Profiling Artists and Their Paintings Using Association Rules - Part 2

Allina Dolor, James Robinson, Cherian Thomas, Matthew Wise

I. Association Rules

The data set used to obtain association rules consists of approximate RGB color values used in a set of paintings. These RGB values were collected from scanned images of these painting found on the websites of *The Museo Nacional Centro de Arte, Reina Sofia* and *The Van Gogh Museum of Amsterdam*. These image scans were then processed to potentially discover interesting relationships between their colors. Association rule were mined and used in determining color relationships to find which color combinations used by artists in different paintings, which had the potential to express some meaningful insight into the dataset. Orange, the open-source machine learning tool, was used as the primary method of efficiently obtaining these association rules to gauge the relationship between the colors. The association rules obtained are shown in full through the separately attached zipped folder called “rules”.

II. Interestingness of a Set of Rules

From the set of gathered association rules, we considered the values of each to determine the most interesting ones from the set. The rules chosen as most interesting are displayed in a separate plain text file. The file contains a table with the corresponding values of support, confidence, and lift for a given artists.

Given the problem domain (that is, the unique identification of a single artist), rules are made interesting not by their content but by their distinctiveness. For this reason, confidence was not directly considered (support was used as a threshold for consideration, as rules with very low support are unlikely to be reliable). Instead, the value of confidence (appropriately binned for use as categorical data) was treated as part of the rule for the psuedo-apriori meta-processing described below. Lift, however, was not a factor, as an artist's distinctive color combination might be used in only a small portion of some paintings while remaining just as significant.

We measured distinctiveness using, as mentioned above, a psuedo-apriori algorithm. What we want to know is how often a certain artist is associated with a certain color association (that is, a rule from traditional association rules processing of each image). This is, essentially, an association rules problem with the following constraints:

1. We're only interested in rules where the antecedent is exactly one artist
2. Each transaction contains exactly two items (the artist, and the rule, plus its associated confidence).
3. Support should be scaled relative to the number of works the artist has in the dataset, rather than the number of total transactions (each rule occurs either 0 or 1 times per work, and it'd throw off the results to allow artists with more works in the set higher support for that reason alone).

Given these constraints, a much simpler variant of the algorithm can be employed. Essentially, the psuedo-apriori algorithm finds the proportion of works of each artist in which each rule is employed. Absent rules are considered to have a support of 0, as they could not reliably be shown in the original dataset. Keep in mind, the goal is not wholly accurate descriptions of the associations among colors, but the development of a distinctive profile.

After applying the psuedo-apriori algorithm, the euclidean distance between the support of each rule for each artist and the average for that rule was used as the determinant of interestingness. The ten most interesting rules for each artist are contained in the included meta_rules.txt file. Note that rules with "N/A" in the confidence, lift, and support columns are interesting because they're absent; that is to say, that well the average artist uses those combinations, the artist in question doesn't, which can be just as distinctive as what they do use.

III. Recommendations

The association rules obtained effectively evaluate the relationship between multiple colors. Through these rules, we are able to make suggestions and recommendations to our client while ensuring the validity and worthiness of our results. The table containing the association rules allows a comparison between the original art and forged art. Picasso frequently associated the colors 78, 128, 128 to the colors 128, 128, 128. Another artist, Zubiaurre associated the colors 76, 76, 76 to 128, 128, 128. Other distinctive associations are laid out in the above mentioned meta_rules.txt file. Our recommendation to any client is that they compare the rules produced from the suspected forgery from those found to be most distinctive for the alleged artist. As several distinctive rules were found for each of the examined artists, the absence of these rules from the suspected forgery should indicate that it is in fact fake, and their presence the reverse. For a more complete measurement, the complete body of the alleged artists work could be sampled and run through the psuedo-apriori algorithm, yielding a vector representation of the artist's style (e.g., the basis set formed by the set of all possible rules with components equal to the respective support). Take the Euclidean distance from this to the suspected forgery's rules, and you have a quantification of the how much that work diverges from the artist's traditional fare.

IV. Other Applications

Our dataset focused on different digital scans of paintings by the following artists: Massanet, Schnabel, Monet, Picasso, van Gogh, and Zubiaurre. Often, famous paintings are interpreted and analyzed visually to understand how they communicate emotions or evoke a certain feeling from the viewer. Through association rule data mining, patterns and relationship in and between a set of painting can be discovered to find the interesting relationships between colors that allows for these feelings to be conveyed. Art analysis delves in the patterns and relationships of colors to convey an interpretation of a given painting.

Color theory provides direction for color mixing and showcases the effects of specific color combinations. Color context is a category within color theory that examines how colors interact in relation to different colors. People may notice that some colors complement others better than other combinations. In contrast, other color combinations may not be as aesthetically pleasing. Our data set and data mining processes could potentially be used by art historians and art enthusiasts as a way to identify what colors certain artists use to complement each other or to aid in conveying a certain feeling in their art. By reviewing the association rules, one can determine the frequent colors used with its specified confidence and support rates as well as the frequent color combinations. Ultimately, our data set with the given association rules could potentially help expand the methods used to analyze and interpret a piece of art.

V. Software Limits and Advantages

Orange was primarily used to extract association rules from our dataset. However, Orange is very limited in its application to this sort of dataset. A series of steps had to be taken before and after the rule extraction to acquire useful results. First, the images were resized such that a pixel represents a fixed area, eliminating variance introduced by the particulars of the digitization process. Second, the colors were binned for treatment as categorical data. The number of possible colors was reduced from 256^3=16777216 to a much more reasonable 4^3=64. Third, the image was sliced into 16 by 16 boxes, forming the set of transactions for processing. Fourth, the RGB values for each pixel in each box were extracted and added to an Orange-compatible .basket file. Fifth, the baskets were run through Orange's association rules module. Sixth, the rules were saved to an intermediary data structure that could be serialized and saved to disk to avoid repetition of the lengthy extraction process. Seventh, the psuedo-apriori algorithm was applied. Eighth, the interestingness was measured as described above.

Orange's primary advantages in our case were its convenient python bindings and powerful interface for rule expression and filtering. Its inability to do the rest of what we required is hardly a failing as much of it falls well outside its scope. However, an interface for comparing rules from different baskets would have been quite useful, and is arguably a useful feature for other association rule problems as well. While individual image data could be viewed and sorted to compare their values, it was difficult for us to see how these values persisted or changed when considering multiple images. In all, it performed its piece of the process effectively and conveniently, which is all that could be expected of it.

References

Massanet, Joan. *"Composición Surrealista."* 1928. Museo Nacional Centro De Arte, Reina Sofia. Web. 28 Feb. 2017. <http://www.museoreinasofia.es/en/collection/artwork/composicion-surrealista-surrealist-composition.>

Massanet, Joan. *"Nacimiento De Venus."* 1927. Museo Nacional Centro De Arte, Reina Sofia. Web. 28 Feb. 2017. <http://www.museoreinasofia.es/en/collection/artwork/nacimiento-venus-birth-venus.>

Monet, Claude. *"Tulip Fields near The Hague."* 1886. The Van Gogh Museum, Amsterdam. Web. 28 Feb. 2017. <https://www.vangoghmuseum.nl/en/collection/s0530N2012.>

Monet, Claude. *"Windmills near Zaandam."* 1871. The Van Gogh Museum, Amsterdam. Web. 28 Feb. 2017. <https://www.vangoghmuseum.nl/en/collection/s0503S2001.>

Picasso, Pablo. "*Figure."* 1928. Museo Nacional Centro De Arte, Reina Sofia. Web. 28 Feb. 2017. <http://www.museoreinasofia.es/en/collection/artwork/figure.>

Picasso, Pablo. *"Buste et palette (Bust and Palette)"*. 1925. Museo Nacional Centro De Arte, Reina Sofia. Web. 28 Feb. 2017 <http://www.museoreinasofia.es/en/collection/artwork/buste-et-palette-bust-and-palette.>

Picasso, Pablo. *"Instruments de musique sur une table (Musical Instruments on a Table)"*. 1924. Museo Nacional Centro De Arte, Reina Sofia. Web. 28 Feb. 2017. <http://www.museoreinasofia.es/en/collection/artwork/instruments-musique-sur-une-table-musical-instruments-table.>

Picasso, Pablo. *"Nature morte (Still Life)"*. 1944. Museo Nacional Centro De Arte, Reina Sofia. Web. 28 Feb. 2017. <http://www.museoreinasofia.es/en/collection/artwork/nature-morte-still-life-1.>

Picasso, Pablo. *"Guernica"*. 1937. Museo Nacional Centro De Arte, Reina Sofia. Web. 28 Feb. 2017. <http://www.museoreinasofia.es/en/collection/artwork/guernica1.>

Schnabel, Julian. *"Al Pueblo De España (To the People of Spain)." Museoreinasofia.es*. N.p., 1991. Web. 28 Feb. 2017. <http://www.museoreinasofia.es/en/collection/artwork/al-pueblo-espana-people-spain.>

Schnabel, Julian. *"Al Pueblo De España (To the People of Spain 0)." Museoreinasofia.es*. N.p., 1991. Web. 28 Feb. 2017. <http://www.museoreinasofia.es/en/collection/artwork/al-pueblo-espana-people-spain.>

Schnabel, Julian. *"Al Pueblo De España (To the People of Spain 1)." Museoreinasofia.es*. N.p., 1997. Web. 28 Feb. 2017. <http://www.museoreinasofia.es/en/collection/artwork/al-pueblo-espana-people-spain.>

Schnabel, Julian. *"Untitled (La Banana E Buona) (Untitled [The Banana Is Good])." Julian Schnabel*. Museoreinasofia.es, 2014. Web. 28 Feb. 2017. <http://www.museoreinasofia.es/en/collection/artwork/untitled-banana-e-buona-untitled-banana-good.>

van Gogh, Vincent. *"A Crab on its Back."* Aug.-Sep. 1887. *The Van Gogh Museum*, Amsterdam. Web. 28 Feb. 2017. <https://www.vangoghmuseum.nl/en/collection/s0124V1962.>

van Gogh, Vincent. *"Basket of Pansies."* May 1887. *The Van Gogh Museum*, Amsterdam. Web. 28 Feb. 2017. <https://www.vangoghmuseum.nl/en/collection/s0180V1962.>

van Gogh, Vincent. *“Irises.”* May 1890. *The Van Gogh Museum*, Amsterdam. Web. 28 Feb. 2017. <https://www.vangoghmuseum.nl/en/collection/s0050V1962.>

van Gogh, Vincent. *“Tree Roots.”* July 1890. *The Van Gogh Museum*, Amsterdam. Web. 28 Feb. 2017. <https://www.vangoghmuseum.nl/en/collection/s0195V1962.>

van Gogh, Vincent. *“Wheatfield with Crows.”* July 1890. *The Van Gogh Museum*, Amsterdam. Web. 28 Feb. 2017. <https://www.vangoghmuseum.nl/en/collection/s0149V1962.>

Zubiaurre, Valentín. *“A las doce (At Twelve).”* 1907. Museo Nacional Centro De Arte, Reina Sofia. Web. <http://www.museoreinasofia.es/en/collection/artwork/doce-twelve.>

Zubiaurre, Valentín. *“Versolaris (Basque Verse Composers).”* 1913. Museo Nacional Centro De Arte, Reina Sofia. Web. 28 Feb. 2017. <http://www.museoreinasofia.es/en/collection/artwork/versolaris-basque-verse-composer>

Zubiaurre, Valentín. *“Autoridades del pueblo (Village Authorities).”* 1920. Museo Nacional Centro De Arte, Reina Sofia. Web. 28 Feb. 2017. <http://www.museoreinasofia.es/en/collection/artwork/autoridades-pueblo-village-authorities.>

Zubiaurre, Valentín. *“Fiesta en Segovia (Fiesta in Segovia).”* 1919. Museo Nacional Centro De Arte, Reina Sofia. Web. 28 Feb. 2017. <http://www.museoreinasofia.es/en/collection/artwork/fiesta-segovia-fiesta-segovias.>

Zubiaurre, Valentín. *“Hombres de Castilla (Men from Castile).”* 1920. Museo Nacional Centro De Arte, Reina Sofia. Web. 28 Feb. 2017. <http://www.museoreinasofia.es/en/collection/artwork/hombres-castilla-men-castile.>