Pertanika J. Soc. Sci. & Hum. 24 (S): 169 - 176 (2016)

SOCIAL SCIENCES & HUMANITIES

Journal homepage: http://www.pertanika.upm.edu.my/

Towards Automatic Customer Purchase Behaviours Prediction through a Social Media Lens Using the Hidden Markov Model

Lasmy*, Chowanda, A., Herman, R. T. and Notoatmojo, B.

Management Program, School of Business Management, Bina Nusantara University, Jakarta 11530, Indonesia

ABSTRACT

In this research article, we present our work on building computational prediction models to dynamically predict users' purchase behaviours by implementing Hidden Markov Models (HMM). The models can be used by decision makers in a company to develop a strategy (e.g. marketing, products development) based on the prediction results. We evaluate the model using our datasets of Facebook. We collected the data by utilising Facebook API. Furthermore, we implemented a Hidden Markov Model (HMM) algorithm to the datasets to provide a dynamic prediction of customers' purchase behaviours over time. In the preliminary evaluation, we implemented our model to the datasets with t=2. In our datasets, we found that the category, electronics, was the most favourite topic to discuss, share and like regarding electronics. Interestingly, we found that a positive direction for its trend appeared in the second run of the model.

Keywords: Customer purchase behaviour, Hidden Markov Model, Facebook datasets, strategic management, computational prediction model

INTRODUCTION

In the era of social networks, more than 74% of Internet users are connected to social networking sites. Most of them (96%) are on Facebook (Pew Research Center, 2013). Social media have been used to interact and engage with people all over the world. Billions of topics, from politics to hobbies, are discussed in social media. People often share their interests, thoughts and ideas on social media. Others, who share the same interests, may post general comments or like a posting.

This phenomenon allows us to leverage on a plethora of information provided on

ARTICLE INFO

Article history:
Received: 19 November 2015
Accepted: 04 May 2016

E-mail address:
Lasmy@binus.edu (Lasmy)
* Corresponding author

ISSN: 0128-7702

social media to improve company strategy in marketing or product development. This is achievable by predicting users' purchase behaviour through a social media lens.

This paper provides the following main contributions:

- We present our work on building computational prediction models to dynamically predict users' purchase behaviours by implementing Hidden Markov Models (HMM). The models, hence, can be used by decision makers in a company to devise strategies (e.g. marketing, products development) based on the prediction results.
- We evaluated the models using our datasets of Facebook with 24 users, 568 posts and 126 likes. This provided preliminary results for the proposed models in this paper.

The rest of the paper is organided as follows: Relevant previous works are presented in the next section. Section 3 outlines our research methodology. In Section 4, we describe our computational prediction models. We present the evaluation results and discussion in Section 5. Finally, we conclude our work and describe future work in Section 6.

LITERATURE REVIEW

Research into customer behaviour provides data to be used for strategic managerial decisions in marketing (e.g. market segmentation, targeting and positioning), customer relationship management (e.g. customer value and satisfaction) and product development (Schiffman et al., 2012). This research provides an insight into customer behaviours in searching for, purchasing and using products and services. In this era of social networks and digital technology, customer behaviour has evolved to a whole new stage. Nowadays, customers can easily buy and sell products through the Internet. Over the decades since the introduction of the Internet, the total revenue of e-commerce has doubled from 15% (2002) to 30% (2012) (United Nations Conference on Trade and Development, 2015) and reached $1.5 trillion in 2014 (The Nielsen Company, 2014).

This new stage provides huge advantages to researchers of customer behaviours, as most of the data are available for public consumption. The only problem is that researchers will need to mine the data with an equitable model or technique to produce relevant information for decision makers within a reasonable amount of time. Several studies have been done in this area. Schafer et al. (1999) and Sarwar et al. (2001) suggested a model with collaborative filtering to filter the users who share similar characteristics among themselves. They argued that users with similar characteristics usually have or like similar products. Similarly, Catanese et al. (2012) proposed a model consisting of similarity detection and community detection to detect groups of related users over time. Those models are widely used to predict or recommend products purchased through e-commerce (Catanese et al., 2012).

With advances in information retrieval technology, content-based search methods applied to web pages, social media, customer reviews, search results etc. are now becoming popular. Certain information can be extracted from these media using machine learning methods. Zhang and Pennacchiotti (2013) utilised datasets of eBay users who have connected their eBay account to their Facebook account. Thus, they can extract their purchase history and their likes on Facebook to obtain a prediction of users' purchase behaviours. Similarly, Sen et al. (2009) utilised users' tags to predict movie preferences. They described the algorithms based on the users' interactions with tags and movies (Sen et al., 2009). The same method was implemented to provide a large set of words based on users' recommendations on Twitter (Netzer et al., 2008). The algorithm was intended to recommend interesting content and provide recommendations or filtering services to Twit users.

In this paper, we built a prediction model based on the Hidden Markov Model (HMM). An HMM model is a tool that represents probability of distributions in time series over sequences of observations with unobserved or hidden states. Although the HMM algorithm is mostly applied to solve computational problems (e.g. speech, handwriting and many other computer-vision problems), it is also a well-known technique to predict behaviour. Netzer et al. (2008) applied an HMM for customer relationship management models. The researchers constructed an HMM to relate latent relationship states to observed buying behaviour; eventually, they could evaluate customer relationships over time. Netzer et al. (2008) proposed a customer relationship model with HMM to build a customer portfolio. In addition, Kiseleva et al. (2013) predicted user intent with an HMM model to understand user intentions based on their browsing behaviours.

METHODOLOGY

We collected datasets from Facebook, as currently Facebook has one billion active users (http://newsroom.fb.com/company-info/). This made Facebook the most used social medium in the world. We collected data from the Facebook API. The data collected were extracted using keywords and the number of likes from posts. Subsequently, the extracted features were categorized into electronics, travelling, motors and fashion. In this paper, we minimised our categories into four general categories, as we aimed to evaluate our models first.

To extract the features, we first implemented a keyword search function in the Facebook API to spot some keywords we wanted to search for (e.g. name of the products, places etc.). The spotted keyword was consequently categorised (e.g. Android or hand phone was categorised under electronics, Bali or London was categorised into places). All the feature data were stored in a single text file.

The next step was to implement the Hidden Markov Model (HMM) algorithm to provide a dynamic prediction of customers' purchase behaviours. In this paper, however, we only provide a dynamic prediction of the trend of the categories. The details of the model are described in the next section.

THE MODELS

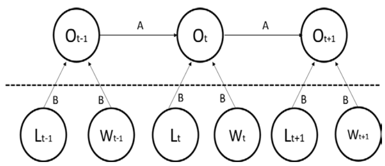

Figure 1: HMM models

Figure 1 presents our model of an HMM with states of categories at time t (0t), and two observations: Like at time t (Lt-1) and Word at time t (Wt-1). Using prior knowledge (t-1) of the model, we were able to determine the unobserved sequence of hidden states for the current state (t). The model definition is:

T = length, in this paper, we set T=2;

N = number of states in the model, where N=4;

M = number of observations in the model, where M=2;

Q = {q0, q1, ..., qN−1} = states of the Markov process

V = {0, 1, ..., M −1} = set of possible observations

A = state transition probabilities

B = observation probability matrix

π = initial state distribution

O = (O0, O1, ..., OT−1) = observation sequence.

Furthermore:

O = {Electronics, Travelling, Motor, Fashion}

$$Q_{i,j} = \begin{bmatrix} q_{0,0} & \cdots & q_{0,N-1} \\ \vdots & \ddots & \vdots \\ q_{3,N-1} & \cdots & q_{N-1,N-1} \end{bmatrix} \quad (1)$$

The state transition probabilities were defined by probability of state q at time t-1 given state q at time t occurred. Hence:

$$A = \{a_{ij}\}; where\ P\left(q_{i,t+1} \middle| q_{it}\right) \quad (2)$$

while the observation probability matrix was defined by probability of observation k at time t given state q at time t occurred. Hence:

$$B = \{b_{jk}\}; where\ P\left(\partial_{k,t} \middle| q_{jt}\right) \quad (3)$$

where the observed probability value could be obtained from the number of likes and spotted keywords in certain categories using the logarithm function as can be seen in (4) and (5).

$$\vartheta_{t,k,c} = \sum \frac{Likes_{k,c}}{max_{c} \in Likes_{k,c}} \times \log \frac{|k|}{|(k,u)|} \quad (4)$$

and

$$\vartheta_{t,k,c} = \sum \frac{Words_{k,c}}{max_{c} \in Words_{k,c}} \times \log \frac{|k|}{|(k,u)|} \quad (5)$$

$$L_t = \{L_0, L_1\} \quad (6)$$

$$W_t = \{W_0, W_1\} \quad (7)$$

$$\vartheta_t = \{\vartheta_0, \vartheta_1\} \tag{8}$$

Finally, from (6)-(8), the mathematical model of the probability of the state sequence was denoted by:

$$P(L) = \pi x_0 b x_0(\vartheta_0) a x_0, x_1 b x_1(\vartheta_1) \tag{9}$$

$$P(W) = \pi x_0 b x_0(\vartheta_0) a x_0, x_1 b x_1(\vartheta_1) \tag{10}$$

EXPERIMENTAL RESULTS

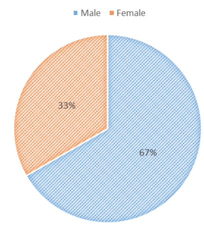

Figure 2: Demographic data

Figure 2 illustrates the demographics of our Facebook datasets. Most of the users were males (67%). In this preliminary evaluation we implemented our model to the datasets with t=2. Table 1 and Figure 3 describe the results of the models.

Table 1
HMM Implementation

Category	t = 1			t = 2		
	P(L)	P(W)	AVG	P(L)	P(W)	AVG
Electronics	0.42	0.39	0.405	0.32	0.23	0.275
Travelling	0.33	0.24	0.285	0.29	0.31	0.3
Motors	0.13	0.16	0.145	0.16	0.19	0.175
Fashion	0.12	0.21	0.165	0.23	0.27	0.25

As we can see from both Table 1 and Figure 3, the top category was electronics (0.42, 0.39, 0.32, 0.23). This indicates that there were quite a number of people who discussed, shared and liked topics or posts regarding electronics. Travelling was the second hot topic, with 0.33 and 0.24 for probability from likes and words the first time we ran the model and 0.29 and 0.31 the second time we ran the model.

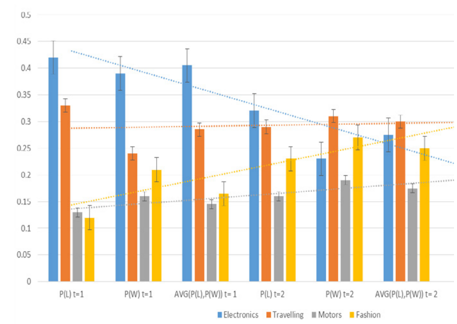

Figure 3: Results

From Figure 3 we can conclude that while electronics was the favourite topic on Facebook, there was an overall decrease trend (see the blue dot). Interestingly, fashion showed a positive direction (see the yellow dot). This data provides strategy makers with additional information to predict customer purchase behaviour based on trends obtained from social media. In this paper, Facebook was used as a study case for our model.

CONCLUSION AND FUTURE WORK

We proposed a computational prediction model to dynamically predict users' purchase behaviours by implementing Hidden Markov Models (HMM). Furthermore, we evaluated the models based on our datasets from Facebook, providing preliminary results for the proposed models explored in this paper.

There were some limitations encountered in this study. Firstly, the datasets were not considerable enough to represent global population. Most importantly, in this research we ran the model only twice, whereas it should have been run over several weeks or even months. Future research into this topic can go on to collect more data at least to represent a much wider, if not the entire population of Internet users. It should also include historical data of the users. Secondly, we need to expand the categories, and add more keyword databases for keyword search functions to enhance the accuracy of the feature extraction and classification.

REFERENCES

Catanese, S., De Meo, P., Ferrara, E., Fiumara, G., & Provetti, A. (2012). Extraction and analysis of Facebook friendship relations. In A. Abraham (Ed.), *Computational Social Networks* (pp. 291–324). London: Springer.

Kiseleva, J., Lam, H. T., Pechenizkiy, M., & Calders, T. (2013). Predicting current user intent with contextual Markov models. In *Proceedings of the 2013 IEEE 13th International Conference on Data Mining Workshops (ICDMW '13)*. Washington DC, USA: IEEE Computer Society.

Netzer, O., Lattin, J. M., & Srinivasan, V. (2008). A hidden Markov model of customer relationship dynamics. *Marketing Science, 27*(2), 185–204.

Pew Research Center: Internet, Science & Tech. (2013). *Social networking fact sheet*. [online] Retrieved 20 August, 2015 from http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web (WWW '01)* (pp. 285–295). NY, USA: ACM.

Schafer, J. B., Konstan, J., & Riedl, J. (1999). Recommender systems in e-commerce. In *Proceedings of the 1st ACM conference on Electronic commerce (EC '99)* (pp. 158–166). New York, NY, USA: ACM.

Schiffman, L., Kanuk, L., & Hansen, H. (2012). *Consumer behaviour*. Harlow, England: Pearson Financial Times/Prentice Hall.

Sen, S., Vig, J., & Riedl, J. (2009, April). Tagommenders: Connecting users to items through tags. In *Proceedings of the 18th international conference on World Wide Web* (pp. 671–680). ACM.

The Nielsen Company. (2014). *E-commerce. Evolution or revolution in the fast-moving consumer goods world*. [online] The Nielsen Company. Retrieved August 20, 2014 from http://ir.nielsen.com/files/doc_financials/Nielsen-Global-E-commerce-Report-August-2014.pdf

United Nations Conference on Trade and Development. (2015). *Information economy report 2015. Unlocking the potential of e-commerce for developing countries*. [online] New York and Geneva: UNITED NATIONS. Retrieved August 20, 2015, from http://unctad.org/en/PublicationsLibrary/ier2015_en.pdf

Zhang, Y., & Pennacchiotti, M. (2013). Predicting purchase behaviors from social media. In *Proceedings of the 22nd international conference on World Wide Web (WWW '13)* (pp. 1521-1532). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland.