Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

Using Machine Learning to Design the Match-Up System between Influencers and Products

[1] Yu-Chen Kuo*, [2] Jong-Yih Kuo, [3] Chen-Li Lin, [4] Pei-Chen Kuo
[1] Department of Computer Science and Information Management, Soochow University, Taiwan
[2][3] Department of Computer Science and Information Engineering, Taipei University of Technology, Taiwan
[4] Department of Computer Science and Engineering, Pennsylvania State University, USA
*Corresponding author Email: yckuo@scu.edu.tw

Article Info
Page Number: 1290-1303
Publication Issue:
Vol. 71 No. 4 (2022)

Article History
Article Received: 25 March 2022
Revised: 30 April 2022
Accepted: 15 June 2022
Publication: 19 August 2022

Abstract
Due to the mature of internet and the rise of the media in these years, online marketing has become a trend to increase the visibility. The traditional marketing methods have been replaced by social media gradually. With the sharing and recommendation of influencers on the social platform, a wider response than traditional marketing can be obtained. Therefore, this research will implement two systems, named the influencer crawler system and the match-up system between influencers and products. First, to obtain the information of influencers in social media through the influencer crawler system, use text mining technology to extract useful information from these social media interactive data, and then use machine learning to matchmaking influencers and products. We use questionnaires to verify and understand the public's expectations for the matchmaking of Influencers and products.

Keywords: Machine Learning, Web Crawler, Text Mining, Recommender System

I. INTRODUCTION

With the booming of social media nowadays, online marketing has become a trend to increase the visibility and reduce the physical stores and overheads. Social media has changed the way of social mode and message receiving for people, now you can get wider responses by sharing and recommending on internet rather than traditional marketing.

The interaction between people has become more diversified with the popularity of mobile devices and the development of social media which have broken the constraints of space of time, you can let your friends and fans know your first hand news by a post or a timeline explore, as long as you have a mobile device and reply on your attractiveness and influence in online community, you can promote products to public and achieve your marketing purpose.

Although there are lots of literatures on social media but only few of them discuss the influence by influencers in social media and the use of machine learning to analyze the relationship between influencers and products is even rarer. This research proposes a system that can analyzes the matchmaking between influencers and products by machine learning, analyzes the contents influencers posted on social media and make predictions from their endorsements, with the help of this system, suppliers can find more suitable endorsers and enhance the

marketing effectiveness.

There are five sections in this research. Section 2 is related work, explaining the relevant background knowledge of endorser's credibility, text mining, Likert scale, and recommender system. Sections 3 explains the system design and implementation. Section 4 introduces the experimental results. Section 5 explains the conclusions of this research.

II. RELATED WORKS

A. Endorser's Credibility

The endorser is often a reference for consumers to decide whether the advertisement information is credible, so the credibility of the endorsers is one of the important factors that affect the effectiveness of the advertisement.

The credibility of endorsers is a method that is often used in advertising to influence consumers' attitudes and purchase intentions [10]. Advertisers use endorsers to persuade consumers to buy the products advertised by the company, or increase consumers' attitudes towards the products through advertising, guide them to a positive understanding of products.

Before Ohanian [13] proposed, there was no consensus on the dimensions of endorser's credibility. Ohanian compiled the research of past scholars and concluded that three dimensions of endorser's credibility are defined as follows:

(1) Attractiveness: Consumers believe that advertising endorsers have external charms, special styles or characteristics for products or services which can attract consumers' attention and make consumers have positive impression on the products or services recommended by them.

(2) Trustworthiness: Consumers believe that advertising endorsers have the characteristics of honesty and integrity, and endorsers with high trustworthiness are often more persuasive.

(3) Expertise: Consumers believe whether the interaction between the endorsers and the products is professional, or whether the endorsers' knowledge on products have been certified or recognized.

Ohanian indicated in the research that expertise is more important than attractiveness and trustworthiness from consumers' buying tendency perspective. If the endorsers have rich professional knowledge and rich experiences on the use of the product, or weighty discussions about the products they endorsed, the advertising effect will be greatly improved. Consumers' trust and support can be obtained.

The attractiveness proposed in the study by Till et al. [15] learned that attractiveness is based on three factors: consumers' familiarity, likeability and closeness.

And in Friis-Jespersen's research [5] mentioned the factors related to the trustworthiness of endorsers which are attraction, expertise, homogeneity, strength of contact, etc. And it also shows that there is a direct and positive relationship between certain factors of endorsers' trustworthiness and advertising attitudes. In addition, the results of Hofstede et al. in this study

[6] show that celebrities' contributions to society, their views on issues of national importance and their contributions to disasters, as well as their humility and non-participation in disputes, will leave a good impression in the minds of consumers.

B. Text Mining

Text mining which purpose is to find useful and relevant information from a large amount of data through automatic or semi-automatic methods. Text mining is an extension of data mining, in data mining, most of them are based on structured data, while the data for text mining belongs to the type of text, and most of them are composed by human natural languages. The data reflect people's daily life, text mining is to extract meaningful information from these unstructured data, and systematically identify, integrate the hidden information behind the text data.

Text mining is mainly divided into three steps: text pre-processing, text processing and text analysis [1], first, in the stage of text pre-processing, the original data will be extracted and prepared which is used to find available data and exclude useless data, then calculate and integrate these pre-processed data, and finally perform statistical analysis on the integrated data to obtain meaningful information.

C. Recommender System

The recommendation system [14] is an information filtering system that can predict users' ratings and preferences for products. Its main purpose is to find the relationship between the user and the product to maximize the interaction between the user and the product. The following will introduce the matrix factorization methods for recommendation systems.

Matrix factorization [8] is to factorize the matrix, find out two or more matrices can get the original matrix when they are multiplied, matrix factorization can be used to find latent features, and these latent features are the interactions between two or more different types of entities.

Take a movie recommendation system as an example, there are a group of users and a group of movies, assuming each user has rated certain movies in the system, hope to predict how users will rate their unrated movies. Table 1 shows the rating matrix of users and movies. The values in the matrix are the ratings given by each user after watching the movie, and the question mark indicates that the user has not rated the movie. Therefore, what the matrix factorization needs to do is to predict these missing values through machine learning, and it is hoped that the predicted matrix is consistent with the rated fields in the original matrix.

Table 1. User-Movie rating matrix

Movie / User	Movie 1	Movie 2	Movie 3	Movie 4
User 1	1	2	?	5

User 2	3	5	?	?
User 3	?	3	4	3
User 4	5	?	?	?

Suppose there have a set U of users and a set M of movies, R is the rating matrix of users and movies, the size is $|U|\times|M|$, and then K latent features are to be found, so what we have to do is to find 2 matrices P and Q. The size of P is $|U|\times|K|$, and the size of Q is $|M|\times|K|$. After they are multiplied, the matrix $\hat{R}$ will be similar to the original matrix R, as (1) shown.

$$R \approx P \times Q^{\mathrm{T}} = \hat{R} \ldots (1)$$

Each row of P would represent the strength of the associations between the user and the latent feature. Similarly, each row of Q would represent the strength of the associations between the movie and the latent feature. First, give random initial values of the two matrices P and Q, and then use optimization method to minimize the error between the product of the P and Q matrices and the original matrix R. When calculating the error, only the data that originally existed in the training set T is taken, which means that there are data rated by the user in the original matrix R. The error calculation is shown in (2) to measure the error between the prediction value and the true value.

$$E = \sum_{(u_i,m_j,r_{ij})\in T} e_{ij}^2 = \sum_{(u_i,m_j,r_{ij})\in T} (r_{ij} - \sum_{k=1}^{K} p_{ik}q_{kj})^2 \ldots (2)$$

It can be used to avoid overfitting through regularization. This can be achieved by adding the regularization parameter λ in (2), and the error calculation is modified to (3):

$$E = \sum_{(u_i,m_j,r_{ij})\in T} \left(r_{ij} - \sum_{k=1}^{K} p_{ik}q_{kj} \right)^2 + \frac{\lambda}{2}\left(||P||^2 + ||Q||^2 \right) \ldots (3)$$

When a user is rating a movie, there may be other factors that affect his rate of the movie. For example, some users may tend to give a higher rating to the movie, while some may be more stringent and generally give a lower rating to the movie. So when making predictions, these possible biases are taken into account, and the way of predicting rating is shown in (4):

$$\widehat{r_{ij}} = b + bu_i + bm_j + \sum_{k=1}^{K} p_{ik}q_{kj} \ldots (4)$$

Where b is the global bias, the calculation is the mean of all ratings in the matrix, bu_i is the bias of user i, and bm_j is the bias of movie j. By adding the bias, the model can converge more quickly.

III. SYSTEM DESIGN AND IMPLEMENTATION

This section will introduce two systems. One is influencer crawler system and the other is

match-up system between influencers and products which will explain how to obtain influencers' social media data and how to match up between influencers and products.

A. Influencer Crawler System

The system is used to crawl the relevant data of the influencers, including the number of fans, the content of the posts, and the feedback of fans, etc., to analyze the credibility of the endorsement by the influencer through the crawled information and find out the previous promoting advertorials or posts.

Figure 1 is the flow chart of influencer crawler system, first of all, this research is based on the top 100 influencers in 2020 in Taiwan listed in the "Top 100 Influential Influencers Key Report" [17] by the digital age as the target influencers for study. Then collect the social platform account links of these 100 influencers, obtain the posts of the influencers through the links, and crawl these posts to obtain the content of the posts, the number of likes, comments, hashtags, etc., save them as excel files and repeat the above steps till the 100 influencers are crawled completely.

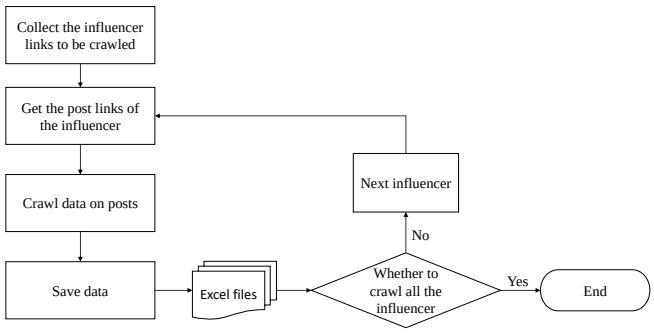

Figure 1. The flow chart of influencer crawler system

This system uses Selenium provided in Python for web crawling. Selenium is a framework for automated testing of web applications. It can automatically open the browser and simulate the operation of a user, and then filter out the required by parsing the HTML of the web page.

The social media referenced in this research is Instagram, and information about influencer posts on Instagram is obtained through this system.

B. Match-Up System Between Influencers and Products

The purpose of the system is to find out the suitable influencers to endorse products or advertorial and seek for the best adverting effect, make recommendations on influencers based on the product types, and the period of the influencers information adopted is from July 1, 2017 to July 31, 2020.

Figure 2 is the flow chart of match-up system between influencers and products, after obtaining the relevant data of the influencers through the above-mentioned crawler system, the original data is calculated and filtered to analyze the credibility of the influencers as endorsers, select the top 10 influencers with the endorsed credibility and find out the posts they have endorsed, use these posts to construct the rating matrix of the influencer and product category, and then

use the matrix factorization to recommend suitable influencers for the product.

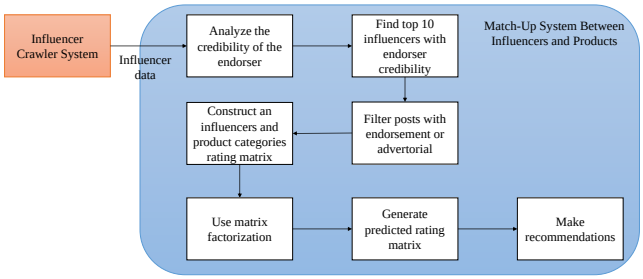

Figure 2. The flow chart of match-up system between influencers and products

C. *Design of Endorser's Credibility Factors*

First, it will summarize the three major dimensions that can represent the credibility of endorsers by Ohanain [13] and analyze the influencers. The three major dimensions are attractiveness, trustworthiness, expertise. Figure 3 is the hierarchy of endorser's credibility factors designed for this research includes three major dimensions and its sub-factors, in attractiveness, the influencing factors are familiarity, likeability, closeness and physical attraction, in trustworthiness, the influencing factors are attractiveness, expertise and tie strength, in expertise, the influencing factors are experience and knowledge, Analyze the credibility scores of the influencer endorsements by these three dimensions.

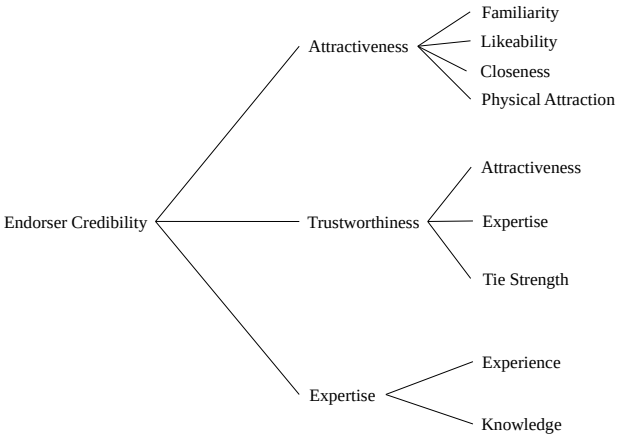

Figure 3. The hierarchy chart of endorser's credibility factors

Below will explain the calculation method of each sub-factor in attractiveness, familiarity is calculated based on the number of posts by influencers and the number of comments by fans to get the average number of posts of all influencers, then divide the number of posts of each influencer by the average to come out the score of each influencer's post, and then count the number of fan comments of each influencer's post, calculate the average number of fan comments of all influencers, then divide the number of fan comments of each influencer by the average to get the comment score, and finally average the post score and the comment score as (5) shown. Likeability is determined by tracking the number of people, calculate the average of all influencers, and then divide the value of each influencer to the average, as (6) shown. The closeness is determined based on the status of the influencers replied to the comments. The

average of the comment response rates of all the influencer is calculated, and then the ratio of each influencer is divided by the average, as (7) shown. Physical attraction is based on the content of fans' comments to see if there are keywords of the compliment about influencers' outlook, such like: handsome, pretty, lovely, etc., count the number of comments with keywords, calculate the average of all influencers, and then divide the number of comments containing keywords by each influencer by the average, as (8) shown.

$$\text{Familiarity} = \frac{1}{2}\left(\frac{\text{Number of posts}}{\text{Average number of posts}} + \frac{\text{Number of comments}}{\text{Average number of comments}}\right) \ldots (5)$$

$$\text{Likeability} = \frac{\text{Number of followers}}{\text{Average number of followers}} \ldots (6)$$

$$\text{Closeness} = \frac{\text{Reply rate}}{\text{Average reply rate}} \ldots (7)$$

$$\text{Physical Attraction} = \frac{\text{Number of comments with keywords}}{\text{Average number of comments with keywords}} \ldots (8)$$

The calculation method of connection strength in the trustworthiness is the average of familiarity in attractiveness and closeness in attractiveness, as (9) shown.

$$\text{Tie Strength} = \frac{\text{Familiarity} + \text{Closeness}}{2} \ldots (9)$$

In the expertise, the experience is calculated based on the time when the first post was posted by the influencer, Use the current time to subtract the time to post the first post to calculate the gap, calculate the average of the time gap of all influencers, and then divide the time gap of each influencer by the average, as (10) shown. The knowledge is to analyze whether the fans have learned something from the influencers by the content of the fans' comments, and grab the keywords, such as teaching, learning, expertise, etc., count the number of comments with keywords, calculate the average of all influencers, and then divide the number of comments with keywords by each influencer by the average, as (11) shown.

$$\text{Experience} = \frac{\text{Current time} - \text{Post time of the first post}}{\text{Average time gap}} \ldots (10)$$

$$\text{Knowledge} =$$

$$\frac{\text{Number of comments with keywords}}{\text{Average number of comments with keywords}} \dots (11)$$

The proportions of the four factors in the attractiveness are familiarity 30% , likeability 30%, closeness 30%, and physical attraction 10% as (12) shown, the reason why only 10% for physical attraction is because this part may be more subjective, so its proportion is reduced, and the proportions of the three factors in the trustworthiness are 1/3 for each, as (13) shown, the proportions of the two factors in the expertise are experience and 50% knowledge 50%, as (14) shown, while the proportions of three major dimensions are attractiveness 30%, trustworthiness 30%, expertise 40%, as (15) shown.

$$\text{Attractiveness} = \text{Familiarity} * 0.3 + \text{Likeability} * 0.3 + \text{closeness} * 0.3 + \text{physical attraction} * 0.1 \dots (12)$$

$$\text{Trustworthiness} = \frac{\text{Attractiveness} + \text{Expertise} + \text{Tie Strength}}{3} \dots (13)$$

$$\text{Expertise} = \frac{\text{Experience} + \text{Knowledge}}{2} \dots (14)$$

$$\text{Endorser's Credibility} = \text{Attractiveness} * 0.3 + \text{Trustworthiness} * 0.3 + \text{Expertise} * 0.4 \dots (15)$$

Figure 4 is based on above calculation which come out the top 10 influencers with the endorsed credibility. These 10 influencers were selected as the research objects of the product suitability analysis. 10 influencers are 蔡阿嘎、蔡桃貴(*A*), Kevin 老師(*B*),王君萍(*C*), Gina Hello(*D*), 林進(*E*), 唐綺陽(*F*), Hello Catie(*G*), 白癡公主(*H*), 阿滴英文(*I*), and 黃阿瑪的後宮生活(*J*).

	Familiarity	Likeability	Closeness	Physical Attraction	Attractiveness	Expertise	Tie Strength	Experience	Knowledge	Trustworthiness	Endorser Credibility
蔡阿嘎、蔡桃貴	6.27	3.76	0.05	15.07	4.53	5.06	3.16	0.91	9.2	4.25	4.66
Kevin老師	1.71	0.35	10.86	1.21	4	3.19	6.28	1.31	5.08	4.49	3.82
王君萍	3.42	2.55	1.2	6.31	2.78	3.46	2.31	1.21	5.71	2.85	3.07
Gina Hello	2.46	1.31	3.84	2.68	2.55	2.68	3.15	1.66	3.7	2.79	2.68
林進	2.87	1.72	2.95	1.76	2.44	1.99	2.91	1.47	2.5	2.44	2.26
唐綺陽	0.79	0.72	0.2	0.44	0.56	3.08	0.5	0.87	5.29	1.38	1.81
Hello Catie	1.3	1.14	0.23	1.54	0.95	2.65	0.76	1.23	4.07	1.46	1.78
白癡公主	1.2	2.66	1.07	1.32	1.61	1.99	1.14	1.49	2.49	1.58	1.75
阿滴英文	0.94	2.86	0.29	0.93	1.32	2.26	0.62	1.23	3.28	1.4	1.72
黃阿瑪的後宮生活	2.96	2.57	0.07	5.78	2.26	1.16	1.52	1.23	1.09	1.64	1.63

Figure 4. Top 10 influencers with the endorsed credibility

D. Design of Match-Up System

This system will analyze the social media data obtained by the web crawler system to filter the posts of influencer endorsements or advertorials. First, all the brands provided by Shopee in Taiwan [18] will be crawled, brands and their corresponding product categories are stored in an excel file to facilitate the subsequent selection of advertorials.

There are a total of 3287 store brands crawled, and 21 categories of products, they are women's clothing, men's clothing, mobile & gadgets, beauty & health, home & living, babies & Moms,

computers & peripherals, women's bags/boutique, women's accessories, shoes, men's bags & accessories, pets, automotive, food & gifts, sports & outdoors, entertainment & collection, miscellaneous, books & creative industries, services & tickets, games, home appliance.

After having the correspondence between the brand and the category, the Hashtag is taken out of the influencers' post and compared with the store's brand to determine whether the post is a post to promote the product. If the related name of the store brand appears in the hashtag, record this post, categorize this post as the advertorial of the category based on the corresponding category of the brand.

As the product category provided by Shopee in Taiwan is very detailed, this study will combine some categories with similar natures, and remove services & tickets, and miscellaneous. There are total 12 after consolidation which are 3C products, men's clothing and accessories, women's clothing and accessories, entertainment, babies, pets, home & living, sports, books & creative Industries, automotive, beauty & health, food & gifts. To analyze the number and number of likes of the advertorials made by influencers on each product category, the rating matrix of the influencer and the product category is constructed, as Figure 5 shown. These ratings are used to determine the effectiveness of the influencers' endorsement or advertorial is good or not. The 0 part of the matrix means that the influencers have never endorsed or promoted the post in the product category and the resulting matrix is used as the data set of this research.

		3C Products	Men's Clothing & Accessories	Women's Clothing & Accessories	Entertainment	Babies	Pets	Home & Living	Sports	Books & Creative Industries	Automotive	Beauty & Health	Food & Gifts
2	蔡阿嘎、蔡桃貴	4.12	5	5	0	5	0	5	1.7	5	0	0	5
3	Kevin老師	0	0.39	1.16	0	0.95	0	0.68	2.88	0	0	5	2.41
4	王君琄	4.43	1.77	1.83	0	0	0	0	1.96	0	0	0.46	0
5	Gina Hello	0	0.67	0.67	0	0	0	0	1.04	0	0	1.68	0
6	林柔	3.42	2.68	2.52	2.65	0	0	0	0	0	0	0	0
7	唐綺陽	2.04	0	0	0	0	0	0.29	0	0	0	0	0
8	黃阿瑪的後宮生活	0	0	1.34	5	0	0	0.61	0	2.63	0	0	0
9	Hello Catie	0	0.17	0	0	0	0	0	0	0	0	0.36	0
10	阿滴英文	5	0	0	0	0	0	0	0	0	0	0.84	0
11	白癡公主	0	0	0	0	0	0	0	5	4.65	0	0.35	0

Figure 5. The rating matrix of the influencer and the product category

After constructing the rating matrix, this research refers to the matrix factorization algorithm written by Yeung on GitHub [16], using the latent feature matrix and bias to predict the missing values in the original matrix, and uses the Stochastic Gradient Descent method to minimize the error and update the latent feature matrices *P*, *Q* and biases. The two latent feature matrices represent the association strength between product categories and latent features, and the association strength between influencers and latent features. The update rules for the latent feature matrices are shown in (16) and (17), and the update rules for the biases are shown in (18) and (19), bp_i is the bias of product category *i*, and bp_i is the bias of influencer j, where α is the learning rate and λ is the regularization parameter.

$$p'_{ik} = p_{ik} + \alpha(2e_{ij}q_{kj} - \lambda p_{ik})\ldots(16)$$

$$q'_{kj} = q_{kj} + \alpha(2e_{ij}p_{ik} - \lambda q_{kj})\ldots(17)$$

$$bp'_i = bp_i + \alpha(e_{ij} - \lambda bp_i)\ldots(18)$$

$$bi'_j = bi_j + \alpha(e_{ij} - \lambda bi_j)\ldots(19)$$

E. Survey Design

The questionnaires in this research is mainly designed by using the Likert scale. First, it will be based on the top 10 influencers with endorser's credibility mentioned in Figure 4., shown as above, and the 12 product categories sorted out, let respondents choose the suitability they think match up the influencers, the options for suitability are: very suitable, suitable, normal, unsuitable, very unsuitable, no comment, this research adopts a five-point scale and adds option of no comment to prevent the respondent from making careless choices without knowing whether it is suitable or not. The suitability investigation shown in the questionnaire which purpose is to investigate the suitability between influencers and product categories by public, in addition, it will also investigate whether the public will increase their willingness to buy products because of influencers' recommendations, then understand if the influence will impact public to purchase product or not.

IV. EXPERIMENT

This section will introduce the system information and explain the results of the experiment.

A. Survey Statistics

There are a total of 202 valid questionnaire samples collected by this research, with a 95% confidence level, about 7% sampling error [7], the statistic shows 68.3% of people will be willing to buy products because of the recommendation by influencers, the gender distribution filled in is 36.1% for males and 63.9% for females, while most of the age groups are 23 to 30 years old, accounting for 53.5%, followed by 19 to 22 years old, accounting for 29.2%.

Based on the score results of the questionnaire statistics, the influencers considered by the public to be the most suitable to endorse or advertorial for 12 categories of products are listed as Table 2 shown.

Table 2. Most suitable influencer for product category

Product Category	Most Suitable Influencer
3C Products	阿滴英文(*I*)
Men's Clothing & Accessories	Kevin 老師(*B*)
Women's Clothing & Accessories	Gina Hello(*D*)
Entertainment	蔡阿嘎、蔡桃貴(*A*)
Babies	蔡阿嘎、蔡桃貴(*A*)
Pets	黃阿瑪的後宮生活(*J*)
Home & Living	蔡阿嘎、蔡桃貴(*A*)

Sports	白癡公主(*H*)
Books and Creative Industries	阿滴英文(*I*)
Automotive	蔡阿嘎、蔡桃貴(*A*)
Beauty & Health	Kevin 老師(*B*)
Food & Gifts	蔡阿嘎、蔡桃貴(*A*)

B. Match-Up Result

After using the matrix factorization algorithm, the top 3 of matchmaking prediction results between 12 categories of products and influencers are shown as Table 3.

Table 3. Top 3 influencers predicted for product category

Product Category	Top 3 Influencers Predicted	
3C Products	1	阿滴英文(*I*)
	2	王君萍(*C*)
	3	白癡公主(*H*)
Men's Clothing & Accessories	1	蔡阿嘎、蔡桃貴(*A*)
	2	白癡公主(*H*)
	3	阿滴英文(*I*)
Women's Clothing & Accessories	1	蔡阿嘎、蔡桃貴(*A*)
	2	白癡公主(*H*)
	3	阿滴英文(*I*)
Entertainment	1	黃阿瑪的後宮生活(*J*)
	2	Kevin 老師(*B*)
	3	唐綺陽(*F*)
Babies	1	蔡阿嘎、蔡桃貴(*A*)
	2	阿滴英文(*I*)
	3	白癡公主(*H*)
Pets	1	蔡阿嘎、蔡桃貴

		(*A*)
	2	白癡公主(*H*)
	3	黃阿瑪的後宮生活(*J*)
Home & Living	1	蔡阿嘎、蔡桃貴(*A*)
	2	阿滴英文(*I*)
	3	白癡公主(*H*)
Sports	1	白癡公主(*H*)
	2	黃阿瑪的後宮生活(*J*)
	3	Kevin 老師(*B*)
Books & Creative Industries	1	蔡阿嘎、蔡桃貴(*A*)
	2	白癡公主(*H*)
	3	阿滴英文(*I*)
Automotive	1	蔡阿嘎、蔡桃貴(*A*)
	2	阿滴英文(*I*)
	3	王君萍(*C*)
Beauty & Health	1	Kevin 老師(*B*)
	2	黃阿瑪的後宮生活(*J*)
	3	唐綺陽(*F*)
Food & Gifts	1	蔡阿嘎、蔡桃貴(*A*)
	2	阿滴英文(*I*)
	3	林進(*E*)

In the prediction results, the top 3 influencers including the first influencer in the questionnaire results have a hit rate of 75%, and the hit rate is calculated as (20) shown.

$$\text{Hit Rate} =$$

$$\frac{\text{The number of categories for top 3 influencers in the predictions including the first influencer in the questionnaires}}{\text{Total number of product categories}} \ldots (20)$$

In other words, the top 3 predicted influencers by 9 out of 12 categories covered the top 1 influencer in the questionnaires. These 9 categories covered are 3C, babies, pets, home & living, sports, books & creative industries, automotive, beauty & health, food & gifts, and the categories of men's clothing & accessories, women's clothing & accessories, entertainment are not covered.

There are 2 out of 9 covered categories without any data in the original rating matrix. These 2 categories are pets and automotive, but after matrix factorization, the possible scores have been predicted which also covered the most suitable influencers in the questionnaires.

The reasons why the categories of men's clothing & accessories, women's clothing & accessories are not covered are because some clothing brands which are both selling men and women's products and the filtered results show the influencers with a large number of advertorials in these two categories are also different from the suitable influencers in the questionnaire. Therefore, there will be a large error when constructs the matrix. As to the category of entertainment, since the numbers of filtered advertorials are not that much and are also different from the suitable influencers in the questionnaire results, so the category of entertainment is not covered.

As the hit rate of covered categories showed, if suppliers find influencers to do advertorials or endorsements based on the predicted result, there is a 75% chance that the most suitable influencers defined by the public can be found from the top 3 influencers in the predicted results.

V. CONCLUSION

The research has two systems, one is influencer crawler system and the other is matchmaking system between influencer and product. First, use the influencer crawler system to crawl the information related to the requested influencers in the Instagram, then the matchmaking system between influencer and product will use these data to analyze the credibility of influencers' endorsement, according to the related information of the posts posted by influencers and sorted brands and categories in Shopee to filter out the advertorials which can construct a rating matrix for influencers and product categories, then use the matrix factorization in machine learning to predict the missing values in the matrix, and finally obtain verified data by questionnaires.

The result of the experiment shows the most suitable influencers for each product category selected by the public in the questionnaire have a 75% possibility of being the top 3 in the predict results, therefore if suppliers give their advertorials to the top 3 influencers as predict result shown then there is a 75% possibility to the most suitable influencers defined by public. Hope the systems proposed in this research can help suppliers find more suitable influencers as their endorsers to improve their marketing effectiveness.

REFERENCES

[1] M. Abdous and W. He, "Using text mining to uncover students' technology-related problems in live video streaming," British Journal of Educational Technology, vol. 42, no. 1, pp. 40-49, 2011.

[2] R. Ahuja, A. Solanki and A. Nayyar, "Movie Recommender System Using K-Means Clustering AND K-Nearest Neighbor," in 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), pp. 263-268, 2019.

[3] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in Noise Reduction in Speech Processing, Springer, Berlin, Heidelberg, pp. 1-4, 2009.

[4] J. DAWES, "Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales," International Journal of Market Research, vol. 50, no. 1, pp. 61-104, 2008.

[5] C. Friis-Jespersen, "Celebrity endorser's credibility: effect on consumers' attitude toward advertisement : Factors influencing vloggers credibility among viewers and their relation with attitude toward advertisement," 2017.

[6] G. Hofstede, G. J. Hofstede, and M. Minkov, "Cultures and organizations: Software of the mind," vol. 2, New York: Mcgraw-hill, 2005.

[7] G. D. Israel, "Determining sample size," 1992.

[8] Y. Koren, R. Bell and C. Volinsky, "Matrix Factorization Techniques for Recommender Systems," in Computer, vol. 42, no. 8, pp. 30-37, 2009.

[9] S. Kosub, "A note on the triangle inequality for the Jaccard distance," Pattern Recognition Letters, vol. 120, pp. 36-38, 2019.

[10] B. A. Lafferty and R. E. Goldsmith, "Corporate credibility's role in consumers' attitudes and purchase intentions when a high versus a low credibility endorser is used in the Ad," Journal of Business Research, vol. 44, no. 2, pp. 109-116, 1999.

[11] R. Likert, "A technique for the measurement of attitudes," Archives of psychology, 1932.

[12] Z. R. Maruf and A. D. Laksito, "The Comparison of Distance Measurement for Optimizing KNN Collaborative Filtering Recommender System," in 3rd International Conference on Information and Communications Technology, pp. 89-93, 2020.

[13] R. Ohanian, "The impact of celebrity spokespersons' perceived image on consumers' intention to purchase," Journal of Advertising Research, vol. 31 no. 1, pp. 46-54, 1991.

[14] P. Resnick and H. R. Varian, "Recommender systems," Communications of the ACM, vol. 40 no. 3, pp. 56-58, 1997.

[15] B. Till and M. Busler, "Matching products with endorsers: attractiveness versus expertise," Journal of Consumer Marketing, vol. 15 no. 6, pp. 576-586, 1998.

[16] A. A. Yeung, "Matrix Factorization in Python." https://github.com/albertauyeung/matrix-factorization-in-python (accessed Apr. 1, 2021).

[17] " Top 100 Influential Influencers Key Report." https://www.kolradar.com/billboard/2020-tw-top100-kol (accessed Sep. 11, 2020).

[18] "Shopee." https://shopee.tw/mall (accessed Feb. 28, 2021).