APPLYING MULTIMEDIA DATA MINING TO AN ASSOCIATIVE CLASSIFICATION-BASED RECOMMENDATION SYSTEM IN E-COMMERCE

Ling-Yi Chou
National Chung Cheng University, Taiwan
choulinyi@gmail.com

Shi-Ming Huang
National Chung Cheng University, Taiwan
smhuang@mis.ccu.edu.tw

I-Heng Wu
Asus, Taiwan
brive3571@gmail.com

Abstract:
With increasing popularity of e-commerce, the growth potential of this business is extraordinary; however, the main challenge for such a paradigm originates from the difficult in finding the most valuable products to recommend customer appropriate information. One of the problems is that business managers generally lack sufficient information to identify whether the product descriptions and images in the website successfully attracted customers in browsing and buying the products. The objective of this research is to develop a recommendation mechanism using multimedia data mining algorithm. Basically, it analyses a large amount of multimedia images information to find the most valuable images and descriptions of products. Our findings show that these images and descriptions match customers' various demands. This paper use the color image semantics applied in the Munsell (HSV) color space and use the content-based image retrieval application to establish a recommendation rule based on an associative classification method for various product images in e-commerce. The prototype has been developed to evaluate the mechanism feasibility and is applied in a real e-commerce case. The result shows that it is useful for on-line sellers to make recommendation decisions. In order to further validate the practicability of this mechanism, the extensive case study with larger trading volumes in apparel e-business areas was conducted. This case study shows that the multimedia data mining approach is an automatic and effective solution, which can be used in marketing to discover consumer habits.

Keywords: data mining, image processing, color semantics, recommendation system, web crawler

1. INTRODUCTION

During the last decades, information technology and the Internet have had a dramatic effect on business operations. With increasing popularity of e-commerce, the growth potential of this business is extraordinary. It is not only convenient to purchase product online for daily use of consumers, but also creates huge market opportunity in each year. Online sellers provide large information to induce consumers to adopt their online channels, and it has become one of the most important way to motivate consumer to purchase repeatedly through these channels.(Chiu, Wang, Fang, & Huang, 2014; Teo & Liu, 2007). For example, a wealth of product title, image and description on the website which were provided from online sellers is in order to increase consumers' the perceive satisfaction.

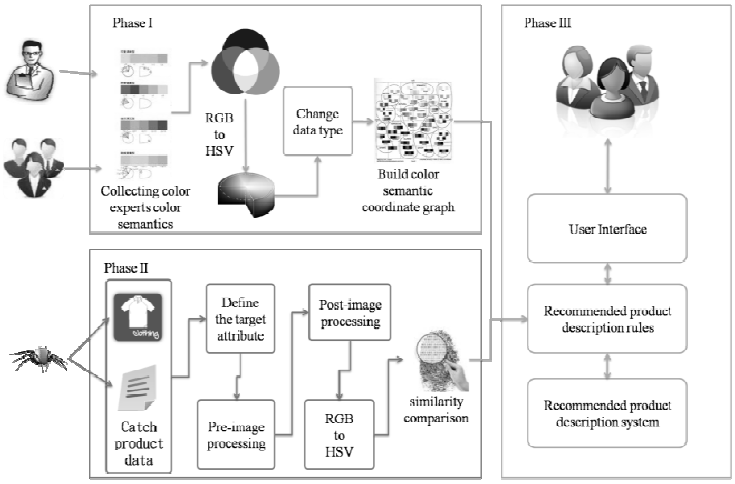
As the result, this study proposes a mechanism to analyse the color semantics in accordant with a large amount of multimedia images information on the online store. We use a multimedia data mining technology and the image retrieval approach to build an associative classification-based recommendation system for products in e-commerce. The objective is to find the most valuable images and descriptions online. First, we collect color list and import it to our database, and we get color change based on HSV model which build our color semantic coordinate graph. Second, we use web crawler and image processing technology to build our system content. Besides, we use image processing to distinguish foreground and background color in order to analyse color semantic weights. Third, we compare to Yahoo search function to evaluate which is more close to customer feeling on color semantic. Finally, we develop an information system to reduce upload time on C2C online platform for product information in the apparel distribution industry.

2. RESEARCH DESIGN AND DEVELOPMENT

In this section, we use the data mining technology and image color semantic analysis method to build the mechanism in the marketing area. However, the main purpose of this study is to build a mechanism to evaluate the mechanism and feasibility. Our design framework is as Picture 1.

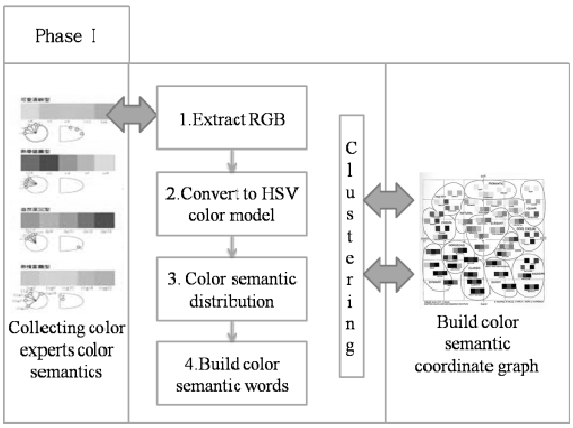
The first stage is the establish a color semantic coordinate graph through collecting color semantic image from experts. The color semantic coordinate graph is used to provide the color semantic in the recommend system. The second stage is to use the web crawler to catch apparel product information on the Yahoo C2C e-commerce platform through image processing techniques to distinguish between the foreground and background, then we use the similarity comparison and data mining to make cluster analysis in order to find the color semantic for user to recommend product to them. The third stage is to bulid product description rules, and the data are stored into recommended product description system. In this stage, the user interface allows users to query and upload image, and at the moment, system would automatically find the matching product descriptions and color semantics in order to save the relevant personnel time to establish product color semantic vocabulary, and enhance changes between the product description vocabularies.

Picture 1: Recommended product description system of flow chart



Source: This study, 2014.

Picture 2: Phase one flow chart



Source: This study, 2014.

2.1. Build a color semantic coordinate graph

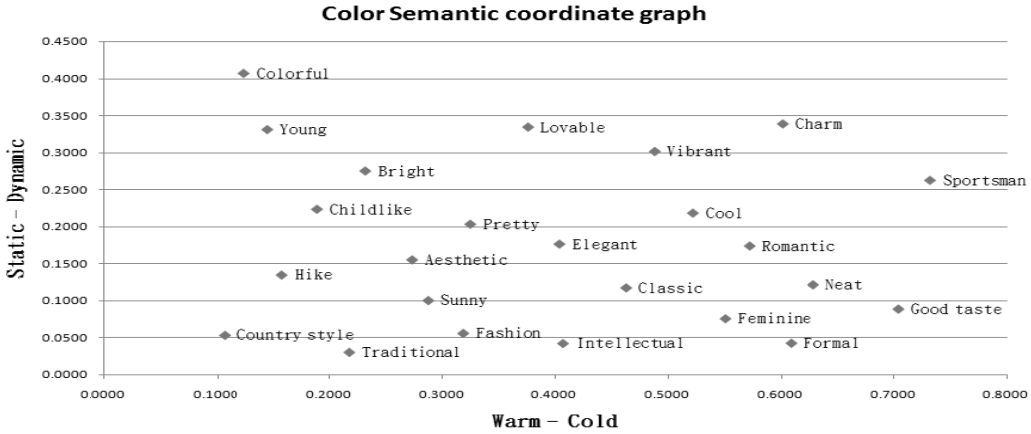
To build a color semantic coordinate graph, we collected color semantics list from the Institute of IRI and related literature to establish process as shown in Picture 2. First, we obtain RGB color list which are adjective color. Next, we convert RGB color into the HSV color and store it in the database. In this stage, we establish the corresponding color adjective phrases. Third, we use the Munsell color system and the ISCC-NBS color system to build the color semantic distribution through the corresponding results of XY axis distribution. X-axis distribution refers to the warm and cold color sense. Y-axis distribution refers to the hard and soft color sense. Finally, its semantic distribution stored in the color semantic database.

Table 1:. several examples RBG convert to HSV

Color name	Color	RGB	HSV
Red		(255,0,0)	(0,100,100)
Green		(0,255,0)	(120,100,100)
Blue		(0,0,255)	(240,100,100)
Gray		(128,128,128)	(0,50,50)
Brown		(150,75,0)	(30,100,59)
Pink		(255,192,203)	(350,25,100)

Source: Smith, 1997.

Picture 3: Color semantic coordinate graph

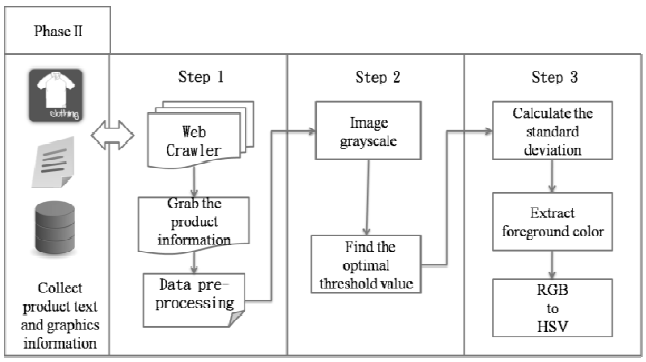


Source: This study, 2014.

2.2. Product description crawling mechanisms

We collect data from the Yahoo C2C platform. The reason that we adopt data from Yahoo platform is not only the rich product information and image, but also high integrity of website for online seller. The process is shown in picuture 4. Finally, we will make similarity matching between output results based on recommended product description rules produced from the third stage.

Picture 4: Phase two flow chart



Source: This study, 2014.

Image processing can be divided into two parts: pre-processing and post-processing. We use the image segmentation techniques for image processing in order to make assure image color semantic close to the human perceptions of clothes color image. The main purpose of image segmentation is to filter the useful image information from original photos. That is, the useful image information was separated from the not useful image information through the statistical computing from image grayscale characteristics.

First, we get the grayscale values of each image. Second, we compute the threshold to decide which color points is belong the gray value. We make the color point as dark spots when the gray value is lower than the threshold; similarly, we make the color point as bright spots when the gray value is higher than the threshold.

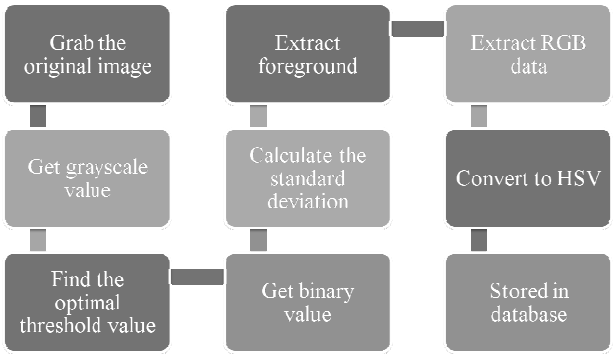
There are many methods to get the threshold such as the percentage method (P-Tile Method), mode method, Otsu method, and the average gray value method (Mean-Value Method). We use the Mean-Value Method to compute the threshold for color images, because the pictures were grasped roughly have same proportion of the semantic distribution between foreground and background on the website. Mean-Value Method helps us to automatically find the optimal threshold value through threshold algorithm, and automatically adjust with the changes of the external light source and does not affect

the segmentation results The grayscale average computation formula of the mean-value method is as follows:

(5)

We distinguish into foreground and background to minimize background interference factors. There are nine process steps during image processing as shown in Picture 5. We actually convert the program implemented according to these steps, and finally we randomly selected images to crawl down the distinction between foreground and background results.

Picture 5: Image processing nine steps



Source: This study, 2014.

Table 2: Image processing

Table 2: Image processing	
Original Image	**Image processing result**
	
	

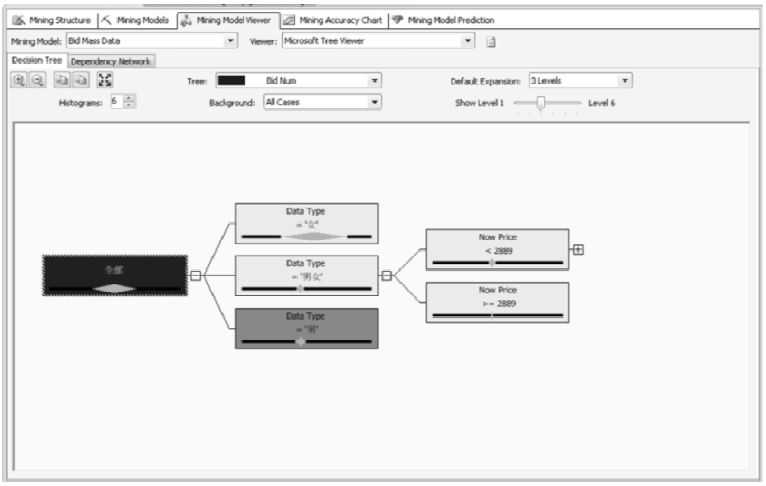
Source: This study, 2014.

2.3. Recommended product description and system development

(1) Decision tree analysis

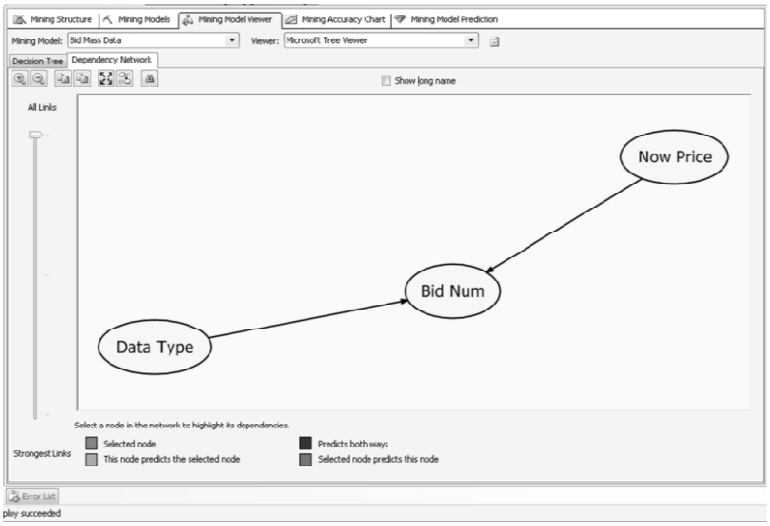
In this section, we use a decision tree analysis tool to help make decision for recommend products. To enhance the system efficiency, this study does not modify the database state, and removed it after the data conversion completed. This study use Microsoft SQL Server2008 to analysis and training data for establishing the number of selling prices decision tree for products. The classification was shown as in Picture 6. COMPLEXITY_PENALTY = 0.50. Picture 7 shows the network dependencies for each attribute while it would use to system development.

Picture 6: Microsoft decision tree system result



Source: This study, 2014.

Picture 7: Microsoft decision tree network dependent surface



Source: This study, 2014.

(2) Recommended the product to describe the system mechanism

We use the convenient features of the Internet, Web development, and Home part of the RIA concept for implementation. Win7 operating system as the system operating platform, the front pages use the html5 Javascript, CSS, Jquery, while the background is PHP and Java and, database using Sqlservre2008, each module features are described below in the table below. Besides, we use alpha testing and beta testing when the prototype system is finished.

Table 3: Recommended product description system module description

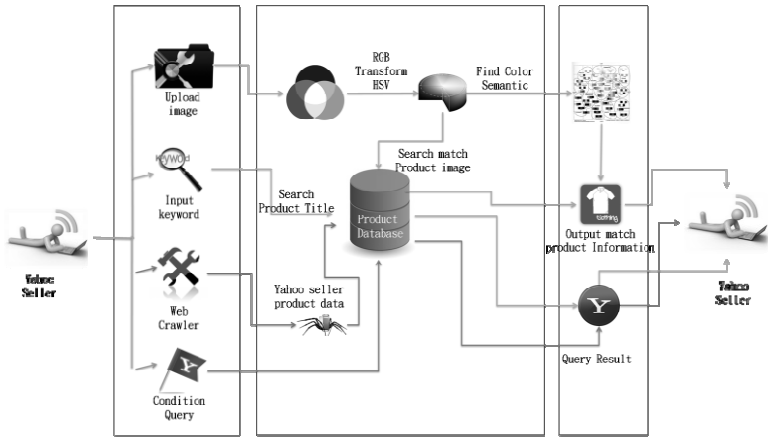
System function	Function description
Upload image function	Upload product images, the system will provide product information on the Yahoo auction site described as a design product.
By Query function	By title keyword search to keyword distribution and correlation analysis of a product within a specific point in time.
Color semantic search function	System through the color semantic ontology to recommend its related adjective sentence, described as the product was added to our text.
Web crawler function	Enter crawling Yahoo apparel product information management capabilities, can enter the URL, the seller name, classification

Source: This study, 2014.

3. RESULTS

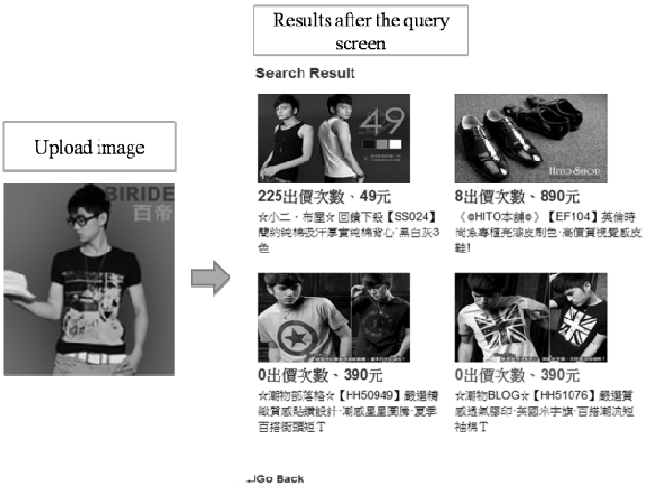
Picture 8 is the research system architecture diagram. We can make color semantic analysis through the Yahoo online product information to recommend product. The mechanism establishes the content of the color semantic through multimedia data mining.

Picture 8: System architecture diagram



Source: This study, 2014.

Picture 9: Upload image function



Source: This study, 2014.

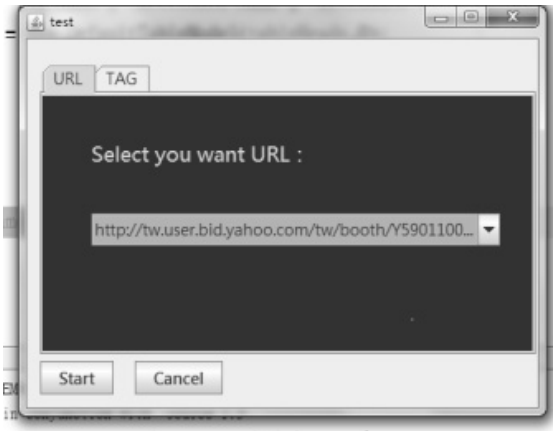
The color semantic search module developed to establish good color semantic maps for users' query, if the Internet seller would like to know the perception of the picture, users can input keywords to query by using the function module as shown in picture 10. Web crawler can create, modify, delete and query Yahoo apparel product information and management capabilities. Online seller and web designer can enter the URL, the seller name, classification, as shown in picture 11.

Picture 10: Query function

賣家名稱(Store Name)	圖片 (Image)	價錢 (Price)	出價次數 (Bid count)	連結(Link)
GUGU&GIGI		120	0	http://tw.page.bid.yahoo.com/tw/auction/c71378100?u=Y3860195802
TAIJI		250	11	http://tw.page.bid.yahoo.com/tw/auction/f333595742?u=Y7195551075
椒糖饅頭		320	0	http://tw.page.bid.yahoo.com/tw/auction/d67270523?u=Y9527790822

Source: This study, 2014.

Picture 11: web crawler link



Source: This study, 2014.

4. CONCLUSION

This study use the multimedia data mining technique to build a mechanism which can help online sellers save time to provide the product description for color semantic of users. Our findings show that these images and descriptions match customers' various demands. This paper use the color image semantics applied in the Munsell (HSV) color space and use the content-based image retrieval application to establish a recommendation rule based on an associative classification method for various product images in e-commerce. The prototype has been developed to evaluate the mechanism feasibility and is applied in a real e-commerce case. The result shows that it is useful for on-line sellers to make recommendation decisions. In order to further validate the practicability of this mechanism, the extensive case study with larger trading volumes in apparel e-business areas was conducted. This case study shows that the multimedia data mining approach is an automatic and effective solution, which can be used in marketing to discover consumer habits.

REFERENCE LIST

1. Akgül, C. B., Rubin, D. L., Napel, S., Beaulieu, C. F., Greenspan, H., & Acar, B. (2011). Content-based image retrieval in radiology: current status and future directions. Journal of Digital Imaging, 24(2), 208-222.
2. Assfalg, J., Del Bimbo, A., & Pala, P. (2000). Using multiple examples for content-based image retrieval. Paper presented at the Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on.
3. Babu Rao, M., Prabhakara Rao, B., & Govardhan, A. (2011). CONTENT BASED IMAGE RETRIEVAL USING DOMINANT COLOR, TEXTURE AND SHAPE. International Journal of Engineering Science & Technology, 3(4).
4. Bahmanyar, R., & Datcu, M. (2013). Measuring the semantic gap based on a communication channel model. Paper presented at the Image Processing, IEEE International Conference on.
5. Berry, M. J., & Linoff, G. S. (2004). Data mining techniques: for marketing, sales, and customer relationship management: John Wiley & Sons.

6. Bhoyar, K., & Kakde, O. (2009). Image retrieval using fuzzy and neuro-fuzzy approaches with fuzzy color semantics. Paper presented at the Digital Image Processing, 2009 International Conference on.
7. Chiu, C. M., Wang, E. T., Fang, Y. H., & Huang, H. Y. (2014). Understanding customers' repeat purchase intentions in B2C e-commerce: the roles of utilitarian value, hedonic value and perceived risk. Information Systems Journal, 24(1), 85-114.
8. Delen, D., & Crossland, M. D. (2008). Seeding the survey and analysis of research literature with text mining. Expert Systems with Applications, 34(3), 1707-1720.
9. Heydon, A., & Najork, M. (1999). Mercator: A scalable, extensible web crawler. World Wide Web, 2(4), 219-229.
10. Lee, Y.-H., Hu, P. J.-H., Cheng, T.-H., & Hsieh, Y.-F. (2012). A cost-sensitive technique for positive-example learning supporting content-based product recommendations in B-to-C e-commerce. Decision Support Systems, 53(1), 245-256.
11. Liu, C., & Song, G. (2011). A method of measuring the semantic gap in image retrieval: Using the information theory. Paper presented at the Image Analysis and Signal Processing (IASP), 2011 International Conference on.
12. Lua, K.-T., & Gan, G. (1994). An application of information theory in Chinese word segmentation. Computer Processing of Chinese and Oriental Languages, 8(1), 115-124.
13. Madaan, R., Dixit, A., Sharma, A., & Bhatia, K. K. (2010). A Framework for Incremental Hidden Web Crawler. International Journal on Computer Science & Engineering, 2(3).
14. Nijssen, S., & Fromont, E. (2010). Optimal constraint-based decision tree induction from itemset lattices. Data Mining and Knowledge Discovery, 21(1), 9-51.
15. Rungsawang, A., & Angkawattanawit, N. (2005). Learnable topic-specific web crawler. Journal of Network and Computer Applications, 28(2), 97-114.
16. Smith, J. R., & Chang, S.-f. (1997). Querying by color regions using the VisualSEEk content-based visual query system. Intelligent multimedia information retrieval, 7(3), 23-41.
17. Speier, C., & Morris, M. G. (2003). The influence of query interface design on decision-making performance. MIS quarterly, 397-423.
18. Tang, J., Zha, Z.-J., Tao, D., & Chua, T.-S. (2012). Semantic-gap-oriented active learning for multilabel image annotation. Image Processing, IEEE Transactions on, 21(4), 2354-2360.
19. Teo, T. S., & Liu, J. (2007). Consumer trust in e-commerce in the United States, Singapore and China. Omega, 35(1), 22-38.
20. Thelwall, M. (2001). A web crawler design for data mining. Journal of Information Science, 27(5), 319-325.
21. Zhang, Y., Jia, Z., & Chen, T. (2011). Image retrieval with geometry-preserving visual phrases. Paper presented at the Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on.