Supplementary Material for Large-scale Tag-based Font Retrieval with Generative Feature Learning

Tianlang Chen[1], Zhaowen Wang[2], Ning Xu[2], Hailin Jin[2] and Jiebo Luo[1]

[1]University of Rochester [2]Adobe Research

{tchen45, jluo}@cs.rochester.edu, {zhawang, nxu, hljin}@adobe.com

In this document, we provide additional materials to supplement our paper "Large-scale Tag-based Font Retrieval with Generative Feature Learning". In the first section, we provide more details about our collected font retrieval dataset. In the second section, we describe the evaluation measurement of Myfonts-test set. In the third section, we provide the weight initialization details of the attention module and the retrieval model. In the fourth section, we comprehensively illustrate the retrieval performance of the proposed model on typical single-tag and multi-tag queries.

1. Dataset Supplementary Information

In Section 3 of our paper, we present a large-scale tag-based font retrieval dataset, which is collected from My-Fonts. After tag preprocessing, the dataset finally contains 1824 tags for font description. Table 1 provides us deeper insight into these tags by showing the top-200 frequent tags. It can be seen that the dataset contains meaningful tags that cover different aspects of a font, such as its category (e.g "sans-serif", "script", "handwrite"), appearance (e.g. "heavy","outline","round"), utility (e.g. "poster", "magazine", "logo") and other special features (e.g. "kid", "romantic", "cartoon").

On the other hand, we collect a tagging set with ranking information via Amazon Mechanical Turk (AMT) as a complement for evaluation. The detailed process of collecting the tagging set is described in the main submission. Roughly speaking, this set contains 1661 groups. Each group includes three fonts related to a specific tag, and is finally labeled a most matching one agreed by all the workers. We show a large number of group examples in Figure 1 to present the tagging set in detail.

2. Measurement of MyFonts-test Set

We evaluate different models' performance on the MyFonts-test set by two standard measures, mean average precision (mAP) and Normalized Discounted cumulative gain (nDCG). For average precision (AP), given a query q, assuming that the total H positive fonts $\{f_1, f_2,..., f_H\}$ in the test set have affinity score ranks $\{r_1, r_2,..., r_H\}$, the average precision score of q (AP_q) is computed as: $AP_q = \frac{1}{H} \sum_{h=1}^{H} \frac{h}{r_h}$. For nDCG, given the font relevance $\{rel_1, rel_2,..., rel_S\}$ for the total S test fonts, which have affinity score ranks $\{1, 2, ..., S\}$ on query q, $nDCG_q$ is computed as: $DCG_q = \sum_{p=1}^{S} \frac{2^{rel_p}-1}{\log_2(p+1)}$, $nDCG_q = \frac{DCG_q}{IDCG_q}$, where $IDCG_q$ is the maximum possible value of DCG_q for different ranking results on q. The font relevance for a positive font is set to 1, for a negative font, it is set to 0.

In our experiments, given a set of test queries, we compute the mean value of AP (mAP) and nDCG for all queries as the final mAP and nDCG scores.

3. Weight Initialization

In the training process, we find that the weight initialization of the attention module and the retrieval model can make effect on the final performance. For the attention module that contains a fully-connected layer followed by a sigmoid unit, the weights of the fully-connected layer are initialized using a normal distribution (μ, σ) where $\mu = 0$ and $\sigma = 5$. The retrieval model contains two fully-connected layers whose dimensions are set as N and 1. N is the total tag vocabulary size. The first layer with a ReLU unit maps the N-dimensional query-based tag probability vector to a N-dimensional feature vector. The second layer with a sigmoid unit then maps it to the final affinity score. We use a $N \times N$ identity matrix to initialize the weights of the first layer, and use a normal distribution with $\mu = 1$ and $\sigma = 0.02$ to initialize the weights of the second layer.

4. Additional qualitative results

In this section, we illustrate a great number of font retrieval results of the proposed model as the supplement of Figure 6 in the main submission. We test the model's performance on typical single-tag and multi-tag queries, the top-20 retrieved fonts for each query are shown as Figure 2. These results demonstrate the effectiveness of our model to retrieve the top corresponding fonts for a wide range of tags.

Table 1. Top-200 frequent tags for the collected dataset.

decorative	display	headline	poster	sans-serif	magazine	modern	legible	retro	script
elegant	informal	serif	handwrite	logo	geometric	funny	vintage	contemporary	clean
bold	sans	alternate	package	sketch	brand	heavy	text	round	invitation
ligature	letter	advertise	fun	friendly	calligraphy	hand	rough	wed	swash
ornament	casual	cool	brush	handmade	fashion	calligraphic	commercial	narrow	cursive
book	signage	comic	grunge	fancy	art-deco	hand-drawn	editorial	corporate	fashionable
cute	condense	kid	organic	multilingual	feminine	monoline	slab-serif	connect	antique
cyrillic	pen	logotype	title	all-caps	german	news	ink	square	symbol
playful	formal	grotesk	soft	futuristic	child	humanist	thin	stylish	classic
scrapbook	technical	light	black	wide	simple	techno	ancient	food	unique
stencil	american	design	grotesque	dingbat	opentype	1930s	unusual	block	picture
cartoon	italic	small-caps	1800s	outline	computer	music	illustration	capital-sharp-s	web
versal-eszett	1950s	masculine	valentine	party	lively	sanserif	creative	irregular	romantic
curly	wood-type	blackletter	1960s	happy	1920s	label	greet	newspaper	gothic
print	distress	sharp	ornamental	fresh	sport	workhorse	delicate	capital	1940s
screen	cap	natural	brush-drawn	love	sign-painting	inline	hipster	engrave	menu
fat	art-nouveau	experimental	oldstyle	industrial	neutral	1970s	bouncy	crazy	wild
sign	caps-only	flourish	versatile	minimal	movie	french	linear	urban	roman
slab	film	publish	angular	modular	beautiful	art	expressive	texture	deco
graceful	mechanical	letterpress	paint	western	wild-west	dynamic	game	greek	draw

Figure 1. Group examples of the collected tagging set. For each group of a tag, the ground-truth font is in the red box.

decorative · **alternate** · **text** · **moden** · **legible** · **display** · **elegant** · **contemporary** · **headline** · **letter**

comic · **funny** · **calligraphy** · **rough** · **wed** · **swash** · **brush** · **fashion** · **signage** · **cursive**

Figure 2. Font retrieval results of the proposed model on typical single-tag and multi-tag queries.