Predicting Semantic Signatures of Fonts

Tugba Kulahcioglu
Rutgers University
Piscataway, NJ, USA
tugba.kulahcioglu@rutgers.edu

Gerard de Melo*
Rutgers University
Piscataway, NJ, USA
gdm@demelo.org

***Abstract*—Towards the aim of semantic font recommendation, we first analyze the relationship between fonts and semantic attributes using a crowdsourced dataset. We deepen this analysis at the level of font categories and font styles, including via a series of interactive visualizations of the relationships between multiple dimensions of this data. Subsequently, we induce semantic signatures for a large number of fonts by computationally predicting attribute values using a k-NN based approach. We evaluate the effectiveness of this approach, and, through a visual exploration, categorize semantic attributes into three groups based on their potential to be conveyed by the fonts. The resulting resource is made available, and we aim for the findings, visualizations, and data to benefit studies that computationally support the challenging but critical process of font selection.**

I. Introduction

It is widely agreed that different typefaces and fonts not only differ with respect to their visual characteristics, but that they also possess different *personas* in terms of their perceived associations and connotations [1], [2], [3]. These range from perceptions of *attractiveness* to evoked affective states such as *happiness* all the way to associations with *confidence* or even *laziness*. Understanding these latent semantic connections is a crucial precondition to using fonts adequately and effectively, as they may affect the perception of a product that the text is describing [4], [5], with consequences ranging from longer response times [6] to product success [7].

There have been user studies aiming to shed some light on the perception of fonts [2], [8], [5], [1], [9]. These have revealed intriguing links, such as the association of sweet taste with certain visual font attributes. However, these studies have been conducted at a small scale. This contrasts with the overwhelming variety of fonts that are now available. For instance, one can quickly collect 50,000 fonts by crawling the web for a few days.[1]

In this study, we aim to induce *semantic signatures* that characterize the semantic attributes of large numbers of fonts.

One way to assess a wider range of fonts is to draw more general conclusions by unearthing connections and trends that apply broadly at a higher level, as in the study that found an association between *sweet* taste and *round* typefaces [8]. To this end, we first analyze the relationships of font categories and font styles with semantic attributes to determine to what extent simple correlations may exist. We find that

Fig. 1: Attributes used in this study, visualized using the font that is predicted to have the highest congruency[3].

certain attributes are indeed manifested most predominantly in specific categories and styles of fonts. However, this does not hold in general, as the interactions of visual features and semantic attributes are not always straightforward. and Thus, this generic approach is not always feasible.

As a second step, overcoming the aforementioned challenges, we develop a semi-automatic computational approach to predict semantic signatures of fonts based on a small set of seed data, with an average error rate of just 8%. With this method, we extend crowdsourced data for 200 fonts to a large dataset, currently covering 1,883 fonts, which we make available online. Figure 1 presents the attributes used in this study, visualized using the fonts found to be most congruent by the extended dataset, excluding the ones from the crowdsourced seed data.[2]

We then proceed to analyze the resulting semantic resource, with the aim of assessing its quality as well as to derive insights regarding the potential of fonts to exhibit desired semantic attributes. The interactive visualization used to carry out this analysis is also made available online.

The paper is organized as follows. In Section II, we review related work. In Section III, we analyze the semantic signatures of fonts based on font categories and styles. In Section IV, we present the method we use to induce signatures for an extended set of fonts, and evaluate its performance. Following this, in Section V, we analyze the resulting dataset. We provide a discussion in Section VI and then conclude in Section VII.

*Gerard de Melo's research is supported by the DARPA SocialSim program.

[1] https://erikbern.com/2016/01/21/analyzing-50k-fonts-using-deep-neural-networks.html

[2] The attribute *attention-grabbing* is shortened as *attention*.

[3] Aiming to showcase different fonts, we used the next most congruent font if the most congruent font is already used for another attribute.

0-7695-6360-0/18/$31.00 ©2018 IEEE
DOI 10.1109/ICSC.2018.00025

II. Related Work

A. Fonts and Semantic Attributes

Previous studies analyze the relationship between fonts and semantic attributes (also referred to as *persona*). The most recent such study that we are aware of, by O'Donovan et al. [2], associates 200 fonts with 37 semantic attributes. Users are requested to pick one of two presented fonts for a given attribute, and then based on the user choices, scores between 0 and 100 are assigned for each font–attribute pair. The resulting dataset is publicly available[4] and is utilized in this study.

While O'Donovan et al. [2] do not investigate the relationship between the underlying properties of typefaces or fonts and semantic attributes, there are previous studies that have shed some light on this relationship. Shaikh et al. [3] carries out an online survey to associate 20 fonts with 15 adjective pairs (e.g., happy and sad). They analyze the findings based on font categories and discover correlations between categories and semantic attributes (e.g., monospace is found to have a strong association with the attributes *dull*, *plain*, *unimaginative*, and *conforming*). Similarly, [9] analyzes correlations of the design characteristics and affective attributes for 36 typefaces, and they come up with some design guidelines. For instance, for a *pretty* affect, the uppercase *F* character should be designed with its lower bar above the middle.

There are a few other studies that investigate such relationships. Brumberger [1] carries out a user study to associate fonts with 20 semantic attributes such as *cold* and *loud*, and generates groups of fonts sharing certain semantic qualities, namely *elegance*, *directness*, and *friendliness*. Other studies [8], [10] analyze the relationship between visual font characteristics and taste attributes (*sweet*, *sour*, etc.) through user studies. One of their findings is that *round* fonts are associated with *sweet* taste. Also, many font-focused websites allow users to tag the fonts they upload with various kinds of labels, including semantic attributes.[5]

B. Font Recommendation and Pairing

O'Donovan et al. [2] aim to recommend fonts to the user that are similar to that user's current font. The authors carry out a user study that asks users to pick the most similar font to a reference font. They subsequently use this data to develop a similarity metric. Wang et al. [11] rely on a deep learning approach to handle the same task. Although the results are not not numerically verified, from the comparison of test results, it is claimed that this approach produces better results than the former approach [2].

In FontJoy [12], font pairings are generated using vector representations. The aim is to find fonts that are both contrasting and complementary. The system can generate new pairs, or select a second font given an already specified one. Their vector representations are provided online.

[4] http://www.dgp.toronto.edu/~donovan/font/

[5] Such as: www.fontsinuse.com, www.dafont.com, www.1001fonts.com

C. Impact of Font Choices

Previous studies investigating the impact of fonts use Stroop or survey-style studies. Stoop-style studies ask users to respond to a given task both correctly and as quickly as possible, while measuring the response time. Lewis & Walker [6] asked users to press the left hand key if the words *slow* or *heavy* appear, and the right one if *fast* or *light* appears. They repeat such tasks with congruent fonts (matching the underlying meaning or theme) and incongruent ones, finding that congruent fonts significantly reduce the response times. Hazlett et al. [13] asked users to assess whether a displayed word is positive or negative. Congruent fonts were found to result in faster responses.

Survey-style experiments gather user ratings for semantic measures. Juni & Gross [4] present two New York Times articles with two different fonts and solicit ratings from users. The results reveal that the same text is perceived as being funnier or angrier when read in a certain font compared to another. Shaikh [5] presents documents in three different fonts (congruent, incongruent, neutral), asking users to assess the personality of the document (e.g., exciting) and the personality of the author (e.g., trustworthy). The findings show strong effects for all three categories of fonts on the perceived personality of documents, whereas congruent and neutral fonts created similar perceptions of the authors' respective personalities. Hazlett et al. [13] displayed the same page with different fonts for 0.7 seconds each, asking users to describe the emotional tone of the page. They found that the latter is strongly influenced by the font type.

Shaikh et al. [14] investigate the effect of fonts on the perception of email and find that fonts with low congruency result in different perceptions of an email than when higher congruency fonts are invoked. A similar study on the perception of a company website [15] reveals that neutral/low congruency fonts negatively affect a company's perception in terms of professionalism, believability, trust, and intent to act.

Many studies in marketing analyze font effects, especially in packaging design. Fligner [7] shows that fonts associated with the attribute *natural* increase the perception of products with respect to *healthfulness* especially if the products' intrinsic (e.g. fat-free) and extrinsic cues (e.g. sold at Whole Foods Market) also support this. The experiments by Childers and Jass [16] show that semantic attributes of fonts affect user perception for both high and low engagement levels; and the effect of a font on the recall performance increases when other factors such as the picture used is consistent with the font. Through experiments using bottled water of a fictional brand, Van Rompay and Pruyn [17] as well found that the congruence between fonts and other design elements influence the perception of brand credibility, aesthetics, and value.

Overall, these studies show that selecting congruent fonts has significant impact on how the content, its authors, and associated entities such as products are perceived. Hence, it is crucial to develop techniques that aid in determining the semantic congruency of fonts.

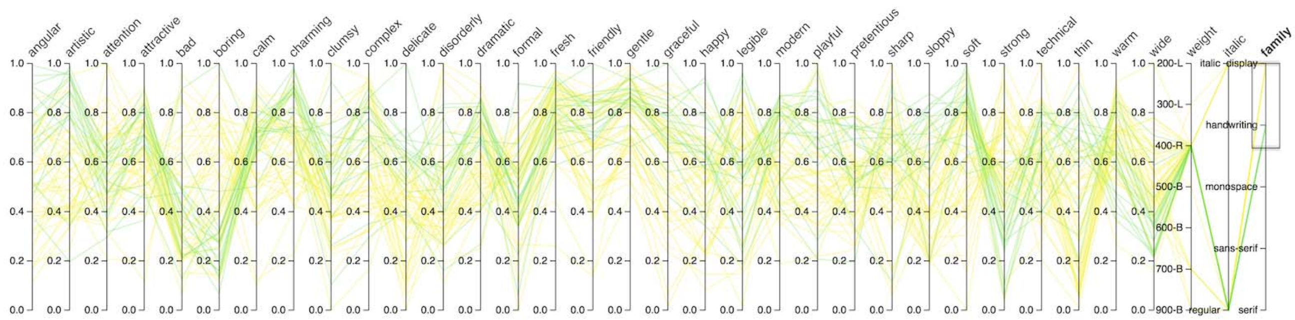

Fig. 2: Semantic signature of the font categories *display* and *handwriting*. This shows the font association values for semantic attributes, and font categories and styles. Each line is colored based on the font categories (*display*: yellow, *handwriting*: green).

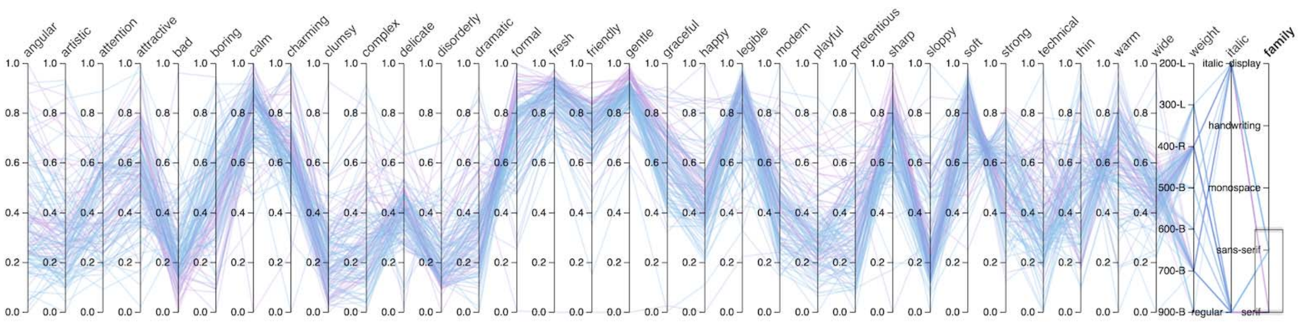

Fig. 3: Semantic signature of the font categories *serif* and *sans-serif*. This shows the font values for the semantic attributes, and font categories and styles. Each line is colored based on the font categories (*sans-serif*: blue, *serif*: pink).

TABLE I: Summary of statistics for the font categories. (SD: Standard Deviation, Att: Attribute, Avg: Average)

		Average		Max		Min	
Category	Count	Avg.	SD	Avg.	Att.	Avg.	Att.
display	45	0.56	0.18	0.80	fresh	0.37	delicate
handwriting	18	0.63	0.15	0.84	gentle	0.29	boring
monospace	8	0.48	0.16	0.81	gentle	0.29	modern
sans-serif	85	0.49	0.14	0.87	gentle	0.24	clumsy
serif	44	0.53	0.16	0.89	gentle	0.19	bad

III. Semantic Signatures of Font Categories and Styles

Our first goal is to expose general associations between font categories or styles [18] and semantic attributes. As a starting point, we consider the crowdsourced data by O'Donovan et al. [2], which associates 200 fonts with 37 semantic attributes (e.g. *happy*, *formal*). We normalize their ratings to the [0,1] range, and derive attributes at the level of font categories, as well as for italic emphasis and font weights, to analyze their relationship with different attributes[6]. The results of this process are depicted in Figures 2, 3, 4, and 5. These plots are taken from an interactive visualization that we have established for the analysis of this data and made available online[7].

[6]We exclude 6 typographic attributes and use the remaining 31 attributes.

[7]Supplementary material can be accessed via http://gerard.demelo.org/fonts/

A. Font Categories

We begin by analyzing the relationship of five coarse-grained font categories.[8] Table I provides a high-level summary of these categories. Although both the averages and maximum scores appear to be close, and the scores of the respective attributes with the maximum average as well, a cursory glance at Figures 2 to 5 reveals that the distributions diverge significantly between particular font categories.

1) Display: Following the highlighted (yellow) lines in Figure 2, we observe that the *display* category appears to have the most scattered attribute scores. Across nearly all considered attributes, we find that its scores lie in a high range. This is also reflected in the summary table with a relatively high standard deviation value of attribute averages.

2) Handwriting: Figure 2 reveals that handwriting fonts appear to show a trend rather different from those of other font categories, especially for the attributes *artistic*, *charming*, *complex*, *dramatic*, *modern*, and *playful*, they score higher than others. The category also has strong associations for

[8]These categories reflect historical origins and typographic properties. *Handwriting* typefaces are designed to create the impression of being hand-rendered. The characters of *monospace* typefaces occupy equal horizontal space. *Serif* typefaces have small lines attached to the end of the strokes in its characters, whereas *sans-serif* denotes typefaces lacking those attached lines. *Display* typefaces do not share typical typographic qualities other than a low level of legibility when used for body text, so they are reserved mostly for headings and other kinds of display purposes.

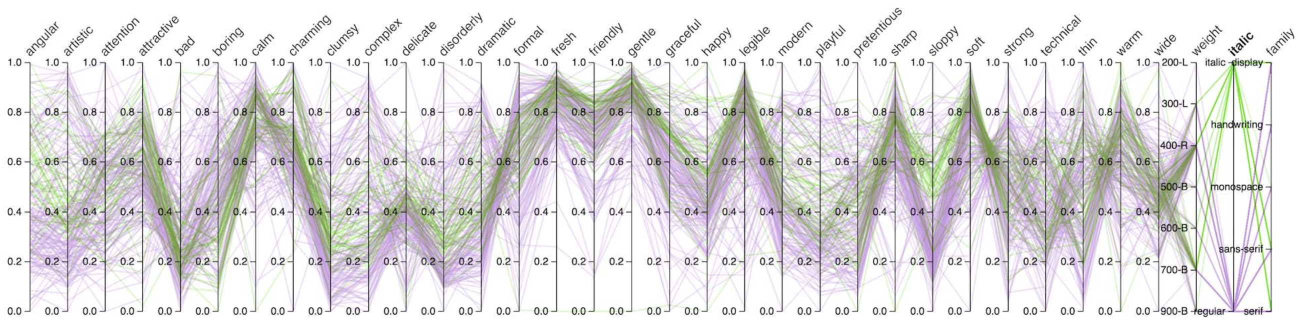

Fig. 4: Semantic signature of the font style *italic*. This shows the font values for the semantic attributes, and font categories and styles. Each line is colored based on the italic emphasis of the fonts (*italic*: green, *regular*: pink).

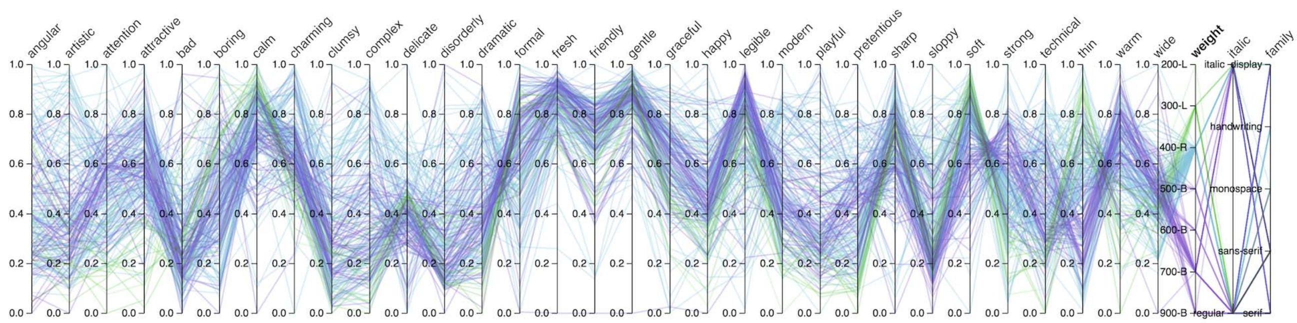

Fig. 5: Semantic signature of the font style *weight*. This shows the font values for the semantic attributes, and font categories and styles. Each line is colored based on the weights of the fonts (*light*: green, *normal*: blue, *bold*: purple).

the attributes *fresh*, *friendly*, and *gentle*, which accords with the general trend. For *boring* and *strong*, in contrast, it has particularly low scores.

3) Monospace: Monospace fonts make up a very small number of instances in the data, only 8 out of 200. Their curve follows the general trend for *fresh* and *gentle* and *calm*, while having atypically high values for *boring*.

4) Sans-serif: This is the largest category in the dataset, with 85 members. The distribution of attributes can be analyzed in detail in Figure 3. We observe high association scores for *calm*, *formal*, *fresh*, *friendly*, *gentle*, *legible*, and *soft*, whereas for *bad*, *clumsy*, *disorderly*, *playful*, and *artistic*, we encounter lower values. The strongest association is for the attribute *gentle*, while the lowest score is seen for *clumsy*.

5) Serif: The serif font category has 44 samples in the dataset, and follows a similar pattern as *sans-serif*, except for showing slightly higher associations for *formal*, *gentle*, *friendly*, *happy*, and *sharp*, and slightly lower values for *bad*. Its highest score is for the attribute *gentle*, just as for sans-serif, while its lowest is for the attribute *bad*.

B. Font Styles

Table II summarizes statistics for the font style properties that we analyze: *italic* emphasis and *weight*. With the exception of the *light* font weight style, the values are very similar across all styles.

TABLE II: Summary of statistics for the font styles.
(SD: Standard Deviation, Att.: Attribute, Avg.: Average)

		Average		Max		Min	
Style	Count	Avg.	SD	Avg.	Att.	Avg.	Att.
italic	42	0.56	0.16	0.87	gentle	0.26	bad
regular	158	0.52	0.18	0.84	gentle	0.32	bad
bold	59	0.52	0.15	0.86	fresh	0.27	disorderly
normal	127	0.54	0.18	0.85	gentle	0.31	bad
light	14	0.44	0.11	0.91	soft	0.14	pretentious

Figure 4 plots the distributions for fonts with *regular* (158 samples) and *italic* (42 samples) styles. For the attributes *artistic*, *complex*, *disorderly*, *dramatic*, and *playful*, *italic* seems to have mid-range values, whereas the *regular* style constitutes the high and low peaks. They both seem to peak for those attributes that also exhibit a general trend of having high values for the fonts in our data, such as *calm*, *fresh*, *gentle*, and *legible*. We found that among the font categories, *serif* fonts suffer the greatest impact when this style property is applied. The most *charming*, *attractive*, and *happy serif* fonts, for example, all use *italic* forms.

Figure 5 plots the distributions for fonts with different weights. Weights below 400 are considered *light* (14 samples), whereas the ones above 400 are considered *bold* (59 samples). The *normal* weight is assumed as 400 and consists of 127 samples for our data. The attributes *thin*, *soft*, and *calm* appear to have high values for fonts in a *light* style. Similarly, the attribute *warm* correlates with the *bold* style. They all seem to have peaks for the attributes *legible* and *gentle*. The least *happy* fonts are those that are *light*. Further analysis reveals that *sans-serif* shows strong interactions with *weight*, e.g., the *calmest* and *softest sans-serif* fonts use *light* forms, whereas *warm* and *legible* fonts use *bold* forms.

IV. Large-Scale Semantic Signature Induction

We now proceed to produce a much larger-scale database of semantic signatures.

A. Method

We assume as input a set of fonts $\mathcal{F}$ described in terms of a set of font attributes $\mathcal{A}$. For this, we again rely on the previously used crowdsourced data by O'Donovan et al. [2], which describes a small set of 200 fonts. For a given font $f \in \mathcal{F}$, it provides scores in $[0, 100]$ for each attribute $a \in \mathcal{A}$. From this data, we derive $|\mathcal{A}|$-dimensional vectors $\vec{f} \in [0, 1]^{|\mathcal{A}|}$ for each $f \in \mathcal{F}$, by transforming the dataset to consider the attributes for a given font while normalizing scores to $[0, 1]$.

Our aim is to predict $\vec{f'}$ for fonts $f' \notin \mathcal{F}$. To achieve this, we use k-nearest neighbors (k-NN) regression. The distance between two fonts, denoted as $d(f_i, f_j)$ is calculated using one of the similarity metrics described in the following subsection.

The unweighted k-NN approach uses the following formula, where $\vec{f_1}$ to $\vec{f_k}$ are attribute vectors for the closest k fonts in $\mathcal{F}$ according to a similarity metric.

$$\vec{f} = \frac{1}{k} \sum_{i=1}^{k} \vec{f_i} \tag{1}$$

The weighted k-NN approach generates weights using the following equation.

$$w_i = \frac{1}{k-1} \frac{\sum_{\substack{j=1 \\ i \neq j}}^{k} d(f', f_j)}{\sum_{j=1}^{k} d(f', f_j)} \tag{2}$$

Subsequently, the weighted values are generated as follows:

$$\vec{f} = \sum_{i=1}^{k} w_i \vec{f_i} \tag{3}$$

B. Similarity Measures

To compute nearest neighbors, we consider four similarity metrics as alternatives.

The first option is to use typographic properties, obtained by parsing a font's glyph outlines to extract italics, thickness, size, area, orientation, stroke width, and spacing. We rely on existing data for this [2]. For some of these features, the data provides an average for all the characters, whereas for others, only selected characters are used.

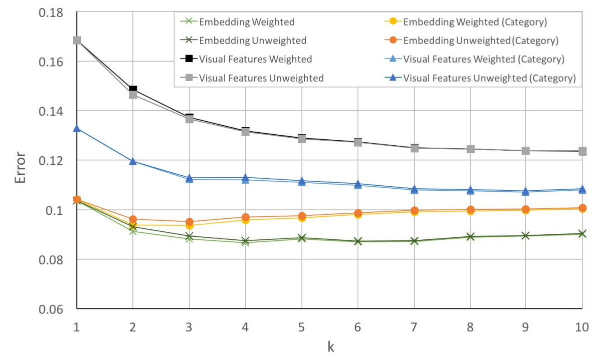

Fig. 6: Error plots of four method–similarity measure combinations using different k values.

TABLE III: Error averages for each attribute.

attribute	e	attribute	e	attribute	e
fresh	0.051	strong	0.086	boring	0.097
gentle	0.051	attention	0.086	playful	0.098
delicate	0.057	bad	0.087	formal	0.099
wide	0.059	modern	0.089	warm	0.101
charming	0.062	legible	0.091	thin	0.101
friendly	0.063	disorderly	0.094	sharp	0.106
calm	0.072	attractive	0.095	angular	0.107
soft	0.076	dramatic	0.095	complex	0.108
graceful	0.077	pretentious	0.096	technical	0.111
sloppy	0.081	artistic	0.097		
happy	0.082	clumsy	0.097		

The second option is to use a deep Convolutional Neural Network (CNN) to induce a font embedding space. For this, we rely on a model[9] that creates images by rendering a set of selected letters (L,a,s,e,g,d,h,u,m,H,l,o,i,v) in a grid, and then feeds them through a pretrained deep convolutional network. Finally, PCA is used for dimensionality reduction to obtain vectors that are compared in terms of cosine similarity. The dataset contains embeddings for 1,883 fonts.

We consider two further alternative similarity measures that effectively restrict the candidate spaces of the above two measures to fonts having the same category as the input font. For example, for an input *handwriting* font, these measures regard all non-*handwriting* fonts as having a similarity of zero.

C. Evaluation

To evaluate this, for each f in $\mathcal{F}$[10], we predict $\vec{f}$ using $\mathcal{F} \setminus f$. We replicate the tests four times, for combinations of similarity measures and methods (weighted, unweighted). Comparing predicted $\vec{f}$ with the ground-truth $\vec{f'}$, an $|\mathcal{A}|$-dimensional vector $\vec{e}$ is calculated as:

$$\vec{e} = \vec{f'} - \vec{f}. \tag{4}$$

For each attribute, we then generate an error value e by averaging the absolute values of errors in $\vec{e}$. The test results

[9] https://github.com/Jack000/fontjoy

[10] We use the 161 fonts that are common to all datasets.

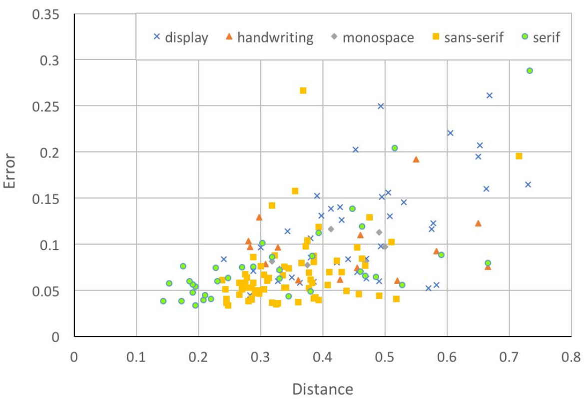

Fig. 7: Scatter plot of the results relating distances of the samples and corresponding error values.

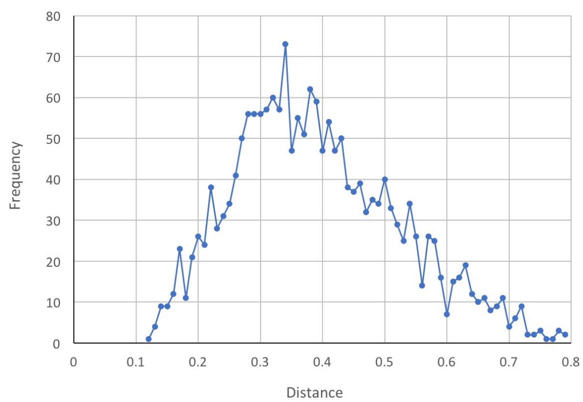

Fig. 8: Distance distribution of the generated dataset.

are summarized in Figure 6. The error scores reported here are averages over e values across all $a \in \mathcal{A}$. The CNN embedding similarity metric results in a lower e for both the weighted and unweighted methods. *Category* based similarities led to slightly improved results for visual features, whereas they did not show any improvements for the embeddings. The lowest error is obtained when $k = 4$ for the weighted version.[11]

Table III lists the e value for each $a \in \mathcal{A}$ using the weighted embedding method where $k = 4$. The most successful predictions are made for *fresh*, whereas the least successful ones are for *technical*. The error scores lie in the narrow range between 0.05 and 0.11, whereas the full value range is between [0,1]. Analyzing these attribute-based error values together with the interactive visualization introduced in the previous section reveals that attributes with lower ranges have a lower degree of error, whereas attributes with high ranges tend to have greater levels. Another factor that appears to have an impact is the distribution of attribute values among different

[11] We attempted to compare the error distribution of the typographical features against the CNN approach to explore to what extent these two metrics might provide complementary signals. However, they both seem to share a similar error pattern. Hence, it was not possible to obtain a significant improvement through a hybrid use of these metrics.

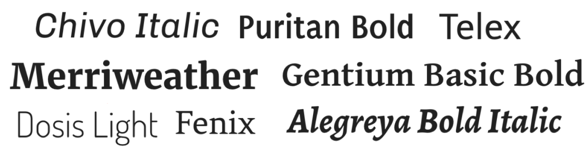

Fig. 9: Font samples with the lowest error averages.

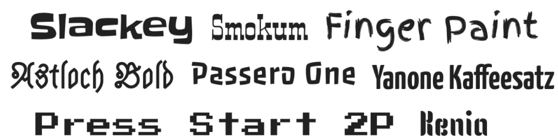

Fig. 10: Font samples with the highest error averages.

font categories. High-ranged values for which different subranges are dense in certain categories seem to be associated with a lower error than ones with mixed such distributions.

Figure 7 shows how e changes with respect to $d(f', f)$. The likelihood of an error increases with increasing distance. However, there are also many cases in which the error is low despite high distances. It is also clear that certain font categories are easier to predict (such as *serif* and *sans-serif*) than others (such as *display*). Figures 9 and 10 show font samples with the highest and lowest e, respectively.

D. Attribute Prediction

Finally, we predict $\vec{f}$ for all fonts covered by the CNN embeddings, using the weighted method with $k = 4$. Figure 8 plots average distance distributions, which have the potential to serve as an indicator for the success of the method, since, based on the previous analysis, the error is found to be low for low distances. We publicly share the resulting dataset.

V. Semantics at a Larger Scale

Next, we analyze the potential of the resulting dataset. This analysis centers around the semantic signatures provided in Figures 11, 12, and 13, and is made available online.

A. Expressive Potential for Attributes

Figure 11 reveals the potential of the included fonts to represent different semantic attributes adequately. For a given attribute, the existence of high-scoring fonts entails a potential to convey that particular attribute effectively. In contrast, a narrower range of values limits this capability. Based on these considerations, we consider three categories of attributes.

1) High Potential: Attributes in this category are associated with fonts with a wide range of association scores, encompassing both very high (>0.8) and very low (<0.2) values. This is a high potential scenario because a well-chosen font can easily distinguish itself from the remaining fonts and may reflect the attribute more strongly. Based on the analysis in Figure 11, the attributes in this category are *angular*, *artistic*, *attention-grabbing*, *attractive*, *boring*, *complex*, *dramatic*, *happy*, *modern*, *playful*, *sloppy*, *strong*, and *thin*.

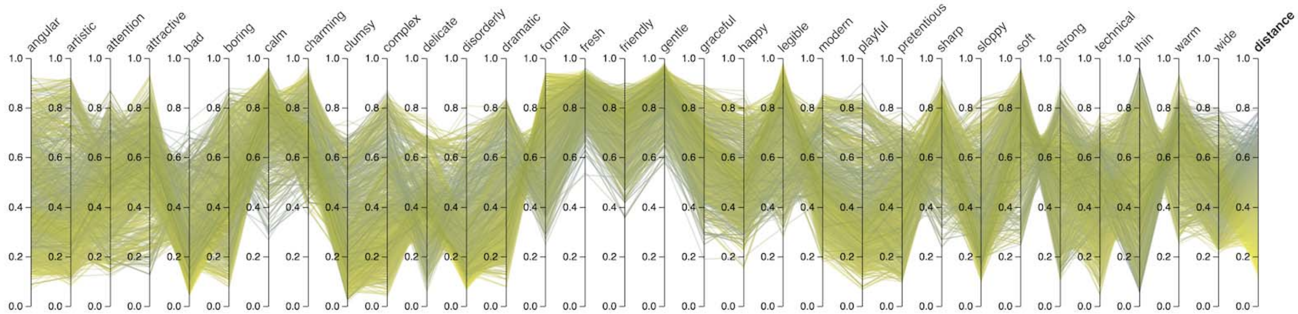

Fig. 11: The semantic signatures of the fonts as generated by our method. This shows the distribution of the font values for the semantic attributes used. Each line is colored based on the distance value of the corresponding font.

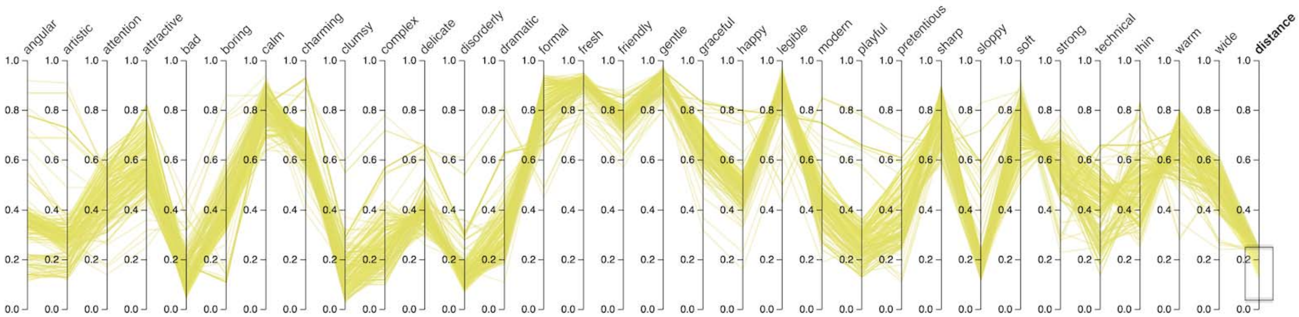

Fig. 12: Semantic signatures of the lowest distance (highest confidence) fonts, as filtered from the visualization in Figure 11

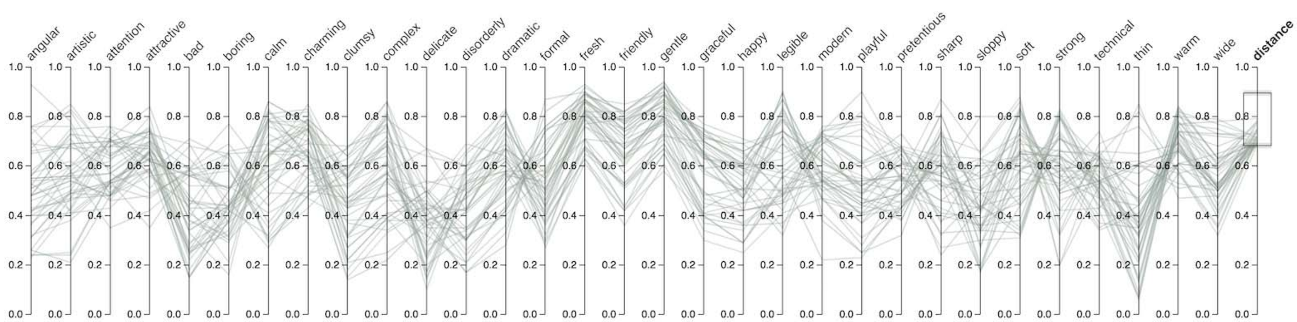

Fig. 13: Semantic signatures of the highest distance (lowest confidence) fonts, as filtered from the visualization in Figure 11

2) Moderate Potential: These attributes possess a high average value, which, at first glance, might be taken as implying a high potential. Yet, this also suggests a potential challenge in emphasizing the attribute more markedly. Still, creating a strong representation may be possible if fonts for other attributes (perhaps opposite attributes) exist in the same context. For this reason, we consider the following attributes as moderate potential ones: *calm*, *charming*, *formal*, *fresh*, *friendly*, *gentle*, *graceful*, *legible*, *sharp*, *soft*, and *warm*.

3) Low Potential: We consider the attributes in this category as having low potential due to an absence of fonts with very high values (>0.8) for them. Specifically, the attributes in this category are *bad*, *clumsy*, *delicate*, *disorderly*, *pretentious*, and *technical*. Despite being categorized as showing limited promise, these attributes might still prove informative as to which attributes to explore as potential candidates for the *moderate potential* category (e.g., opposites of these attributes).

B. Quality

We use the nearest neighbor distances to further assess the quality of the dataset. As discussed in the previous section, our algorithm uses the most similar four fonts to determine the values for a new font. The success of the algorithm increases when the average distance to these similar fonts decrease. For this reason, the distance value may be interpreted as a confidence value (inverse relationship), although the evaluations reveal that in some cases it is still possible to have a successful prediction with a high distance value.

Figure 12 depicts a filtered plot considering only fonts with a low distance value (lower than ~0.27) and thus in the highest confidence bracket. The interesting finding here is that many of the fonts at the high or low end of the range that determine the category of a given attribute remain. In other words, these are the fonts that possess an important relationship with the attribute, and, in the case of being at the high end, are the strongest candidates to be selected.

Figure 13 depicts a filtered plot considering only fonts with a high distance value (higher than ~0.7) and thus in the lowest confidence bracket. The results, again, appear favorable, as these fonts have scores that lie mostly closer to the middle of the respective ranges for each attribute. Thus, they have a smaller chance of being selected to represent those attributes. Nevertheless, this does not preclude the possibility of them in reality having values closer to the ones at the ends of these ranges, which would mean that we might be overlooking a font that could be a good candidate to represent an attribute.

Taken together, this suggests that the categorical organization of the attributes provided above is overall fairly reliable. In conjunction with the finding from Section V that most fonts have low-to-mid range distances, the fonts in our dataset, especially when picked from the ends of the ranges, tend to have very representative attribute values (see Figure 1).

VI. Discussion

We now review and discuss the findings of this paper, starting with the analysis in Section III. Although there are some general trends in the data (such as high values for *gentle*), fonts appear to show characteristic biases. This is expected, as font categories are defined based on combinations of certain design metrics (contrast, x-height, etc.), which give rise to a particular perception with shared semantic characteristics. This is also confirmed by the scattered distribution of the font category *display*, since it is the only one among these categories with a very wide range of characteristics, complying with general design knowledge. Our results are also in line with the previous user study by Shaik et al. [3]. Both studies find *serif* to be *formal*, *monospace* to be *boring*, and *handwriting* to be *happy*.

Despite being able to reflect these category-based biases, the crowdsourced dataset is not large enough and the correlations not sufficiently clear to give rise to generalized metrics or models. To overcome this challenge, in Section IV, we use a k-NN approach. Our evaluation shows that this method has very low error rates for the font attribute prediction problem at hand. Another interesting point here is that the CNN embeddings are found to be a better similarity measure for attribute prediction compared to the typographical features.

Section V attempts to approximate the quality of the generated dataset, and makes predictions about the potential of the fonts to represent these semantic attributes. A point that should be noted is that all attributes have values growing away from the center (0.5). This is important because it shows that there is a high risk to unintentionally represent these attributes at different levels (high or low) if the font selection process does not consider these associations.

VII. Conclusion and Future Work

Starting with a crowdsourced dataset, we first analyzed the relationship between font categories/styles and semantic attributes, and reported a series of novel findings. We published an interactive online visualization that provides further insights from the dataset. Secondly, we induced a large-scale repository of semantic signatures for nearly 2,000 fonts, based on a weighted k-NN approach via a CNN embedding based similarity measure. Finally, we analyzed the resulting data to assess its quality, and provided an interactive visualization to allow for exploring it. We also characterized the potential of these fonts to represent different groups of semantic attributes. In future work, we intend to further extend the dataset in the semantic dimension, i.e., adding novel attributes via the existing associations, as well as by capturing thousands of additional fonts, which our approach already supports.

References

[1] E. R. Brumberger, "The rhetoric of typography: The persona of typeface and text," *Technical Communication*, vol. 50, no. 2, pp. 206–223, 2003.

[2] P. O'Donovan, J. Libeks, A. Agarwala, and A. Hertzmann, "Exploratory font selection using crowdsourced attributes," *ACM Trans. Graph.*, vol. 33, no. 4, pp. 92:1–92:9, 2014.

[3] A. D. Shaikh, B. S. Chaparro, and D. Fox, "Perception of fonts: Perceived personality traits and uses," *Usability News*, vol. 8, no. 1, pp. 1–6, 2006.

[4] S. Juni and J. S. Gross, "Emotional and persuasive perception of fonts," *Perceptual and motor skills*, vol. 106, no. 1, pp. 35–42, 2008.

[5] A. D. Shaikh, "Psychology of onscreen type: Investigations regarding typeface personality, appropriateness, and impact on document perception," Ph.D. dissertation, 2007.

[6] C. Lewis and P. Walker, "Typographic influences on reading," *British Journal of Psychology*, vol. 80, no. 2, pp. 241–257, 1989.

[7] A. Fligner, "The effect of packaging typeface on product perception and evaluation," Ph.D. dissertation, The Ohio State University, 2013.

[8] C. Velasco, A. Salgado-Montejo, F. Marmolejo-Ramos, and C. Spence, "Predictive packaging design: Tasting shapes, typefaces, names, and sounds," *Food Quality and Preference*, vol. 34, pp. 88 – 95, 2014.

[9] Q. Qiu, S. Watanabe, and K. Omura, "Images and anatomy of latin typefaces," *International Journal of Affective Engineering*, vol. 16, no. 2, pp. 121–130, 2017.

[10] C. Velasco, A. T. Woods, S. Hyndman, and C. Spence, "The taste of typeface," *i-Perception*, vol. 6, no. 4, 2015.

[11] Z. Wang, J. Yang, H. Jin, E. Shechtman, A. Agarwala, J. Brandt, and T. S. Huang, "Deepfont: Identify your font from an image," in *Proc. 23rd ACM MM*. ACM, 2015, pp. 451–459.

[12] J. Qiao, "Fontjoy – Font pairing made simple," http://fontjoy.com/, 2017, accessed: 2017-07-29.

[13] R. L. Hazlett, K. Larson, A. D. Shaikh, and B. S. Chaparo, "Two studies on how a typeface congruent with content can enhance onscreen communication." *Information Design Journal*, vol. 20, no. 3, 2013.

[14] A. D. Shaikh, D. Fox, and B. S. Chaparro, "The effect of typeface on the perception of email," *Usability News*, vol. 9, no. 1, pp. 1–7, 2007.

[15] A. D. Shaikh, "The effect of website typeface appropriateness on the perception of a company's ethos," *Usability News*, vol. 9, no. 2, 2007.

[16] T. L. Childers and J. Jass, "All dressed up with something to say: Effects of typeface semantic associations on brand perceptions and consumer memory," *Journal of Consumer Psychology*, vol. 12, no. 2, pp. 93–106, 2002.

[17] T. J. Van Rompay and A. T. Pruyn, "When visual product features speak the same language: Effects of shape-typeface congruence on brand perception and price expectations," *Journal of Product Innovation Management*, vol. 28, no. 4, pp. 599–610, 2011.

[18] T. Rabinowitz, *Exploring Typography*, 2nd ed. Boston, MA, United States: Course Technology Press, 2015.