Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 7, July 2021: 7402 - 7411

Research Article

Analysis of TV Advertising Video Color Using Color Extraction Based on K-means Clustering using Big Data

Seung-Yeon Hwang[a], Jeong-Joon Kim[b]

[a] Dept of Computer Engineering, University of Anyang, Anyang-si, Gyeonggi-do, Republic of Korea
[b] Corresponding Author, Dept of Software, University of Anyang, Anyang-si, Gyeonggi-do, Republic of Korea

Abstract

Color is an emotional element that moves people's minds. Advertising stimulates people 's emotions, raising the value of goods and promoting consumers' desire to purchase. The color strategy in advertising has become an important element of marketing. In today's competitive environment, color marketing is used as a key management strategy that differentiates itself from other products and determines the success or failure of the company. Based on the Big Data collected by Creative Advertising Center of Korea Advertising Association, this paper analyzed to find elements other than the color marketing strategy used in the existing advertisements. In this paper, we propose a new method to collect color information using K-means clustering based on the advertisement image collected by web crawling, and analyzed the correlation between the consumers 'involvement' and the color of the advertisement according to the advertisement category.

Keywords: Big Data, Color Marketing, Advertising, K-means Clustering, Web Crawling, Involvement

1. Introduction

Marketing is becoming more important in modern society. Depending on the strategy of marketing, a company's sales or management vary. Types of marketing include viral marketing, noise marketing, coz marketing, PPL marketing, and color marketing[1]. In this paper, we will focus on color marketing. Color marketing is the most important variable in purchasing a product by setting the color to increase the purchasing power of consumers. Humans can get various feelings such as warmth, purity, and cleanness through color. These colors instantly give consumers a variety of feelings, such as brand images and product images. In addition, according to the study on the emotional factors determinants of products by the Color Research Institute of America, the proportion of vision among the five senses is 87%, and the proportion of color among these visions is very high at 60%[2]. Colors can be seen very clearly in advertising videos. This is because these commercial videos require a short time of 15 to 30 seconds to draw attention from many people. Although there are many factors such as product design, color, music, and person in order to attract the attention of consumers in advertising videos, color is a very important factor in visuals according to the study of emotional factors of products.

In 2016, Japanese food company Clorretz made two advertisements, one of which was advertised by AI and the other by man-made advertisements. It is said that advertisements made with AI were made by adding big data and product information from existing advertisement videos. The results of consumer preference were surprising. The preference for advertising videos made by human directors was 54%, while that of advertising videos made by artificial intelligence directors was 46%. The difference between the two preferences is very

narrow, although human director has won[3]. Therefore, we would like to analyze the information of big data advertising videos that are combined with AI in detail.

In this paper, colors were extracted using K-means clustering based on advertisement images collected using web crawling at Creative Advertising Information Center of the Korea Advertising Federation, and the relationship between advertisements and colors was analyzed.

II. Related research

1. Web Crawling with Python

For data collection, we collected 2016 advertising videos using web crawling. If you look at the video URL(https://www.adic.or.kr/ad/tv/show.cjsp?ukey=1581322&oid=)) of the Advertising Information Center, you can see that there is a unique key to the advertisement, such as uki=1581322. The ukey values in 2016 ranged from 1576914 to 1593626, and all advertisement videos and information in this range were imported. General information of advertisement video in html was crawled using beatuifulsoup, The ad video, which is an extension of the flv file, was crawled using PhantomJS and Selenium to help access the script because it was written in the script language. During crawling, the crawling was performed while using sleep at intervals of 1 to 3 seconds due to periodic access from the same IP.

2. Extract as an advertising video image

Image extraction from the video was the most obvious scene change, when visually verified, to extract images in units of 100 frames. As a result, approximately 33,000 images were extracted from 5272 advertised videos in 2016 on a 100-frame basis. Image extraction was created in Python using OpenCv as a jpg file format.

3. Image color extraction with K-means clustering

From the image extracted from the advertisement video, we extracted the color using OpenCv and then using the K-means clustering algorithm to extract the color in RGB form. The simple concept of k-means clustering is to cluster data among k similar properties.

Figure 1 shows the units of color extraction from images in cluster 5, cluster 10, and cluster 15. When visually checked, the color extraction unit was decided as Cluster 10 because the color was clearly distinguished and various colors were pulled out.

We have 10 clusters of colors in all images, extracted the RGB value, and made it into a txt file with the ratio information and separator of the corresponding RGB value. The txt file contains the id, RGB value, and ratio information of the image.

Fig 1. Color Extraction According to Size of Clustreing

4. Video color TOP10 using mapreduce

After moving all image color txt files to the Hadoop system, we created a color TOP10 for one video via mapreduce. One thing to be aware of when selecting the TOP10 is that it is the same color when viewed with the human eye, but the RGB value is slightly different by 1 or 2, so it is sometimes treated as a different color. Therefore, the results also showed that TOP10 was all selected in similar colors. Two methods were used to exclude such cases. The first is a method of processing K-means in order to once again extract the color of the image selected by K-means. However, since this method is a method of applying clustering once again to the color that has already been clustered, the exact desired color did not come out. The second method applies an algorithm that captures similar colors of RGB colors. The algorithm selected TOP10 after treating each data of red (R), green (G), and blue (B) as having a reference value of 32 different values as the same color. After many trials and errors, the value of 32 was determined to be the most appropriate range when visually verified. If you check the image of the video color extraction result in Figure 2, four pictures are images of one advertising video. The color graph at the bottom of the image is the result of selecting the video color TOP10 using the algorithm. When visually checked, In the advertisement video, the representative colors were selected, and the colors were not similar, and the results of various colors were confirmed.

Fig 2. Result of Color Extraction from Advertising Video

5. Convert RGB to HSB

The colors extracted from the advertisement video are all RGB. The principle of RGB is color-coded by the weight of the color, ranging from 0 to 255 for each of the following: Red(R), Green(G) and Blue(B). Therefore, it is difficult to find out representative colors from color data. For example, if there are dark yellow, yellow, and light yellow, the representative colors here are yellow. However, RGB's color representation method makes it difficult to identify the representative color, yellow. To address this problem, we transformed colors into HSB forms. The principle of HSB is represented by hue (H), brightness (S), and brightness (B). H represents a color in the range of 0 to 360, and the brightness is in the range of 0 to 100. The higher the number, the more intense the color. The brightness is also white in the range of 0 to 100. If it's 0, it's black, if it's 100, it's close to white. The part to be used in the HSB color format is H, and the expression for changing RGB to HSB was processed using mapreduce referring to the conversion formula in Figure 3.

$$H = \begin{cases} \text{undefined}, & \text{if } MAX = MIN \\ 60° \times \frac{G-B}{MAX-MIN} + 0°, & \text{if } MAX = R \text{ and } G \geq B \\ 60° \times \frac{G-B}{MAX-MIN} + 360°, & \text{if } MAX = R \text{ and } G < B \\ 60° \times \frac{B-R}{MAX-MIN} + 120°, & \text{if } MAX = G \\ 60° \times \frac{R-G}{MAX-MIN} + 240°, & \text{if } MAX = B \end{cases}$$

$$S = \begin{cases} 0, & \text{if } MAX = 0 \\ 1 - \frac{MIN}{MAX}, & \text{otherwise} \end{cases}$$

$$B = MAX$$

Fig 3. Transformation from RGB to HSB

III. Related research

1. Animated Movie Poster Colour Analysis

The project is the BOAZ 2014 project. This is a project that collects and analyzes poster images of 330 animated movies released between January 2000 and June 2014 by the Korean Film Promotion Commission and the country version of the movie through Google image search. The color of the poster was extracted through KSCA Korean standard color analysis. The extracted color information and movie information data were defective and analyzed. The analysis was analyzed by country and genre, and the analysis by genre showed that various colors such as strong red and blue were used in action, and that blue was overwhelming in melodrama and adult categories, and that there were also many pink categories. It was expected that more meaningful results would be obtained if big data technology was applied to advertising video analysis while confirming the meaningful data of the related study[4].

2. Techniques related to web crawling(Selenum, PhantomJS)

Web Crawling is the import of information from a computer's web into software technology. Pages linked to web pages can also be visited to extract data from multiple pages[5].

2.1 Beautifulsoup

The library that parses html is a library that allows users to parse code from html[6].

2.2 Selenum

It is a library provided by Selenum, and it helps the part that cannot be crawled only with Beautifulsoup by requesting the server with dynamic web page or login information.

2.3 PhantomJs

An interfaceless browser is imported from the website to source code, which can be run in JavaScript on the page to extract parts that do not exist in html.

3. degree of involvement

The degree of involvement is the importance of the product, service or object, the degree of perception, or interest. For example, how carefully a product is purchased when it is purchased. High-end products are products that are purchased while looking for a lot of information and thinking about in order to purchase the product, such as cars and home appliances. On the other hand, low-involvement products are products that can be purchased anytime, anywhere without having to search for large information such as beverages or food or think about it a lot. These involvement determinants are personal, product, and situational. First, the personal factor is that the interest in the product varies depending on the values and goals you have. Secondly, the product factor refers to the characteristics of the product, such as price. The higher the price, the higher the level

of concern consumers have, making it a high-intention product. The third is that the level of involvement varies depending on society's interest or situation due to situational factors. For example, fine dust is a big social problem recently, and this problem is causing interest in air purifiers[7].

4. K-means clustering

k-means clustering is an algorithm for grouping many data into groups of similar data. It is clustering by grouping data close to the center of the cluster. Here, the meaning of k denotes the number of clusters to divide. If there are three clusters, they form clusters among themselves close to any point[8].

5. OpenCv

It is a library developed by Intel that focuses on computer image processing. It can be used in various ways such as object, face, behavior recognition, and tracking. It can also be used on many platforms, including Windows and Linux[9].

6. HSV

Hue, Saturation, and Value are represented by 0 to 360 degrees, with 0 degrees red, 60 degrees yellow, 120 degrees green 180 degrees turquoise, 240 degrees blue, and 300 degrees red purple. Saturation is expressed as 0-100% and is the amount of color. The higher the number of saturation, the more intense the color, and the lower the number, the weaker the color. The brightness is 0-100% white. 0% is black and 100% is white[10].

IV. Analysis Results

The purpose of this paper is to analyze the correlation between consumers' 'involvement' and the color of advertising according to the category of advertising.

1. Color analysis based on engagement

Fig 4. Images of High Involvement Product Advertisement

Fig 5. Low Involvement Product Advertisement

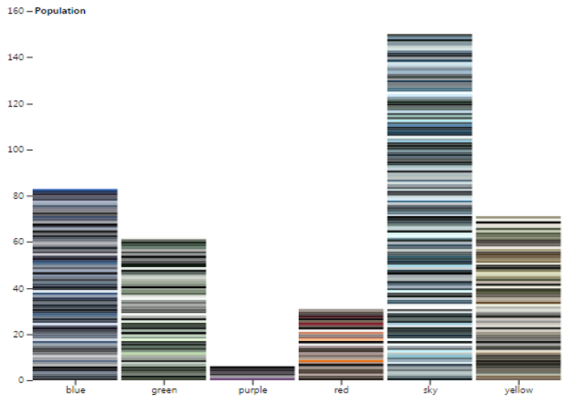

Fig 6. Color Extracted from Images of High Involvement Product Advertisement

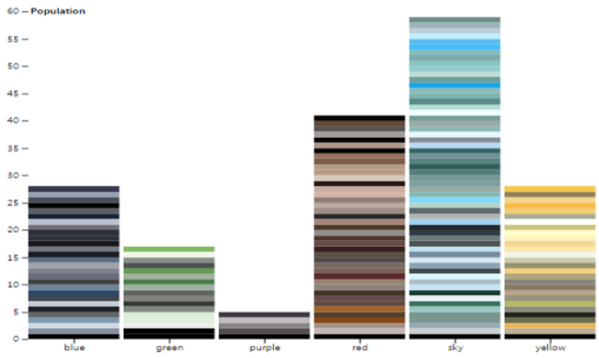

Fig 7. Color Extracted from Images of Low Involvement Product Advertisement

The image in Figure 4 is a high-intention product, and the image in Figure 5 is a low-intention product. Comparing the images in Figure 4 and Figure 5, it is the same kitchen background, but Figure 4 uses a lot of darker colors than Figure 5. On the other hand, Figure 5 showed a lot of relatively bright and bright colors. Prior to analyzing the results, the low involvement product was analyzed and visualized after expecting a brighter and brighter color than the high involvement product.

Figure 6, Figure 7, is a color extraction using OpenCv and k-means. Figure 6 is a high-contained product and Figure 7 is a low-contained product. For visualization, we extract representative colors based on HUE from color data that changed RGB to HSB via mapreduce. Red is 0-30, 330-360, yellow is 30-90, green is 90-150, sky is 150-210, and purple is 270-330 using mapreduce.

We take the processed data locally and draw a cumulative bar graph using d3js. If you check the stacked bar graph similar to the expected result, Compared to the low-involvement product, the high-involvement product extracted dark colors, and the low-involvement product showed relatively bright and bright colors. Although sky, and yellow tones could confirm the most distinct results, The same sky, yellow tone, but high-ranking sky was also drawn in many dark sky blue and black or gray tones, and yellow could be seen in many dark yellow or beige tones rather than light yellow. However, in the sky tone of the lower part of the body, one could see sky blue or bright sky blue close to fluorescence, and in the yellow tone, many bright yellow colors were pulled out.

2. Color Analysis by Category

This paper analyzes the correlation between category information and color, such as transport devices and computers in advertising.

2.1. Color Analysis by Category-Air Purifier

Fig 8. Images from Air Purifier Advertisement

The four images in Figure 8 are from different brands of advertising in the field of air purifiers. We could all identify many of the same green lineages.

If you check the advertisement, the color of the actor's clothes, the product or product background, and the background of the person are all green, and then The air purifier product was expected to be able to check the green side a lot.

Fig 9. Color Extraction From Air Purifier Advertisement

Figure 9 shows the color extracted from the advertisement for air purifiers in a donut graph and shows the TOP10 color in the air purifier field. As expected, we could confirm that the green line does not fill much in the TOP10 to show cleanness and purity, which are characteristics of air purifiers. It can also be seen that not only the green ones but also the clean and reliable sky blue ones were used a lot.

2.2 Color Analysis by Category 2-Air Conditioner

Fig 10. Images from air conditional advertisement

Figure 10 shows an image extracted from an air conditioner ad. 모It is an advertisement of two different brands, but the same as the air purifier, CG, person's clothing color props, and background could all be seen in

the sky blue and blue colors. After checking the advertisement video, I analyzed the results in anticipation that there would be many sky blue or blue colors.

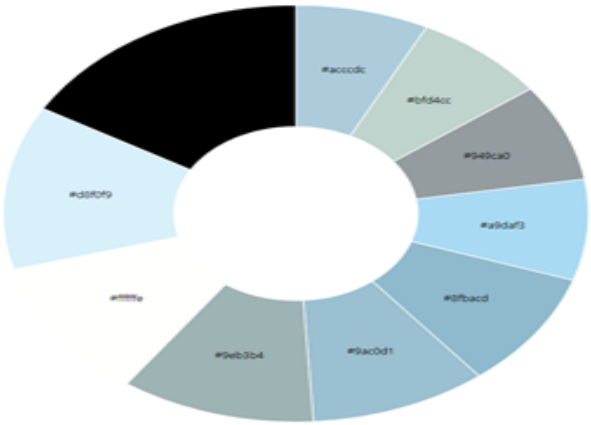

Fig 11. Color Extraction from air conditional advertisement

Figure 11 shows TOP10 of the category in a doughnut-type graph of colors extracted from air conditioning advertisements. I could see that sky blue was used a lot to show coolness which is characteristic of air conditioner. Compared to air purifiers, we could see that only the sky blue line was used more clearly.

2.3 Color analysis by category-toys

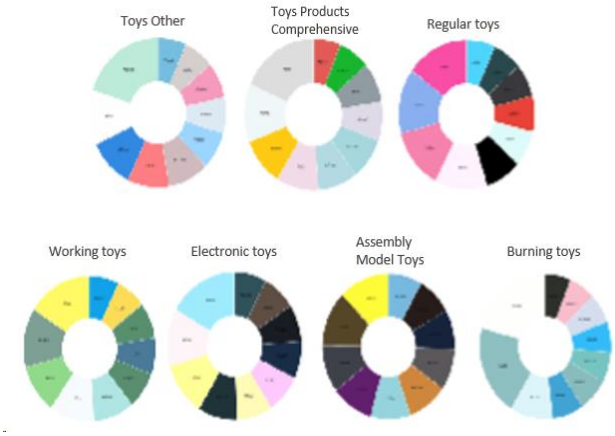

Fig 12. Color Extraction from Advertisement of Field of Toys

Figure 12 shows a doughnut-type graph of color extracted from advertisements for different types of toys. Pastel colors can be found mainly for young children's attention. Also, as can be seen in all categories of toys, it can be seen that many different colors are used, not just one color. In addition, in other areas, the black series could always be identified within the TOP10, but was relatively unused in the category of toys.

2.4 Color analysis by category-car

The analysis analyzed the color of each category of large/medium/subcompact/light cars divided by the size of the passenger car advertising. The launch price of large/medium-sized cars was usually between 25 million won and 50 million won. In addition, light and semi-compact cars could be found to have a market price of between 10 million won and 25 million. We extracted TOP 10 of each category with a doughnut-type graph of colors. Large/medium products, which are expensive, were dark and could be seen in many beige and black colors. On the other hand, it can be seen that light/subcompact cars, which are cheaper than large/medium-sized cars, have a lot of bright and active sky blue and blue colors. Similar to involvement, there are a large number of

products that are relatively expensive in high-intention products, and low-intention products have low-cost products. Earlier, according to the results of the high/low involvement, many dark colors were pulled out of the high involvement and many bright colors were pulled out of the low involvement products. Applying the cost-effective aspect of the passenger car category in question, it can be seen that dark colors are extracted from large/medium car advertisements, and lighter colors are extracted from light/submedium car advertisements, similar to analysis.

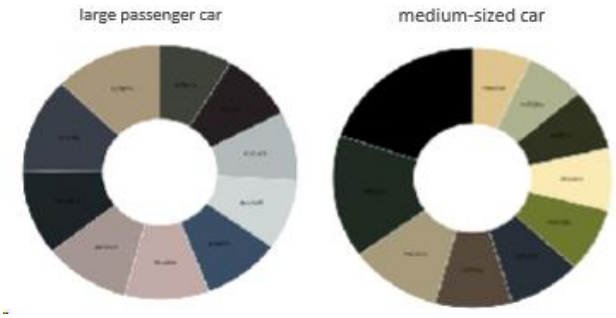

Fig 13. Color Extracted from Large/Medium Passenger Cars

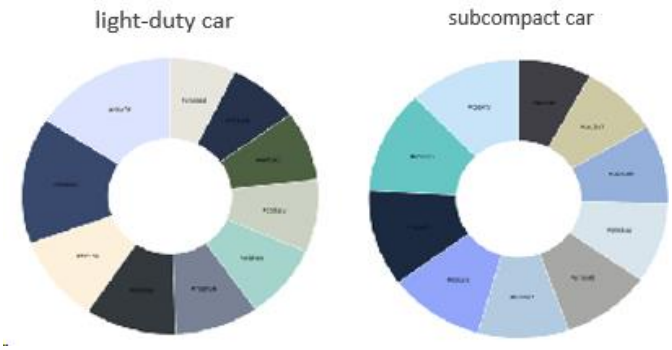

Fig 14. Color Extracted from Light / Semi-heavy Passenger Cars

V. Conclusion

In this paper, images were extracted from advertisements provided by the Korea Advertising Federation Advertising Information Center Creative using web crawling. We extract colors from extracted images with OpenCv using the K-means clustering technique to identify color marketing strategies inherent in advertising according to their involvement and product types.

In past research, color has been used in marketing to stimulate consumer sentiment. Colors reflect the unique image of the brand, stimulating consumers' sense of belonging, allowing them to consume the values and images of the color. As such, color has a great influence on marketing.

As introduced earlier in the introduction, ads made by artificial intelligence advertising creative directors did not differ much in terms of ads and preferences made by human advertising creative directors. If the results of the analysis in this paper and color marketing using AI are combined, we expect to establish a better effective advertising strategy than before.

REFERENCES

[1] Min-Young Jeong and Mi-Na Jo(2011), “The Effect of Hospitality Viral Marketing Perception on Satisfaction after using Viral Marketing," Journal of Tourism and Leisure Research, 23(3):359-376

[2] OyungerelGankhuyag, Ho-Gyu Choi, Woo-Geuk Shim, Nam-Goo Park(2013), “A Study on the Influence of Color Marketing on Brand Awareness and Brand Attitude : Focus on Mongolian Cashmere Brand,” Korean Review of Corporation Management, 4(2):45-67

[3] Young Ran Cha(2018), “Artificial Intelligence Strategy for Advertising and Media Industries : Focused on In-depth Interviews,” Journal of The Korea Contents Association, 18(9):102-115
[4] Gui-Yeon Han, Seung-Hee Oh, Jin-HeeChoi(2014), “A Comparative Analysis of Space Image and Colors Used across Brand Coffee Chains - Focused on Angelinus, Caffé Bene, Coffee Bean, Starbucks -,” Journal of Korea Society of Color Studies, 28(1):219-228
[5] Kyung-Seob Yoon, Yeon-Hong Kim(2018), “Designing and implementing web crawling-based SNS web site,” Proceedings of the Korean Society of Computer Information Conference, 26(1):21-24
[6] Seong-Hwan Cho(2018), “A Study on Analysis of the Trend of Blockchain by Key Words Network Analysis,” Journal of Korea Institute of Information, Electronics, and Communication Technology, 11(5):550-555
[7] Se-Hoon Park, Eun-Jung Kang(1999), “A Study on the Effectiveness of Two - Sided Ads ; The Role of Type of Attribute , Correlation , and Involvement,” Journal of Korean Marketing Association, 14(3):1-24
[8] Sun Park, Yeon-woo Lee, Chun-Sik Shim, Seong- Ro Lee(2012), “Generic Document Summarization using Coherence of Sentence Cluster and Semantic Feature,” Journal of the Korea Institute of Information and Communication Engineering, 16(12):2607-2613
[9] OpenCv, https://opencv.org/
[10] Ho-SikPark(2015), “Vehicle Tracking System using HSV Color Space at nighttime,” Journal of Korea Institute of Information, Electronics, and Communication Technology, 8(4):270-274