Logo Design Analysis by Ranking

Takuro Karamatsu*, Daiki Suehiro*†, and Seiichi Uchida*
*Kyushu University †RIKEN

***Abstract*—In this paper, we analyze logo designs by using machine learning, as a promising trial of graphic design analysis. Specifically, we will focus on favicon images, which are tiny logos used as company icons on web browsers, and analyze them to understand their trends in individual industry classes. For example, if we can catch the subtle trends in favicons of financial companies, they will suggest to us how professional designers express the atmosphere of financial companies graphically. For the purpose, we will use top-rank learning, which is one of the recent machine learning methods for ranking and very suitable for revealing the subtle trends in graphic designs.**

***Keywords*-Logo design; Graphic design analysis; Top-rank learning; TopPush algorithm; Machine learning**

I. Introduction

Our daily life is surrounded by many *graphic designs*. Typical graphic designs are poster and advertisement, signboard, logo and icon, book cover, the label of goods, web page, and so on. Graphic designs often give us some special impression that affects our reaction to them, such as decision, behavior, and feeling. For example, we often make our decision on buying a commercial product just by watching its label design. In other words, professional designers always try to give a certain impression to the customers through their graphic design.

The relationship between graphic design and its impression is still an interesting open problem. Of course, there are many graphic design textbooks and they show various examples and use-cases. However, most textbooks just give subjective and intuitive explanations to the examples and thus give neither objective nor quantified facts revealed through scientific analysis with enough amount of data.

Graphic design analysis is not only meaningful as a scientific exploration of the relationship between graphic designs and their impression but also useful for various application. Clearly, the results will help the process of creating various graphic designs. In addition, the relationship between graphic designs and their use-cases will be useful for image understanding. An example is Jolly et al. [1] where book cover designs are utilized to understand book genres. An example in more general image understanding research is Movshovitz et al. [2], where that visual design of storefront is used for recognizing the store class.

In this paper, we analyze logo designs objectively and automatically by using machine learning, as a promising trial of graphic design analysis. Especially, we will focus on *favicon images*, instead of dealing with arbitrary logo images. Each favicon image is a tiny square logo and used as a company icon on web browsers. This property of favicons is useful to limit the variation of logo images since the definition of logo is wide [3] and thus their variation is too large to analyze.

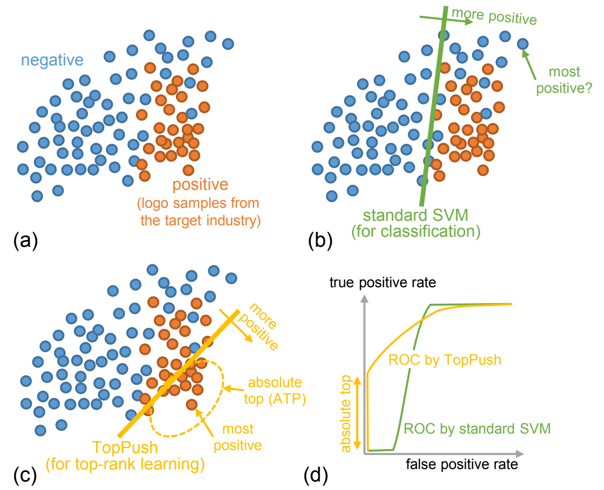


Figure 1. Overview of top-rank learning for logo (favicon) analysis. (a) Logo sample distribution. Samples from the target industry are shown by orange dots. (b) Standard SVM for classification. Its result is not suitable to rank the samples. (c) Top-rank learning by the TopPush algorithm. It can give "absolute top" samples that contain the most positive sample and other "very positive" samples. (d) Relationship to ROC. Absolute top samples by top-rank learning form the vertical-rise of ROC.

Our main purpose of analyzing favicon designs is to catch their trends in individual industry classes, such as finance and transportation. For example, we want to analyze what kind of favicons are distinctive to financial companies and how their trend is different from the other industries. If we can find any trend specific to finance, it will reveal how professional designers express the impression of financial companies through their favicon design.

Since we anticipate that difference of the favicon design trends among industries is very subtle, we will employ a machine-learning strategy to *pick up the most distinctive favicon samples for each industry*. More specifically, we will use TopPush algorithm [4], which is a top-rank learning algorithm that can give a ranking (i.e., an order) of samples according to their distinctiveness.

Fig. 1 illustrates our favicon analysis formulated as a top-rank learning problem. Fig. 1 (a) is an example of the distribution of favicon image samples; they are comprised of positive samples from the target industry and negative samples from the others. One naive strategy to find the most distinctive samples (i.e., the most positive samples) is to use the classification boundary between positive and negative sample sets by, for example, a standard SVM. The furthest positive samples from the boundary are expected to be the most distinctive samples for the target class. However, as shown in Fig. 1 (b), this strategy fails to find the most positive samples; this is simply because the standard SVM is designed not for finding them but for higher classification accuracy of all samples.

In contrast, as shown in Fig. 1 (c), top-rank learning by TopPush provides a function to rank the samples so that the most positive samples achieve the top ranks and the negative samples result in lower ranks. As an advantage of TopPush over other top-rank learning algorithms, it is theoretically guaranteed that TopPush maximizes the number of *absolute top positives*, which are positive samples distinguished from all negative samples in their ranking. Consequently, TopPush will be powerful for our favicon analysis task, by finding the most distinctive positive samples as absolute top positives, even when the difference between positive and negative samples is very subtle.

The main contributions of this paper are summarized as follows.

- To the authors' best knowledge, this is the first trial of the objective and automatic analysis of the trends of graphic design in individual industry classes.
- We formulate the problem of selecting the most distinctive favicon images to each industry class as a top-rank learning problem and applied the TopPush algorithm to the favicon images as a solution to the problem. Fortunately, we could catch their trends in color and shape successfully, even though there are huge variations in the favicon designs and thus their trends are too subtle and unclear.

II. Related Work

A. Logo analysis

In [3], *logo* is defined as "a symbol, a graphic and visual sign which plays an important role into the communication structure of a company" and has three types: "Iconic or symbolic logo", "text-based logo", and "mixed logo." The same paper also defines seven conditions (such as legibility) and seven functions. Several functions are related to the impression of the logo; for example, its signification function will appeal some emotion to the observer and its aesthetics function will inspire pleasure.

The recent progress of neural network technologies enables us to generate graphic designs. For example, automatic font generation is a hot topic [5]–[8]. However, to the authors' best knowledge, automatic logo generation has not been tried except for Sage et al. [9]. They use generative adversarial networks (GAN) and a public logo dataset called Large Logo Dataset (LLD), for generating logo images and favicon images. LLD contains more than 600 thousand logos and thus larger than WebLogo-2M [10] which contains 194 different logos shown in 1,867,177 images.

Research for correlating graphic designs and their impressions has been analyzed for 100 years and always relied on subjective and small-scale experiments. For example, an old study [11] has made a subjective experiment to understand the impression of typefaces (i.e., fonts). Even recent trials still rely on subjective impressions. For example, Fontmatcher [12] for correlating typefaces and images uses a font dataset with subjective impression [13]. In this paper, we, therefore, try to reveal the correlation objectively and automatically.

B. Learning-to-rank

Learning-to-rank is an important machine learning task, especially in the field of information retrieval [14], [15]. The bipartite ranking problem is a basic problem in learning-to-rank, which assumes binary-labeled (i.e., positive and negative) training data [15], [16]. Many algorithms to solve the bipartite ranking problem have been proposed, and the applied to the ranking tasks for various domains such as recommendation, search engines, and financial risk analysis [15]–[18], in addition to classification tasks for maximizing AUC such as [19], [20].

Top-rank learning we apply in this paper has been theoretically and practically developed in the last decade or so [4], [21], [22]. TopPush [4] is one of the novel learning algorithms for top-rank learning, which maximizes the number of absolute top positives rather than standard classification algorithm, as shown in Fig. 1 (d).

Our key idea is that we can discover the distinctive samples for an industry class from a large number of favicon image data by only observing samples that are highly ranked by TopPush. Although the design trends of the favicons in the class is very subtle, we can still catch the trends by observing highly-ranked images because these images (including absolute top positives) are very distinctive to the class. To the best of our knowledge, top-rank learning, as well as learning-to-rank, has not been applied to graphic design analysis so far.

III. Top-rank learning

We employ a learning-to-rank approach to discover distinctive samples and some trends for each industry class. In this section, we first introduce the bipartite ranking problem and then its extension, i.e., the top-rank learning problem. TopPush, an algorithm to solve the top-rank learning problem, is also explained briefly. Note that we assume linear classifiers throughout this paper for making the analysis results more understandable.

A. Bipartite ranking problem

The bipartite ranking problem is one of the basic problems in the learning-to-rank task. We briefly described the standard setting of the bipartite ranking problem as below: We assume a training sample set that consists of p positive data

$\mathbf{x}_1^+, \ldots, \mathbf{x}_p^+$ and n negative data $\mathbf{x}_1^-, \ldots, \mathbf{x}_n^-$. The goal of the bipartite ranking problem is to find a ranking function f that gives positive data higher values than negative data. That is, we want to find f that maximizes the number of pairs $(\mathbf{x}_i^+, \mathbf{x}_j^-)$ which satisfies $f(\mathbf{x}_i^+) > f(\mathbf{x}_j^-)$. Consequently, the optimization of f is the maximization of AUC because AUC is given as follows:

$$\mathrm{AUC} = \sum_{i=1}^{p} \sum_{j=1}^{n} \frac{I(f(\mathbf{x}_i^+) > f(\mathbf{x}_j^-))}{pn},$$

where I denotes the indicator function. Since AUC is a well-known performance measure for tasks with imbalanced data, the ranking function f is robust to those tasks.

Among many algorithms for solving the problem, RankSVM (or Ranking SVM) [15] is well-known and has been applied to many imbalanced-classification tasks. The key idea of RankSVM is to reformulate the AUC maximization to the minimization problem of the following loss:

$$L_{\mathrm{RankSVM}} = \sum_{i=1}^{p} \sum_{j=1}^{n} \ell(f(\mathbf{x}_j^-) - f(\mathbf{x}_i^+)), \tag{1}$$

where ℓ is the hinge-loss for given samples.

It should be emphasized that the above AUC maximization evaluates the overall ranking accuracy from the bottom to top. In our graphic design analysis, we need to focus only on top-ranked samples; this is because only top-ranked samples will show the subtle trends and low-ranked samples will not. Therefore, we do not need to push such low-ranked samples to upper ranks.

B. Top-rank learning

The top-rank learning problem is an extension of the bipartite ranking problem. Given binary labeled training samples, its goal is to find a top-ranking function that maximizes the ranking accuracy *at the top* [21]. Specifically, we want to maximize the rate of absolute top positives (hereafter called ATP, shown in Fig. 1(c)) in whole positives [23]. ATP is formulated as follows:

$$\mathrm{ATP} = \frac{1}{p} \sum_{i=1}^{p} I\left(f(\mathbf{x}_i^+) - \max_{j=1,\ldots,n} f(\mathbf{x}_j^-) > 0 \right). \tag{2}$$

ATP evaluates the number of positive data which is ranked higher than the highest negative data.

TopPush [4] is an efficient algorithm for maximizing ATP for a linear ranking function (i.e., $f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x}$). TopPush attempts to minimize the following loss which is a relaxed version of ATP:

$$L_{\mathrm{TopPush}} = \frac{1}{p} \sum_{i=1}^{p} \ell\left(\max_{j=1,\ldots,n} \mathbf{w} \cdot \mathbf{x}_j^- - \mathbf{w} \cdot \mathbf{x}_i^+ \right), \tag{3}$$

where ℓ is a convex and non-decreasing loss function and we used the truncated quadratic loss $\ell(z) = [1+z]_+^2$. Finally, the optimization problem of TopPush is formulated as the minimization problem of (3) with a regularization term for $\mathbf{w}$, namely:

$$\min_{\mathbf{w}} \frac{\lambda}{2} \|\mathbf{w}\|^2 + \frac{1}{p} \sum_{i=1}^{p} \ell\left(\max_{j=1,\ldots,n} \mathbf{w} \cdot \mathbf{x}_j^- - \mathbf{w} \cdot \mathbf{x}_i^+ \right),$$

where λ is a regularization parameter. If λ becomes smaller, positive samples are more pushed to be top-ranked; if λ is set at a very small positive value, it may give a result strongly overfitted to the training samples.

IV. Experiments

In order to understand the correlation between favicon images and individual industry classes, an experiment was conducted with 3,100 favicon images and TopPush.

A. Data collection

For the experiment, we collected 3,100 favicon images using web scraping technique since Large Log Dataset (LLD) [9] does not have industry class information. Specifically, we searched for the official website of each company on the list by NASDAQ[1]. After finding the official website, we extract its favicon image analyzing HTML of the website. Some group companies share the same favicons and we do not unify them. Each favicon image is resized to 32×32 pixels[2].

As industry classes, we follow the definition in the NASDAQ list. In the list, all companies are classified into 12 industry classes, such as "Basic Industry" and "Transportation," as shown in Fig. 2. In other words, each of 3,100 favicons is classified into one of 12 classes.

B. The detailed procedure for ranking favicons by TopPush

In order to rank the favicon images of a certain industry class, we use the implementation[3] of TopPush. TopPush is applied to all 3,100 favicons by considering a certain target industry class as the positive class and the other 11 classes as the (big) negative class. Then we can expect that favicon images with the most distinctive designs to the target class will be ranked at the top. We perform this one-vs-others ranking task for each of 12 industry classes. Although this one-vs-others task is an imbalance problem, top-rank learning is theoretically robust for imbalanced tasks and we do not need to give any weight to the positive samples.

In order to understand our analysis result easily, we use the bitmap feature to represent each favicon image. Since each favicon is a 32×32 RGB image, its feature becomes a 3,072-dimensional vector. We also conducted other experiments using different feature vectors, such as an RGB histogram and a Histogram-of-gradient (HOG) feature. Those results are given at `https://github.com/karamatsutakuro/logo_TopPush_supplementary`.

[1] https://www.nasdaq.com/

[2] The dataset will be published upon the paper acceptance.

[3] http://lamda.nju.edu.cn/files/TopPush.zip

Industry	Industry example	Favicon example	Industry	Industry example	Favicon example
Basic Industries (215)	· Agricultural Chemicals · Major Chemicals · Precious Metals · Steel/Iron Ore		Finance (521)	· Business Services · Investment Bankers/Brokers/Service · Major Banks · Savings Institutions	
Capital Goods (231)	· Auto Manufacturing · Electrical Products · Industrial Machinery/Components · Metal Fabrications		Health Care (532)	· Hospital/Nursing Management · Industrial Specialties · Major Pharmaceuticals · Medical/Dental Instruments	
Consumer Durables (92)	· Automotive Aftermarket · Home Furnishings · Specialty Chemicals · Telecommunications Equipment		Miscellaneous (95)	· Business Services · Industrial Machinery/Components · Multi-Sector Companies · Publishing	
Consumer Non-Durables (136)	· Apparel · Beverages (Production/Distribution) · Electronic Components · Packaged Foods		Public Utilities (131)	· Electric Utilities: Central · Natural Gas Distribution · Power Generation · Telecommunications Equipment	
Consumer Services (467)	· Broadcasting · Hotels/Resorts · Real Estate Investment Trusts · Restaurants		Technology (427)	· Computer Software: Prepackaged Software · Diversified Commercial Services · EDP Services · Semiconductors	
Energy (179)	· Industrial Machinery/Components · Integrated oil Companies · Natural Gas Distribution · Oil & Gas Production		Transportation (74)	· Air Freight/Delivery Services · Marine Transportation · Oil Refining/Marketing · Trucking Freight/Courier Services	

Figure 2. Twelve industry classes defined by NASDAQ. The parenthesized number represents the number of favicon images of each industry class.

C. Qualitative evaluation

Fig. 3 shows six ranking results. In each result, all 3,100 favicon images are ranked in the order from top-left to bottom-right. The ranking orders are different in those results because of the difference in the target industry class and the parameter value of λ. Among 12 classes, we picked up three classes, "Energy", "Technology" and "Transportation" in Fig. 3 because the top-ranked samples have smaller variances than the other classes. In other words, each of these three classes has more similar top-ranked samples. The parameter λ is set at 0.1 or 10. In the figure, top-10 images of the target class are also shown along with their average image. Fig. 4 shows the top-10 ranked favicons and their average for all 12 classes (at $\lambda = 10$).

In Fig. 3, a pair of images $\mathbf{w}^+$ and $\mathbf{w}^-$ are also shown to visualize the learned $\mathbf{w} = \mathbf{w}^+ - \mathbf{w}^-$. In $\mathbf{w}^+$, brighter pixels correspond to larger positive-valued elements of $\mathbf{w}$. Black pixels correspond to negative-valued elements. Similarly, in $\mathbf{w}^-$, brighter pixels correspond to larger negative-valued elements. If a favicon image $\mathbf{x}$ has an RGB value similar to $\mathbf{w}^+$ (i.e., if $\mathbf{x} \cdot \mathbf{w}^+$ is large), the favicon will be ranked higher. If $\mathbf{x} \cdot \mathbf{w}^-$ is large, it will be ranked lower.

From those figures, the following facts can be observed:

- The favicons of the target class are ranked around the top as expected since the black boxes indicating the rank distribution of the target favicons gather around the top. Especially when $\lambda = 0.1$, most favicons belong to ATP. (This implies the rank result is overfitted to the samples with $\lambda = 0.1$.)
- Since $\mathbf{w}^+$ and $\mathbf{w}^-$ become a random dot-like pattern and does not show any structure, there is no common structure or part in the favicon images in any class. This suggests that favicon images have huge variations even if we consider only a single industry class.
- In spite of the huge variations, TopPush still could catch subtle color trends in their design, as shown by the average of top-10 images. Specifically, the top-ranked "Energy" favicons have more yellow and blue pixels whereas "Technology" favicons have more orange and green and "Transportation" have more blue and red.
- A similar observation of the average images suggests that TopPush also catches a subtle shape trend for each class. As shown in Fig. 4, several classes have an average image capturing a roundish structure. For example, "Transportation", "Consumer Services" and "Health Care" have such a roundish structure. In contrast, the average images of "Finance", "Public Utilities", and "Technology" are much less roundish.

About the last two facts, it should be emphasized again that they are not just an average image. They are an average image of top-ranked favicons, which are the distinctive favicons to the class. (There are many favicons which show different designs from the average images but they are ranked lower.) TopPush could catch the subtle trends specific to the class and make the favicons showing the trends distinctively to be ranked higher.

D. Quantitative evaluation by comparing TopPush to the standard SVM

Finally, as a quantitative evaluation of TopPush for our favicon image ranking task, we compare TopPush to the standard SVM. Although TopPush is robust to class imbalanced tasks, the standard SVM is not. We, therefore, use a simple remedy by Akibani et al. [24] to deal with the class

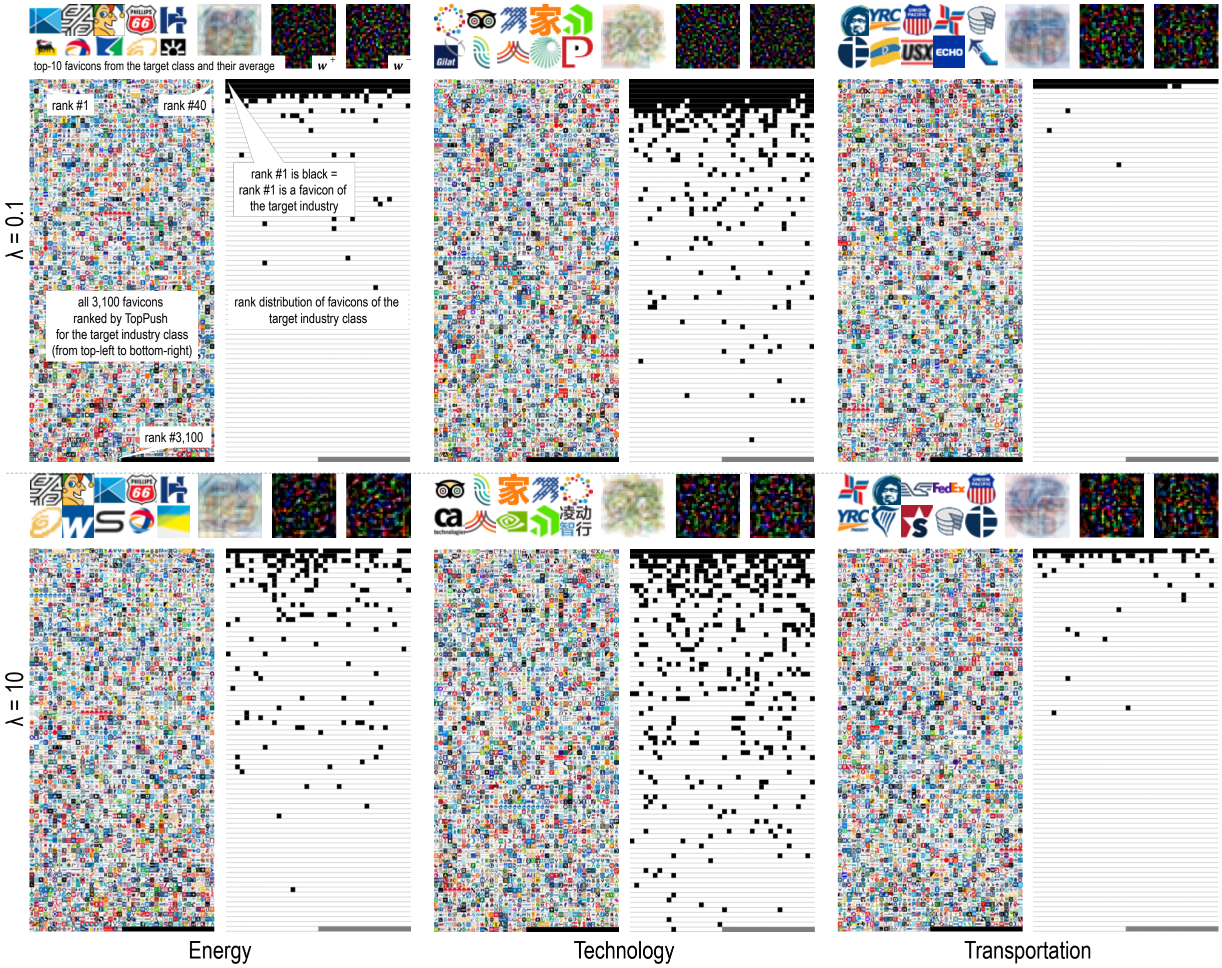


Figure 3. Ranking results of three industry classes, "Energy", "Technology" and "Transportation." Those results are given by TopPush under two parameter values $\lambda = 0.1$ and 10. For each result, top-10 favicon images from the target industry class, their average image, a bitmap representation of $\mathbf{w}^+$ and $\mathbf{w}^-$, the rank of all 3,100 favicons from top-left to bottom-right, and the rank distribution of the favicons of the target industry class (indicated as black boxes).

imbalance. With the remedy, the standard SVM formulated as the following optimization problem:

$$\min_{\mathbf{w},b,\boldsymbol{\xi}^+,\boldsymbol{\xi}^-} \quad \frac{1}{2}\|\mathbf{w}\|^2 + \frac{nC}{p}\sum_{i=1}^{p}\xi_i^+ + C\sum_{j=1}^{n}\xi_j^-,$$

$$\text{sub.to } (\mathbf{w}\cdot\mathbf{x}_i^+ + b \geq 1-\xi_i^+) \wedge (\xi_i^+ \geq 0) \text{ for } i=1,\ldots,p,$$
$$(-(\mathbf{w}\cdot\mathbf{x}_j^- + b) \geq 1-\xi_j^-) \wedge (\xi_j^- \geq 0) \text{ for } j=1,\ldots,n,$$

where p and n are the number of positive and negative samples and $p \ll n$. Finally, we rank the samples $\{\mathbf{x}\}$ by the inner-product value between $\mathbf{x}$ and the classification hypeArplane $\mathbf{w}$ for different values of C.

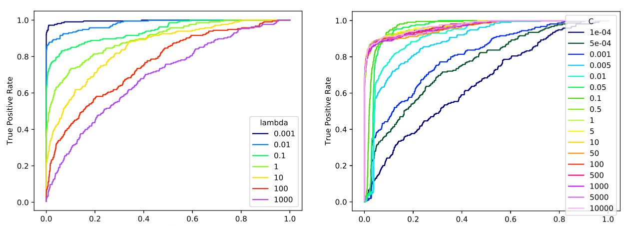
Figs. 5 (a) and (b) show the ROC curves by TopPush and the standard SVM, respectively, for the class "Basic Industries." (We omit the plots for other classes because they also show similar ROC curves as those figures.) From the leftmost side of the ROC curves in (a), we can understand that TopPush could find absolute top. We can see that the ATP of TopPush is significantly larger than the standard SVM. Note that several ROC curves of the standard SVM show a short horizontal shift in their leftmost side. This means that the very top-ranked samples come from non-target classes like Fig. 1 (b). In contrast, the ROC curves of TopPush do not have such a part.

V. CONCLUSION

In this paper, we analyzed 3,100 favicon images to understand the trends in their design at individual industry classes objectively and automatically. For the analysis, we used TopPush, which is an algorithm for a top-rank learning problem. The analysis result shows that TopPush could catch very suitable trends in favicon designs in spite of huge graphical variations in them. We also confirmed quantitatively that TopPush shows better performance than the standard SVM in the ranking task.

Future work will focus on further trend analysis in some different feature space, which will show the trends more distinctively. Instead of the bitmap feature, color histogram and

Figure 4. Top-10 favicon images and their average image (at $\lambda = 10$). The last value is the variance of pixel values of those 10 images.



(a) TopPush (b) standard SVM

Figure 5. ROC curves for favicons of "Basic Industries."

bag-of-visual-words representation will be the first choices as the feature space to be examined. Analysis of other graphic designs is also a promising research direction.

Acknowledgement

We appreciate all anonymous reviewer's helpful comments. This work was supported by JSPS KAKENHI Grant Number JP17H06100 and JP18K18001.

References

[1] S. Jolly, B. K. Iwana, R. Kuroki, and S. Uchida, "How do convolutional neural networks learn design?" in *Proc. ICPR*, 2018.

[2] Y. Movshovitz-Attias, Q. Yu, M. C. Stumpe, V. Shet, S. Arnoud, and L. Yatziv, "Ontological supervision for fine grained classification of street view storefronts," in *Proc. CVPR*, 2015.

[3] G. Adîr, V. Adîr, and N. E. Pascu, "Logo design and the corporate identity," *Procedia-Social and Behavioral Sciences*, vol. 51, pp. 650–654, 2012.

[4] N. Li, R. Jin, and Z.-H. Zhou, "Top rank optimization in linear time," in *Proc. NIPS*, 2014.

[5] P. Lyu, X. Bai, C. Yao, Z. Zhu, T. Huang, and W. Liu, "Auto-encoder guided GAN for chinese calligraphy synthesis," in *Proc. ICDAR*, 2017.

[6] A. K. Bhunia, A. K. Bhunia, P. Banerjee, A. Konwer, A. Bhowmick, P. P. Roy, and U. Pal, "Word level font-to-font image translation using convolutional recurrent generative adversarial networks," in *Proc. ICPR*, 2018.

[7] Y. Guo, Z. Lian, Y. Tang, and J. Xiao, "Creating new chinese fonts based on manifold learning and adversarial networks," in *Eurographics*, 2018.

[8] S. Azadi, M. Fisher, V. G. Kim, Z. Wang, E. Shechtman, and T. Darrell, "Multi-content GAN for few-shot font style transfer," in *Proc. CVPR*, 2018.

[9] A. Sage, E. Agustsson, R. Timofte, and L. Van Gool, "Logo synthesis and manipulation with clustered generative adversarial networks," in *Proc. CVPR*, 2018.

[10] H. Su, S. Gong, and X. Zhu, "WebLogo-2M: Scalable logo detection by deep learning from the web," in *Proc. ICCVW*, 2017.

[11] A. T. Poffenberger and R. Franken, "A study of the appropriateness of type faces." *J. Applied Psychology*, vol. 7, no. 4, pp. 312–329, 1923.

[12] S. Choi, K. Aizawa, and N. Sebe, "Fontmatcher: font image paring for harmonious digital graphic design," in *Proc. IUI*, 2018.

[13] P. O'Donovan, J. Lībeks, A. Agarwala, and A. Hertzmann, "Exploratory font selection using crowdsourced attributes," *ACM Trans. Graphics*, vol. 33, no. 4, 2014.

[14] W. W. Cohen, R. E. Schapire, and Y. Singer, "Learning to order things," *J. Artificial Intelligence Research*, vol. 10, pp. 243–279, 1999.

[15] T. Joachims, "Optimizing search engines using clickthrough data," in *Proc. SIGKDD*, 2002.

[16] Y. Freund, R. Iyer, R. E. Shapire, and Y. Singer, "An efficient boosting algorithm for combining preferences," *JMLR*, vol. 4, pp. 933–969, 2003.

[17] X. Zhao, G. Li, M. Wang, J. Yuan, Z.-J. Zha, Z. Li, and T.-S. Chua, "Integrating rich information for video recommendation with multi-task rank aggregation," in *Proc. ACM MM*, 2011.

[18] X. Xu, C. Zhou, and Z. Wang, "Credit scoring algorithm based on link analysis ranking with support vector machine," *Expert Systems with Applications*, vol. 36, no. 2, pp. 2625–2632, 2009.

[19] U. Brefeld and T. Scheffer, "AUC maximizing support vector learning," in *Proceedings of ICML Workshop*, 2005.

[20] R. Cruz, K. Fernandes, J. S. Cardoso, and J. F. P. Costa, "Tackling class imbalance with ranking," in *Proc. IJCNN*, 2016.

[21] C. Rudin, "Ranking with a p-norm push," in *Proc. COLT*, 2006.

[22] H. Narasimhan and S. Agarwal, "SVMpAUCtight: A new support vector method for optimizing partial AUC based on a tight convex upper bound," in *Proc. SIGKDD*, 2013.

[23] S. Agarwal, "The infinite push: A new support vector ranking algorithm that directly optimizes accuracy at the absolute top of the list," in *Proc. SDM*, 2011.

[24] R. Akbani, S. Kwek, and N. Japkowicz, "Applying support vector machines to imbalanced datasets," in *Proc. ECML*, 2004.