

クラウドソーシングを用いたアイコンの主観的属性と視覚的特徴量の関係性の調査

吉谷 拓真^{1,a)} 矢谷 浩司^{1,b)}

概要: アイコンは GUI や資料製作において重要な要素である。同じテーマのものを描いたアイコンであっても、そのスタイル（例：フォーマル、かわいい、近代的）は多岐に渡りそれぞれ異なる印象を与える。しかし、どのような視覚的特徴量がアイコンの持つ主観的属性に寄与するのかに関しての十分な調査はなされていない。この関係性を明らかにすることによって、例えば主観的な概念を用いて望みのアイコンを検索できるシステムのような、アイコンデータとの新たなインタラクションが可能となる。本研究ではアイコンの主観的属性と視覚的特徴量の関係を定量的に分析した。まずアイコンのデータセットを Web 上のデータを用いて構築した。次にクラウドソーシングを用いて 3,380 個の主観的属性候補を収集し、手動で整理を行い最終的に 45 個の主観的属性を得た。更に、1 つの主観的属性と 2 つのアイコンコレクションを提示しどちらのアイコンコレクションがより与えられた主観的属性に合っているかに関する主観的評価のデータの収集も行った。このデータを機械学習のモデルに学習させ、アイコンの主観的属性と視覚的特徴量の関係性を明らかにした。その結果、アスペクト比、黒さ、線幅の一定度合いが主要な視覚的特徴量であることが分かった。本研究は視覚的特徴と主観的属性を結びつけるための定量的手法を提示するものである。

1. はじめに

アイコン（ピクトグラム）はインフォグラフィクス、プレゼンテーションスライド、看板など様々な資料や制作物において重要な要素である。これらの資料などの製作においてスタイルの似たアイコンを選ぶことは視覚的な一貫性を与えるために重要である。しかし、そのような作業には多くの試行錯誤やプロのデザイナーの経験が必要である。その結果、制作物とよく合うアイコンを探すことは非常に煩雑な作業となっている。更に、自由に使えるアイコンの数は現在ますます増えている。例えば The Noun Project^{*1} は 2 万個以上のアイコンコレクション（あるテーマのもとで複数のアイコンを束ねたもの）を提供している。人々のアイコンの選択肢は広がっている反面、これらを効率的に検索できるようにすることは未だ実現されておらず大きな課題となっている。

このようなアイコンデータのアクセシビリティの欠如はアイコンには標準的なデータセットが存在しないこと、ま

たスタイルに関する定量的な指標が存在しないことに起因する。O'Donovan ら [5] は主観的属性と視覚的特徴量を用いたクラスタリングによってフォントの検索に対する障壁を大きく下げること成功した。我々は同様の手法をアイコンにも使うことができると考えている。しかし、彼らの手法をそのまま適用することの可否に関しては未だ明らかになっていない。例えば、O'Donovan らは既存研究から抜き出した 37 個の主観的属性を用いたが、これをそのままアイコンにも使えるかに関しては追加の調査が必要である。アイコンはフォントに比べて視覚的多様性が非常に大きいためである。

本研究では O'Donovan ら [5] の手法を踏襲しつつ、アイコンの主観的属性と視覚的特徴量の関係を定量的に分析した。まず 4,828 個のアイコンコレクションを The Noun Project から収集し、アイコンの内容のテーマごとに手動で分類した。その結果 26 のテーマ、計 286 個のアイコンコレクションを得た。次に主観的属性の候補となる形容詞をクラウドソーシングを通して収集した。集まった 3,380 個の単語から形容詞のみを抜き出し、同じもしくは反対の意味を持つものを併合していくことで 45 個の形容詞からなる主観的属性集合を構築した。更に 1 つの属性につき 2 つのアイコンコレクションを提示し、どちらがより与えられた属性によってよく表されるかという質問への答えをク

¹ 東京大学大学院 工学系研究科
Interactive Intelligent Systems Laboratory,
Graduate School of Engineering, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan

a) takuma@iis-lab.org

b) koji@iis-lab.org

*1 <https://thenounproject.com/>

クラウドソーシングで収集した。これらのデータから機械学習モデルを構築することで、各アイコンコレクションについて各主観的属性がどのような相対値を持つのかを推定した。更に同じクラウドソーシングデータを用いて、主観的属性と視覚的特徴量の関係性の調査も行った。その結果、アスペクト比、黒さ、線幅の一定度合いが主要な視覚的特徴量であることを発見した。本稿では視覚的特徴量と主観的属性を結びつける定量的な手法について述べる。本研究はアイコンの主観的属性とそれらの視覚的特徴量の関係性に関する定量的な知見を提供するものである。

2. 関連研究

スタイル属性を応用したシステムは数多く研究されている。そのようなシステムを用いることでユーザは望みのスタイルを得るための煩雑な作業から解放される。Chaudhuriら [1] は 3D キャラクターやオブジェクトの予め指定された部分を、ユーザが形容詞の程度を表すスライダーを調節することによって入れ替えることができる *AttribIt* というシステムを開発した。このシステムの実現に必要なデータはクラウドソーシングを通じて収集された。まず形容詞を、次にそれらのオブジェクトに対する程度を評価したデータを収集した。これらのデータから機械学習を用いてオブジェクトの特徴量と形容詞の程度を関連付けた。形容詞の収集において、彼らはクラウドソーシングのワーカー達に 2 つの 3D オブジェクトを同時に見せそれらを比較させることで効果的に形容詞を引き出した。このようにして集められた各形容詞について 3D オブジェクト A, B の 2 つを同時に見せ、どちらがその形容詞によってよりよく表されるかという質問も用意した。これらの質問に関する応答もクラウドソーシングにて収集した。*AttribIt* ではこのようにして形容詞によって表現されるスタイルの量的な違いを収集・表現することができる。また、彼らはこの手法を Web デザインにも適用することで 3D オブジェクトだけでなく他のメディアにも使えることを示した。Yumerら [6] は同様の手法を用いて形容詞と 3D オブジェクトの関係を収集し、これらを用いて 3D オブジェクトを連続的に変形させることを試みた。彼らの開発したシステムには複数の形容詞の程度を表すスライダーが用意されている。ユーザはそれらのスライダーの値を調整することで 3D オブジェクトの見た目を連続的に変化させることができる。O'Donovan ら [5] は主観的属性の属性値をクラウドソーシングによって推定する方法を提案し、更にフォントのスタイルの距離関数も定義することでフォントを効果的に探索できるアプリケーションを製作した。本研究では彼らの手法をベースにアイコンのスタイルの定量化を行った。

3. アイコンデータセットの構築

アイコンデータセットの構築にあたり、The Noun

Project というアイコン共有 Web サービスからアイコンを取得した。The Noun Project ではアイコン製作者が複数のアイコンをまとめた「アイコンコレクション」を作ることができ、多くの場合それらは同じテーマやスタイルのアイコンがまとまっている。我々は単体のアイコンよりも複数のアイコンを同時に提示する方がスタイルという抽象的なものをより認識しやすいと考え、データセットの構築に際しアイコンコレクションを最小単位として用いることにした。我々はまず 12 個から 24 個のアイコンを含んでいるアイコンコレクション 4,828 個をクローリングにより収集した。その後、Amazon Mechanical Turk (AMT) を用いたクラウドソーシングと著者らの目視によって、含まれるアイコンのスタイルが全て同じであるアイコンコレクション 1,097 個を抽出した。次に、これらのアイコンコレクションをテーマごとに手動で分類し、それぞれが 11 個のアイコンコレクションを持つ 26 テーマを選択しアイコンデータセットとした。最終的なデータセットは 286 個のアイコンコレクションを含む。

4. 主観的属性集合の構築

アイコンの主観的なスタイルの属性の獲得にあたり、我々は O'Donovan らの手法 [5] を採用した。彼らの手法には 2 段階ある。1) 各フォントに属性値を割り当てる、2) フォントのスタイルの距離関数を作る。本研究では 1) を対象としている。アイコンの標準的なスタイル属性は今まで提案されてこなかったため、我々はアイコン用の属性集合を構築するところから始めた。属性集合の構築は形容詞の収集と合併によって行った。

形容詞の収集

形容詞の収集は Chaudhuri らの手法 [1] に従って行った。形容詞を効果的に引き出すために、我々はアイコンコレクションの比較セットを各テーマごとに 55 個、合計 1,430 個用意した。

データの収集には AMT を使用した。2 つのアイコンコレクション A, B を提示し、ワーカー達に質問文の空白部を埋める 3 つの形容詞を答えるよう指示した。質問文は「A is (more) ___ than B.」(「A は B よりも ___ である」) とその逆の 2 文を用意し、1 つのペアに対し合計 6 個の形容詞を収集した。各 Human Intelligent Task (HIT)*2 に 11 のタスクを含めた。回答の質を確保するため、各 HIT には 3 人のワーカーを割り当てた。合計 217 人のワーカーに作業を依頼した。各 HIT に対し 0.10 USD を支払った。

この結果 3,380 個の属性候補を得た。しかしそれらのうち多くは意味が似ているか、もしくは一度しか現れないものであった。そこで代表的な属性を得るために次に述べる合併を行った。

*2 AMT 上のタスクの単位。ワーカー達は HIT ごとに報酬を受取る。

表 1: 最終的な属性集合. 二値的なものと連続値的なものに分けた.

二値的	hand-drawn, outlined, shaded, symmetrical
	accurate, angular, attractive, bad, boring, charming, cheerful, chubby, clear, crowded, curvy, dark, elegant, energetic, expressive,
連続値的	formal, friendly, gentle, heavy, long, masculine, messy, modern, noticeable, organic, realistic, scary, sharp, shiny, silly, simple, smooth, soft, solid, strong, technical, thin, transparent, understandable, warm, wide

形容詞の合併

形容詞の合併は次の3段階で行った. 1) 前処理, 2) 出現頻度によるフィルタリング, 3) 手作業による意味的な分類. まず前処理として著者の一人が収集された形容詞を確認して明らかなスペルミスを修正した. 更に複数単語からなる回答はハイフンで繋ぐように統一し (例: attention-grabbing), 形容詞ではない単語を取り除いた (例: food). 形容詞でない単語の特定には NLTK (Natural Language Tool Kit) と WordNet の part-of-speech タグ機能を使用した. 以上の前処理の結果, 2,486 個の属性が残った.

次に出現頻度が低い属性を取り除いた. まず各属性を出現回数の大きい順に並べ, 総出現回数の 80% に達するまで先頭から和を取っていった. 和が 80% に到達するまでに使用した属性 441 個を抽出した.

更に属性を絞るために著者の一人と外部の一人とで属性を精査し, 同じもしくは反対の意味を持つと両者が合意した属性を, 出現回数が大きいものから順に手作業でまとめていった. その結果, 45 のグループが生まれた. 各グループごとに代表する属性を選び, それぞれ二値的か連続値的かを先程の二人で協議し決定した. 構築された主観的属性集合を二値的, 連続値的に分けてまとめた結果を表 1 に示す.

5. 属性値の割当

続いて各アイコンコレクションに対して, 属性値を求める. 属性値を決める単純な方法としては評価者の投票数をそのまま使うものが考えられる. 例えば 100 人中 90 人があるアイコンコレクションが “formal” という属性によりよく表されると合意した場合, その属性値を 0.90 と決めることができる. しかしながら, この方法ではアイコンコレクションと属性のすべての組み合わせに関してデータを集める必要があるため, データ収集の規模が膨大になってしまう. また, すべての評価者を等しく信頼できないという問題も存在する. そこで我々は O’Donovan ら [5] に倣い属性値の尤度を用いることにした.

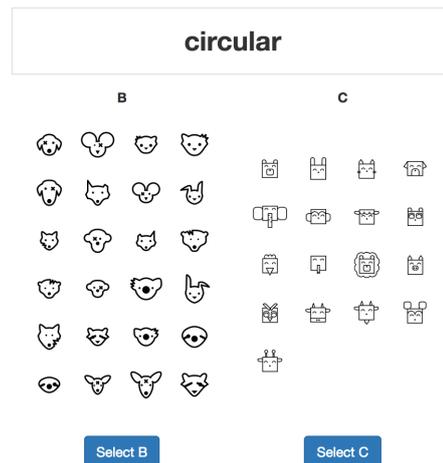


図 1: 属性値割り当てのためのデータ収集用インターフェースのスクリーンショット. ワーカーに 1 つの形容詞 (主観属性) と 2 つのアイコンコレクションを提示し, どちらのアイコンコレクションがより与えられた属性のスタイルをうまく表しているかを強制的に選択させる.

5.1 推定モデル

図 1 に示すように, AMT 上のワーカー達に 1 つの形容詞と 2 つのアイコンコレクションを提示し, その形容詞がどちらのアイコンコレクションをより表しているかという質問を与え片方を強制的に選択させる.

このようにして集まるデータに対し, 次のモデルを用いて属性値の推定を行う. 評価者 (ワーカー) の応答は $D = \{a, c_i, c_j, u, q\}$ として表される. a, c_i, u, q をそれぞれ属性, アイコンコレクション (ID: i), 評価者の ID, 評価者の回答とする. $q = 1$ は評価者が c_i の方が c_j よりも a によりよく表されると答えることを意味し, $q = 0$ はその逆を意味する. また, r_u を評価者の信頼度を表す未知数と, $v_{j,a}$ をアイコンコレクション j に対する a の属性値とする. 評価者の応答のモデルとして以下のロジスティック関数を用いる.

$$p(q = 1 | c_i, c_j, u, a) = \frac{1}{1 + \exp(r_u(v_{j,a} - v_{i,a}))} \quad (1)$$

全ての比較応答の集合 D が与えられた時, 負の対数尤度目的関数は以下の通りとなる.

$$\begin{aligned} E(\mathbf{v}, \mathbf{r}) &= -\ln p(D | \mathbf{v}, \mathbf{r}) \\ &= -\sum_k q^k \ln p(q = 1 | c_i^k, c_j^k, a^k, u^k) \\ &\quad - \sum_k (1 - q^k) \ln (1 - p(q = 1 | c_i^k, c_j^k, a^k, u^k)) \end{aligned} \quad (2)$$

$$(3)$$

ここで k は全訓練タプルに対するインデックスである. この目的関数を全属性値 \mathbf{v} と評価者信頼度 \mathbf{r} に関して Adam [3] により最小化する. 得られた属性値は $[0, 100]$ の間にスケールされる.

5.2 データ収集

41 個の連続的主観属性についてデータ収集を行った。各アイコンコレクションにつきランダムに 8 個のペアを生成し各属性と組み合わせることで計 93,808 個の三つ組が生成された。これらの三つ組に対する回答データを AMT を利用し収集した。各 HIT につき 110 個の三つ組質問を含ませた。これらのうち 5 個は答えが明らかなものであり、更に別の 5 個は他の三つ組と重複させたものである。今回は答えが明らかな 5 個の質問を品質保証に利用し、5 個中 3 個以上の質問に対して予め決められている答えと逆の回答をした応答を拒否した。実際の学習ではより厳しい基準を当てはめており、2 個の不正解を含む応答は破棄した。7,536 人のワーカーから計 18,342 の応答を得た。AMT 上での拒否率は 22.9%、実際の応答データの破棄率は 33.3%となった。この結果 12,242 個の承認された応答を得た。各三つ組に対する承認された応答の数の平均を取ると 13.0 となった。

5.3 モデルの学習

集まった全データを用いて繰り返し学習を行った。図 2 に求まった属性値の例を示す。また表 2 の二列目に各属性ごとの負の対数尤度の平均値と全属性での平均値を示す。

5.4 考察

図 2 に示されるように、多くの場合においては属性値の割り当ては妥当であるように見える。しかし、予期しない結果となることもあった。例えば三番目に大きな属性値は“long”となっているが、このアイコンコレクションに含まれるアイコンで長いものは見当たらない。この他のアイコンコレクションでも“long”が大きな属性値を持つケースが複数見られた。これは“long”という形容詞の意味についての共通理解が無かったこと、またクラウドソーシングを通じたデータ収集によりノイズが混入したことなどが原因であると考えられる。クラウドソーシングのタスクにおいてアイコンは予め決められた大きさの領域に表示されたこと、またどのアイコンも極端なアスペクト比を持っていないこと（最小：0.39、最大：2.75）からワーカー達は“long”なアイコンがどのようなものか理解することができなかった可能性がある。このことから、各属性に十分な定義を持たせることが今後の課題として発見された。

表 2 の二列目に示されるように、負の対数尤度は属性ごとに大きく異なる。負の対数尤度は小さいほど推定誤差が小さかったことを意味し、例えば“dark”などは推定が容易であったことを示す。この表を観察すると、見た目に関係する属性ほど負の対数尤度が小さくなるのが分かる。一方、“modern”などのより主観性の度合いが高い属性は負の対数尤度が高くなっていることが見て取れる。この傾向を説明する仮説として、主観性の度合いが高いほど評価者

間での意味合意が形成されにくかったことが考えられる。この仮説の検証も今後の課題とする。

6. 属性予測

次に属性と視覚的特徴量の関係について調査を行った。具体的には、28 次元の視覚的特徴量ベクトルから属性値を予測するようなモデルを構築した。

6.1 予測モデル

\mathbf{x}_i と \mathbf{w}_a をそれぞれ各アイコンコレクション (ID: i) に対し予め計算された視覚的特徴量ベクトルと属性 (a として示される) ごとの重みベクトルとする。これらの定数と変数を用いて評価者の三つ組への応答は以下のようにモデル化される。

$$p(q = 1 | c_i, c_j, u, a) = \frac{1}{1 + \exp(r_u(\mathbf{x}_j \cdot \mathbf{w}_a - \mathbf{x}_i \cdot \mathbf{w}_a))} \quad (4)$$

属性値割り当て時と同様、Adam [3] を用いて負の対数尤度関数による目的関数を属性ごとの重みベクトル \mathbf{w} と全評価者の信頼度 \mathbf{r} に関して最小化する。

6.2 視覚的特徴量

我々は以下の 19 個の視覚的特徴量を定義する。各特徴量の計算においては解像度 512×512 ピクセルの二値画像を使用した。

サイズと領域

各アイコンのバウンディングボックスを求め、それらの幅、高さ、アスペクト比を特徴量として計算した (それぞれ *bbox width*, *bbox height*, *bbox aspect ratio*)。各画像に含まれる黒画素の総数も計算し (*area*)、それをアスペクト比で割ったものも特徴として追加した (*area to bbox aspect ratio*)。

輪郭

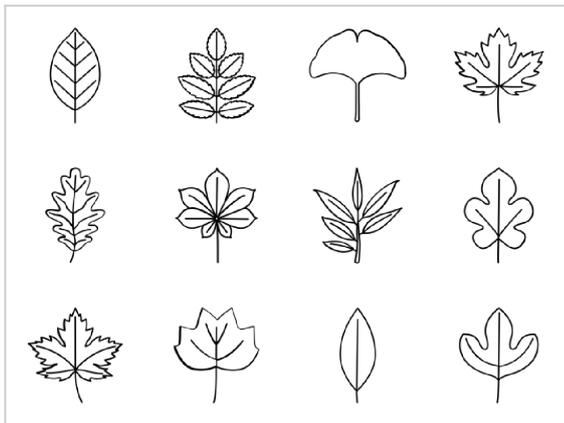
各アイコンの輪郭を求め、それらの個数 (*contour count*) と全周長 (*arc length*) を求めた。更に 10-bin 曲率ヒストグラム (*curvature histogram 0-9*) とそのエントロピー (*curvature histogram entropy*) も求めた。輪郭上の各サンプリング点において、その点と一定距離離れて隣り合う別の 2 つのサンプリング点からなる 2 つのベクトルから曲率を計算した。

黒さ

Liu ら [4] の研究においてはアイコンのスタイルを 4 つに分割するのに *blackness* が使用された。これは画像の凸閉包内の黒画素の割合として計算される。本研究においても特徴量として追加した。

コーナー

検出されたコーナーの数も特徴量として追加した (*corner count*)。



organic	(98.83)	friendly	(50.08)
elegant	(92.62)	energetic	(49.92)
long	(92.05)	angular	(47.04)
realistic	(91.76)	messy	(46.30)
attractive	(88.42)	expressive	(45.13)
accurate	(87.16)	technical	(43.38)
gentle	(86.38)	strong	(40.66)
thin	(81.80)	crowded	(40.64)
warm	(78.05)	masculine	(35.35)
charming	(77.31)	wide	(35.11)
soft	(75.94)	shiny	(32.54)
formal	(75.51)	dark	(27.50)
transparent	(74.43)	solid	(25.49)
clear	(69.78)	heavy	(24.53)
curvy	(68.01)	modern	(22.34)
cheerful	(63.82)	boring	(20.35)
noticeable	(63.66)	chubby	(19.93)
sharp	(58.85)	scary	(14.53)
understandable	(58.29)	bad	(14.22)
smooth	(57.72)	silly	(7.13)
simple	(50.23)		

図 2: 推定された属性値の例。左にアイコンコレクション、右にそれに対しモデルが割り当てた属性値を降順に示す。“long”が上位に来るなど誤った例も見られるが、概ね理にかなった属性値が割り当てられていることがわかる。

線もしくは領域の幅

Stroke Width Transform (SWT) [2] SWT アルゴリズムは二値画像から、その画像と同じサイズで各画素がその画素が含まれる領域の幅の値であるような画像を生成する。大きさの影響を避けるため、各アイコン画像につき4つの異なる大きさ(100%, 50%, 25%, 12%)に変換し、それぞれ SWT を計算し結果の画像の平均と標準偏差を求め、特徴量とした (SWT 100/50/25/12 mean と SWT 100/50/25/12 std)。

曲率ヒストグラムが10次元ベクトルなので、以上の特徴量すべて合わせて28次元のベクトルが得られた。

6.2.1 モデルの学習

三つ組比較データの上でモデルを繰り返し学習させた。その結果、*bbox aspect ratio*, *area*, *area to bbox aspect ratio*, *blackness*, *SWT 100 std*, *SWT 50 std*, *SWT 12 mean* などの視覚的特徴量が属性予測において支配的であることが分かった。また各属性における分類率を表2の三列目に示す。なお分類率は O'Donovan ら [5] が予測モデルの性能指標として導入したもので、1つのアイコンコレクションに関する全ての比較データを取り出しテストデータとし、それ以外の比較データを訓練データとしモデルを学習させた上でテストデータとモデルの予測の合致率(モデルが人間の評価者と同じ評価を下した率)として計算される。

6.3 考察

まず属性予測においていくつかの属性と強い相関を示した視覚的特徴量について考察する。*bbox ratio* (バウンディングボックスのアスペクト比) はアイコンの内容と深い関係を持つ可能性がある。例えば音楽プレイヤーの再生ボタンなどを含むテーマにおけるアイコンは多くの場合正方形もしくは円形の外形を持つため *bbox ratio* は似たような値を取ると考えられる。この関係性の調査は今後の課題

である。*area*, *area to bbox ratio*, *blackness* はアイコンの黒さを表す。SWT の平均は黒い領域の幅を示し、標準偏差はアイコン内の線もしくは領域の幅が一定かどうかを示す。平均よりも標準偏差の方が属性予測においてよく効いていたことから、線幅が一定であることは知覚されるアイコンのスタイルにより寄与することが考えられる。

表2の三列目にはテストデータに対する分類率が示されている。表自体は属性値割り当ての負の対数尤度の昇順に並んでいるが分類率も概ね同じ順番で降順になっていることが分かる。5.4で既に考察したように、属性の主観性の度合いが高ければ評価者間での意味理解の相違がより大きく、それが分類値の悪さにも反映されていると考えられる。更に、視覚的特徴量の観点から、主観性の度合いが高い属性は視覚的特徴量と属性の意味のつながりが複雑であるということも考えられる。例えば“dark”という属性は明らかに黒さを表す視覚的特徴量と深い関係を持つが、“modern”と深い関係を持つ視覚的特徴量が何かということに関しては人によって意見が別れると考えられる。視覚的特徴量と属性の主観性度合い関係性について今後の研究で更なる調査が必要である。

7. おわりに

アイコンのスタイルを定量化することによりアイコンデータをより便利に活用することができるようになる。本研究ではフォントに関する既存研究の手法を拡張し、アイコンの主観的属性と視覚的特徴量に関して定量的な分析を行った。

そのためにまず286個のアイコンコレクションからなるデータセットを構築し、45の形容詞からなる主観属性集合を構築した。100万個以上のアイコンを提供する The Noun Project から1097個のアイコンコレクションを選定し、これらのアイコンコレクションを意味ごとに手動で分類・選定することで286個のアイコンコレクションか

表 2: 属性ごとの属性値割当て時の負の対数尤度と、属性予測時の分類率の一覧。負の対数尤度の昇順に並べている。負の対数尤度は低い方が誤差が小さく、また分類率は高いほど予測精度が高いということを表す。“dark”, “thin” など視覚的特徴量と直結しそうな比較的客観的な属性ほど割り当て誤差が小さくまた予測精度も高くなり、逆に“modern”などのより主観性の高い属性は誤差が大きく精度も低くなっていることが分かる。

属性	負の対数尤度	分類率
dark	0.2532	0.8773
heavy	0.3929	0.8399
solid	0.3974	0.8491
thin	0.4466	0.8300
transparent	0.4522	0.8514
strong	0.4566	0.8073
silly	0.4624	0.6348
cheerful	0.4740	0.6699
technical	0.4883	0.6905
boring	0.4929	0.6758
organic	0.4942	0.7063
expressive	0.4944	0.6376
noticeable	0.4960	0.7509
friendly	0.4972	0.6667
curvy	0.5026	0.6684
charming	0.5058	0.6953
chubby	0.5090	0.6904
realistic	0.5128	0.6636
angular	0.5235	0.7083
scary	0.5282	0.6542
gentle	0.5295	0.7462
masculine	0.5306	0.6677
soft	0.5343	0.7537
long	0.5407	0.6422
crowded	0.5408	0.6355
wide	0.5419	0.6535
attractive	0.5435	0.6823
energetic	0.5476	0.6021
warm	0.5513	0.6650
simple	0.5516	0.6755
understandable	0.5659	0.6031
bad	0.5690	0.6535
formal	0.5695	0.6224
elegant	0.5699	0.6352
clear	0.5757	0.7054
sharp	0.5870	0.6757
accurate	0.5887	0.5786
messy	0.5927	0.5980
shiny	0.5962	0.6154
modern	0.6162	0.6569
smooth	0.6190	0.6403
平均	0.5181	0.6897

らなるデータセットを構築した。主観的属性集合はクラウドソーシングを通じて集めた 2,486 個の形容詞から構築した。出現頻度の低いものを除き、更に意味ごとに手で併合することで最終的に 45 個のアイコンの主観的属性を決定した。

アイコンの主観的属性と視覚的特徴量の関係性を調査するにあたり、2 つのロジスティクスモデルをクラウドソーシングから得た大規模比較データを用いて学習したまず各アイコンコレクションについて属性値を割り当てた。次に予測モデルを用いて視覚的特徴量と属性の関係を調査した。その結果、アスペクト比、黒さ、線幅の一定さなどが主観的属性の予測において大きく寄与することが分かった。

今後は本研究の結果を利用したアプリケーションの開発を行っていく予定である。例えばスタイルを表す主観的属性を用いてアイコンを検索できるシステムを構築することを考えている。このようなシステムの開発と評価を通じて今回の我々の発見の有用性に対する更なる評価を下すことができる。

参考文献

- [1] Siddhartha Chaudhuri, Evangelos Kalogerakis, Stephen Giguere, and Thomas Funkhouser. 2013. Attribit: Content Creation with Semantic Attributes. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13)*. ACM, New York, NY, USA, 193–202. DOI: <http://dx.doi.org/10.1145/2501988.2502008>
- [2] B. Epshtein, E. Ofek, and Y. Wexler. 2010. Detecting text in natural scenes with stroke width transform. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2963–2970. DOI: <http://dx.doi.org/10.1109/CVPR.2010.5540041>
- [3] D. P. Kingma and J. Ba. 2014. Adam: A Method for Stochastic Optimization. *ArXiv e-prints* (Dec. 2014).
- [4] Yiming Liu, Aseem Agarwala, Jingwan Lu, and Szymon Rusinkiewicz. 2016. Data-driven Iconification. In *Proceedings of the Joint Symposium on Computational Aesthetics and Sketch Based Interfaces and Modeling and Non-Photorealistic Animation and Rendering (Expresive '16)*. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 113–124. <http://dl.acm.org/citation.cfm?id=2981324.2981341>
- [5] Peter O'Donovan, Jānis Lībeks, Aseem Agarwala, and Aaron Hertzmann. 2014. Exploratory Font Selection Using Crowdsourced Attributes. *ACM Trans. Graph.* 33, 4, Article 92 (July 2014), 9 pages. DOI: <http://dx.doi.org/10.1145/2601097.2601110>
- [6] Mehmet Ersin Yumer, Siddhartha Chaudhuri, Jessica K. Hodgins, and Levent Burak Kara. 2015. Semantic Shape Editing Using Deformation Handles. *ACM Trans. Graph.* 34, 4, Article 86 (July 2015), 12 pages. DOI: <http://dx.doi.org/10.1145/2766908>