A method to predict color preferences through data mining process

Lee, Tien-Rein, Sun, Ching-Wen and Tsai, Cheng-Min

Graduate Institute of Information Communications, Chinese Culture University, 55, Hwa-Kang Rd. Taipei, Taiwan, R.O.C 11192 Corresponding author: T.R. Lee (trlee@staff.pccu.edu.tw)

Keywords:

Color preference, data mining, predicting model, market basket.

ABSTRACT

In seeking an effective method to extract color preference knowledge from large amounts of data, this study aimed to find out whether there are relations between color preferences and other preferences through a data mining process. Five sets of images were collected to represent seasons, geographical landforms, time periods, fruits, and flowers. Objects such as mugs, T-shirts, chairs, motorcycles, floppy disks, and backpacks in the 8 NCS colors were also used as survey stimuli. To find out their color preferences as well as preferences of environmental settings, this study surveyed 309 college students of Chinese Culture University. After data were collected, a series of "market basket analyses" statistical procedures were performed to derive hidden patterns of color preferences and their association rules. A color preference knowledge discovery model was also introduced to predict trends of the targeted people's color preferences.

1. INTRODUCTION

In order to search for alternative methods of exploring people's color preferences, this research attempted to find out whether there is a way to know someone's color preference without conducting a direct color preference survey. Another question is: Can we know someone's color preferences by knowing his other preferences?

Conventional colour preference related information is usually gathered through a survey type of investigation. Usually, amounts of survey data are very large. Knowing how to effectively manage or analyse data from surveys and how to organize such data to get new information is essential. Han & Kamber (2000) point out that data mining refers to extracting or "mining" knowledge from large amounts of data. Therefore, in this study we conducted a series of preference related surveys to collect enough data to use a data mining method to derive people's color preferences, and build a knowledge database on these preferences which can enhance communication invovling color.

The subjects in this research were asked to choose their most favorite and least favorite colors or color images (products) on the monitor screen. The survey was conducted using an electronic format in a controlled environment with calibrated monitors.

2. METHOD

Subjects. The subjects were 309 undergraduates (143 males) from the College of Journalism and Communications at Chinese Culture University. All of them were screened for color deficiencies using the Ishihara Color Vision Test.

Stimuli. A set of 8 digital color chips was prepared. The 8 colors selected (R, Y50R, Y, G50Y, G, B50G, B and R50B), were chosen because they represent an even coverage of the outer edge of the NCS hue circle. Color chips, mugs, T-shirts, mopeds, floppy disks, and backpacks were integrated with 8 colors to provide the stimuli for investigating subjects' color preferences. In addition, image sets of seasons, geographical landforms, time periods, fruits, and flowers were presented to elicit subjects' preferences (figure 1).



S/S, S/A, A/W, and W/S, are seasonal transition periods.

Figure 1: All images used in stimulus sets for the experiment.

Objects. The first set of color photographs containing six objects (designated: mug, T-shirt, chair, moped, disk and backpack) was prepared. Each of the object photos was then colorized to match the 8 colors of the digital chips. The second set of photos contains 5 categories of images (designated: seasons, day time periods, landforms, flowers, and fruits).



Figure 2: Layout of survey images shown on the screen, left-mug, right-seasons.

These sets were displayed as shown in figure 2.

Procedure. After subjects reported to the laboratory and passed the color vision test, they read the instructions of the experiment. An electronic survey was then conducted in a controlled environment. After being briefed, subjects provided information about their gender, primary region of residence in Taiwan, favorite leisure time activity, and personality. Subjects then ranked their most and least favorite colors on each page displayed on the computer screen using a web-enabled survey system. All 309 subjects participated in the survey. These subjects were asked to select their most favorite colour and least favorite colour out of 8 color choices. After this color survey, various preferences and demographic data were collected to build up a storehouse of other relevant information.

Cup	dawn								morning								noon							
	C1	C2	C3	C4	C5	C6	C7	C8	C1	C2	C3	C4	C5	C6	C7	C8	C1	C2	C3	C4	C5	C6	C7	C8
Q1	7	6	2	2	6	8	18	0	0	2	3	0	2	2	4	4	2	1	1	0	0	1	4	0
Q2	3	3	0	2	1	2	0	1	3	1	0	0	0	0	0	0	0	0	1	0	0	0	3	0
Q3	5	5	4	1	1	5	7	0	1	1	1	0	0	2	0	0	0	1	0	1	0	2	1	0
Q4	2	8	3	1	1	6	8	3	2	0	0	0	0	1	0	0	0	0	0	0	1	1	0	1
Cup	afternoon							dusk								evening								
Cup				anci	noon							uu	SK							CVC	inng			
Cup	C1	C2	C3	C4	C5	C6	C7	C8	C1	C2	C3	C4	C5	C6	C7	C8	C1	C2	C3	C4	C5	C6	C7	C8
Q1	C1 2	C2 1	C3 2	C4 0	C5 2	C6 3	C7 2	C8 1	C1 2	C2 1	C3 2	C4 1	C5 2	C6 2	C7 7	C8 5	C1 5	C2 2	C3 3	C4 1	C5 1	C6 6	C7 11	C8 1
Q1 Q2	C1 2 0	C2 1 0	C3 2 0	C4 0 0	C5 2 0	C6 3 1	C7 2 0	C8 1 0	C1 2 1	C2 1 0	C3 2 1	C4 1 0	C5 2 0	C6 2 3	C7 7 3	C8 5 0	C1 5 2	C2 2 1	C3 3 1	C4 1 0	C5 1 0	C6 6 2	C7 11 1	C8 1 2
Q1 Q2 Q3	C1 2 0 2	C2 1 0 3	C3 2 0 0	C4 0 0 0	C5 2 0 0	C6 3 1 1	C7 2 0 4	C8 1 0 0	C1 2 1 5	C2 1 0 2	C3 2 1 1	C4 1 0 0	C5 2 0 3	C6 2 3 0	C7 7 3 2	C8 5 0 0	C1 5 2 3	C2 2 1 1	C3 3 1 0	C4 1 0 1	C5 1 0 0	C6 6 2 2	C7 11 1 2	C8 1 2 2
Q1 Q2 Q3 Q4	C1 2 0 2 1	C2 1 0 3 2	C3 2 0 0 0	C4 0 0 0 0	C5 2 0 0 0	C6 3 1 1 0	C7 2 0 4 0	C8 1 0 0 0	C1 2 1 5 3	C2 1 0 2 4	C3 2 1 1 2	C4 1 0 0 0	C5 2 0 3 1	C6 2 3 0 2	C7 7 3 2 3	C8 5 0 0 0	C1 5 2 3 3	C2 2 1 1 1	C3 3 1 0 1	C4 1 0 1 0	C5 1 0 0 0	C6 6 2 2 1	C7 11 1 2 2	C8 1 2 2 0

Data collection. After a series of preference surveys was completed, data were collected from the 309 subjects (table 1), and structural data matrices were constructed (figure 3, 4). The Market Basket Analysis (sometimes called Affinity Grouping) was chosen over many other data mining techniques to perform the process of exploring relations among the above factors.





Figure 3: The 3-D data cube representation of the data in Table 1, preferences of recreation types, time period, and mug color.

Figure 4: Lattice of cuboids, making up a 4-D data cube for the dimensions recreation type, color preference, time period, and object. Each cuboid represents a different degree of summarization.

3. RESULTS

Preferences. Ranking orders of subjects' preferences and some of the sample findings of the study are listed on table 2, and 3 below:

1. In general, blue (B) and blue products are the most popular; red (R) and orange (Y50R) products are second and third, respectively. Purple (R50B) and purple products are the least favorite, preceded by turquoise and green.

2. Turquoise (B50G) is the least favorite chair color, while green is the least favorite disk color.

3. A person who likes a certain color tends to favor objects with that color. For example, 55.6% of those who like red mugs will also favor red T-shirts; 50.6% of persons who favor blue mugs tend to like blue T-shirts.

4. Regional and seasonal preference information can be used to predict object color preferences. For example, of those who live in southern areas and favor the seashore, more than 50% favor blue T-shirts. 55% of males who like the transition period between spring and summer, also like blue T-shirts.

5. Some colors are disliked by most people. In terms of gender, 66.9% of females and 52.4% of males dislike blue-green chairs. More than half of the people who live in northern and central Taiwan dislike blue-green chairs.

6. From negative preference sets, one can find positive preferences, too. For example, 31.1% of those who dislike winter and noon time period, like red T-shirts. Among those who dislike summer and noontime, 35.1% favor blue chairs.

Color Chip			Mug			T-shirt			Chair			Moped			Floppy disk			Backpack		
Color	Μ	L	Color	Μ	L	Color	Μ	L	Color	М	L	Color	Μ	L	Color	Μ	L	Color	М	L
Y	3	3	Y	5	2	Y	5	4	Y	6	3	Y	4	5	Y	1	8	Y	3	3
Y50R	3	7	Y50R	4	2	Y50R	3	6	Y50R	2	6	Y50R	3	6	Y50R	7	3	Y50R	2	7
R	2	4	R	2	8	R	1	8	R	3	7	R	2	7	R	4	5	R	4	6
R50B	5	2	R50B	7	1	R50B	6	1	R50B	4	2	R50B	6	1	R50B	6	2	R50B	7	1
В	1	8	В	1	7	В	2	7	В	1	4	В	1	8	В	2	6	В	1	5
B50G	6	1	B50G	3	4	B50G	4	2	B50G	8	1	B50G	5	3	B50G	3	7	B50G	6	2
G	8	4	G	6	5	G	7	5	G	5	7	G	8	2	G	8	1	G	8	4
G50Y	6	4	G50Y	8	6	G50Y	7	3	G50Y	7	5	G50Y	7	4	G50Y	5	4	G50Y	5	8

Table 2: Ranking order of subject's favorite and least favorite colored product.

"M"- most favorite; "L" - least favorite.

Sea	son	Land	form	1	Time p	perio	bd	Frui	t		Flower			
item	М	L	item	Μ	L	item	Μ	L	item	Μ	L	item	Μ	L
Spring	4	7	Coast	1	8	Dawn	1	6	Watermelon	1	10	Carnation	1	5
S/S	1	5	Plain	2	5	Dusk	2	3	Cherry	2	4	Tulip	2	11
Summer	5	2	Glacier	3	7	Evening	2	4	Wax apple	3	11	Rose	3	8
S/A	2	8	River	4	5	Morning	4	5	Strawberry	4	8	Lily	4	8
Autumn	3	6	Mountain	5	4	Afternoon	5	2	Apple	5	12	Lotus	5	11
A/W	6	3	Volcano	6	3	Noon	6	1	Guava	6	6	Araceae	6	3
Winter	7	1	Canyon	7	2				Banana	7	9	Osmanthus	7	2
W/S	8	4	Desert	8	1				Tangerine	8	6	Orchid	8	7
									Grapes	8	5	Bird of paradise	9	1
									Pineapple	10	2	Narcissus	10	3
									Papaya	11	3	Sunflower	11	6
								Starfruit	12	1	Plum	12	10	

Table 3: Ranking order of subject's favorite and least favorite natural scenes.

"M"- most favorite; "L" - least favorite.

Data mining results. After analyzing data from the survey of 309 subjects, structural data matrices were constructed. Examples of predictability exceeding 40% for people having same color preferences based on demographic considerations:

1. 45.24% of males who live in northern Taiwan like blue mopeds.

2. 44% of those who think they are introverted like blue chairs, and 44% of them dislike purple T-shirts.

3. 40% of males who like sport also like blue chairs. Of females who like recreation life, 48.72 % dislike green disks.

4. Among males who think they are extroverted, 42.65% like blue mopeds. Meanwhile, 48.53% of them dislike purple T-shirts.

5. Half of senior male students dislike purple mugs. 42.50% of those who favor audio-visual as their leisure activity dislike purple mugs.

Examples of predictability exceeding 50% for people having the same color preference based on natural scene conditions:

1. 52.94% of those who like tulips and seashores like red mopeds.

2. Of those who dislike noontime, and also dislike pineapples, or starfruits, or bird of paradise flowers, or deserts, or winter, or summer, 50% also won't like turquoise chairs.

3. If a person dislikes deserts, and also dislikes summer, or winter, or pineapples, or starfruits, or bird of paradise flowers, there is a greater than 50% chance he won't like turquoise chairs either.

4. CONCLUSIONS

In general, people's color preference varies with the category of object viewed. We can't predict someone's preferences toward different objects based only on their chip color preference.

Using the data mining process, we found commonalities among subjects having the same association rules. However, by increasing our subject sample size, demographic-related information, number of surveys, and database size, we should be able to discover more patterns, and also achieve greater accuracy.

Based on what we found from this exploratory data mining study, we conclude that by joining demographic data along with preference associations, one can construct a model to predict someone's color preference. By reversing the model, we might be able to predict someone's other preferences or demographic information by knowing his color preference. The pattern we found did not enable 100% predictability, but did enable certain percentages of predictability.

References

- 1. Thuraisingham, B., Web data mining and applications in business intelligence and counterterrorism (CRC Press, Florida, U.S.A., 2000)
- 2. Han, J. & Kamber, M., Data mining: Concepts and Techniques (Academic Press, 2000)
- 3. Lee, T. R., Color preference and its applications (Asian-Pacific Publishing, Taipei, Taiwan, 2002).