

“Yes, we need to use it wisely. That’s why what kind of information is available to us is very important”

“Of course, we are talking about the useful data, not those confusing and inaccurate.”

Challenge and Difficulties

To ensure an accurate data being analysed, manager needed to digest all the data once before processing.

“It is still okay if we need to process thousand or ten thousand records manually, turn out you will find that you can also do it the same to a million records, just you are sacrificing the accuracy”.

Usually, products listed on the shopping site are sold in predefined packing: they are either sold in an individual pack, or sold in multiple quantity (e.g., buy one get one free), or it is sold in a bundle setting (e.g., a shampoo plus a shower gel plus a shaver cream) (Figure 1).

Figure 1. Usual (predefined) product packing

From the subscribed sales records data source, these kinds of descriptions are in-putted by many types of people, for example shop keeper, freelance sales assistant, or banding manager, etc. They are also written in many kinds of languages.

“Why don’t they use a standardize product name to promote?”

“Simple. How these people create an attractive topic title is fundamentally important to the sales volume.”

“That’s right. An eye-catching title usually wins. Lots of better sales opportunities.” (Table 1), “We will use Emoji, too □□□” (Ge & Gretzel, 2018).

Table 1: Sample topic title for the product description

Qty	Example
Solo item	Best Seller! Loped Good Effect Ageing Prevent Day Cream 200g
Multiple item	Glory Bright Skin Lift Double Action Night Cream SPF 80 / PA ** Skin Care - 500ml – Day Cream [For Dry Skin Looks Pretty] x2
Bundle	Hiya Amazing 7-in-1 Aging-Prevent + Life’s Good Cream 80g Moisturizer with SPF 30. Best Deal Ever-rrr!

COMPUTATIONAL WORKFLOW

As seen from the communication with the marketing professional, it is understood that an accurate data pool can improve a lot in their decision-making process. In this section, we will discussion the computational workflow in detail, how existing technology can assist in the problems above (Figure 2).

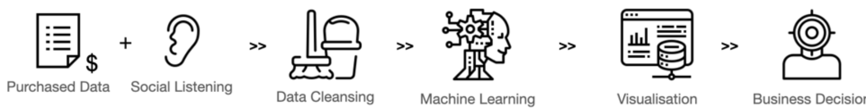

Figure 2. Computational workflow for automated (partial) business decision making

We will talk about how social listening (Killian & McManus, 2015) engage in the data collection process, it's technical constrain and the storage format for the further computation. We will discuss the general approach on how text analysis, sentimental analysis and classification can be used to understand how the topic title, same to the customer's comments are being presented in the online shopping site. Classification helps on identify what kind of product packaging is belongs to.

Purchased Data and Social Listening

For a conglomerates, data purchase for business use is very common. In most case, a paid service can deliver what the business professionals want, except certain scenario such as time limitation, communication barrier, and some strictly confidential commercial information process. Therefore, some of the companies will keep an internal software development team, to serve the internal request for such requests (or any other kind of technology needs).

Social listening is a way to monitor, capture, and understand certain kind of predefined keywords which end-users are interested into (Pomputius, 2019), they are usually targeted to the social media network such as Facebook, Instagram, MeWe, (Statista, 2021) etc. Definition become border and right now software, apps, tools, which previously considered another name, e.g., Instant Messenger (IM): WhatsApp, WeChat, Facebook Messenger, or Video streaming platform: YouTube, TikTok; may also considered as part of the social media network.

Social media platform may provide an Application Programming Interface (API) for third party's external development (Preibisch, 2018), e.g., Facebook for Developers (https://developers.facebook.com). It's Graph API provides a functionality on showing the relationship between object and object, say the relationship diagram between you and your friends' network. One of the practical uses is to check if the commenter is either a real person, or a robot; with the help of additional information such as profile activity (Ghosh, Surachawala & Lerman, 2011), and behaviour pattern (Conti, Poovendran & Secchiero, 2012).

Data Cleansing

Data obtained from the above, usually stored in an XML or JSON format for cross-device communication purpose (Wang, 2011), will then undergone a data cleansing step.

It is always a good discussion that when this step to be perform, and what it should be called. Some people will name it “Text Cleaning”, some people will keep it “Data Cleansing”, some people will prefer a more technical term such as “Normalizing Text” (Sproat & Jaitly, 2016). We argue that when it takes place and what it should be called is entirely up to the design of the computational process. We design this step to be more neutral, while always preserve the original data source for future use, or possible rollback purpose.

Text Normalization, Stop Words Removal, Stemming and Lemmatization are essential steps for text analysis. We perform normalization to convert capital case letters into small letters, here we ignore the consideration of sentimental as this is not our focus in this process, thru when sentiment analysis will be done in the next step, we could also re-move those Unicode characters and focus on “word form”.

Stop word is an interesting topic, the word list will evolve from time to time (Table 2) (Reporter, 2020), officer need to determine what kind of words should be appeared from the list as it will greatly affect the outputs (Wilbur & Sirotkin, 1992). Removal decision should be based on the goal what you are talking into.

Table 2: Sample stop words in different kind of subject area

Type	**Stop words**
Cosmetic	Clean, Beauty, Pretty, Amazing, Skincare
Drinks	Drink, Cheers, Bottle
Hygiene	Wash, Virus, Bacteria, Clean, Green, White

Stemming allows the machine grouping different kind of “word form” into the same type, e.g., for a cosmetic whitening product, stemming allows the program to understand “whitening”, and “white” are describing the same things. Lemmatization takes a similar approach but in terms of root definition in English language (Korenius, Laurikkala, Järvelin & Juhola, 2004).

Machine Learning

Introduction of no-code, and low-code programming (Harvard Business Review, 2021) now extended to the area of machine-learning. Predefined function calls and library allow an easy manipulation on structured data, that is what we have done in the previous steps. In this experiment, scikit-learn, a machine learning library is mainly used.

Key library and function calls used in the machine-learning automation process. References are mainly extracted from scikit-learn.

```
Import pandas pd
Import numpy as np
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import MultinomialNB
```

Sample code above showed part of the learning function we used, such as "CountVectorizer", a converter to convert text documents into a matrix (Kulkarni & Shivananda, 2021), and "MultinomialNB", a Naive Bayes classifier for multinomial models (Xu, Li & Wang, 2017). Technical detail will be skipped here; further code will be revealed at a later stage upon the beta version completed. We aimed to elaborate the business know-how in terms of the technical procedure in this writing, and ultimately open-source part of the code for research use.

BUSINESS DECISION, AND FURTHER STUDY

According to NielsenIQ (2021), there are about 30,000 new products launched each year. There are lots of the e-commence platform providing various kind of business data: manufacturer, pack-type, volume, quantity, grammage tier, demand space, benefit space, unit sold, listed price, revenue, market share, etc.; investment to the automatic text analytical process is criterial to the company success in the future.

In this case study, the branding and marketing manager experimented their raw data, with the machine learning pilot test program: aimed to automatically classify the package type from the big data source; this is the first step. The accuracy of this procedure is over 85%, we decided to move into next step to predict a totally brand-new item which the existing trained records do not have this knowledge: what is the demand space of a new product from a new competitor. This experiment is meaningful because it plays a preventive measurement when deploying with social listening, brand manager can be aware instantly if a new product directly competing to their items is launched to the market.

CONCLUSIONS

This paper discuss how machine learning is used in the business decision process. We discussed the concerns, difficulties, challenges and technical know-how on how artificial intelligence applied in the fast-moving consumer goods industry. Although certain in-depth know-how and procedure detail are skipped in this paper, we generally discuss the workflow how it can be made. Further disclosure will be made upon the completion of the product; we served this writing as a presentation about how technology can possibility change the daily work practice in the workplace which heavily rely on human interpretation. We concluded that the primarily experimental results are success and accepted to the invited business user.

ACKNOWLEDGMENTS

This conference fee is supported by the research team of University of Glasgow, Singapore. Thank you very much.

REFERENCES

Conti, M., Poovendran, R., & Secchiero, M. (2012, August). Fakebook: Detecting fake profiles in on-line social networks. In 2012 IEEE/ACM International Conference on Ad-vances in Social Networks Analysis and Mining (pp. 1071-1078). IEEE.

Ge, J., & Gretzel, U. (2018). Emoji rhetoric: a social media influencer perspective. Journal of marketing management, 34(15-16), 1272-1295.

Ghosh, R., Surachawala, T., & Lerman, K. (2011). Entropy-based classification of'retweeting'activity on twitter. arXiv preprint arXiv:1106.0346.

Killian, G., & McManus, K. (2015). A marketing communications approach for the digital era: Managerial guidelines for social media integration. Business horizons, 58(5), 539-549.

Korenius, T., Laurikkala, J., Järvelin, K., & Juhola, M. (2004, November). Stemming and lemmatization in the clustering of finnish text documents. In Proceedings of the thirteenth ACM international conference on Information and knowledge management (pp. 625-633).

Kulkarni, A., & Shivananda, A. (2021). Converting text to features. In Natural language processing recipes (pp. 63-106). Apress, Berkeley, CA.

McKendrick, J. (2018, December 19). How Fast Is Artificial Intelligence Growing? Look At The Key Bellwethers. Forbes. https://www.forbes.com/sites/joemckendrick/2018/12/19/how-fast-is-artificial-intelligence-growing-look-at-the-key-bellwethers/?sh=7642878a474a

NielsenIQ. (2021, May 19). Bursting with new products, there’s never been a better time for breakthrough innovation.

https://nielseniq.com/global/en/insights/analysis/2019/bursting-with-new-products-theres-never-been-a-better-time-for-breakthrough-innovation/

Norrie, T., Patil, N., Yoon, D. H., Kurian, G., Li, S., Laudon, J., ... & Patterson, D. (2021). The Design Process for Google's Training Chips: TPUv2 and TPUv3. IEEE Micro, 41(2), 56-63.

Pomputius, A. (2019). Can you hear me now? Social listening as a strategy for under-standing user needs. Medical reference services quarterly, 38(2), 181-186.

Preibisch, S. (2018). API Development. Apress.

Reporter, G. S. (2020, June 27). L'Oréal to remove words like "whitening" from skincare products. The Guardian. https://www.theguardian.com/world/2020/jun/27/loreal-to-remove-words-like-whitening-from-skincare-products

Rozman, A. F., & Fellhoelter, K. J. (1995, March). Circuit considerations for fast, sensitive, low-voltage loads in a distributed power system. In Proceedings of 1995 IEEE Applied Power Electronics Conference and Exposition-APEC'95 (Vol. 1, pp. 34-42). IEEE.

Simon, P. (2013). Too big to ignore: the business case for big data (Vol. 72). John Wiley & Sons.

Solomonoff, Ray J. "Machine learning-past and future." Dartmouth, NH, July (2006).

Sproat, R., & Jaitly, N. (2016). RNN approaches to text normalization: A challenge. arXiv preprint arXiv:1611.00068.

Statista. (2021, November 16). Global social networks ranked by number of users 2021. https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

Tanwar, M., Duggal, R., & Khatri, S. K. (2015, September). Unravelling unstructured data: A wealth of information in big data. In 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions) (pp. 1-6). IEEE.

Wang, G. (2011, April). Improving data transmission in web applications via the transla-tion between XML and JSON. In 2011 Third International Conference on Communica-tions and Mobile Computing (pp. 182-185). IEEE.

When Low-Code/No-Code Development Works — and When It Doesn't. (2021, June 22). Harvard Business Review. https://hbr.org/2021/06/when-low-code-no-code-development-works-and-when-it-doesnt

Wilbur, W. J., & Sirotkin, K. (1992). The automatic identification of stop words. Journal of information science, 18(1), 45-55.

Xu, S., Li, Y., & Wang, Z. (2017). Bayesian multinomial Naïve Bayes classifier to text classification. In Advanced multimedia and ubiquitous engineering (pp. 347-352). Spring-er, Singapore.