

プレゼンテーションスライドの客観評価と印象予測 Evaluation and Impression Prediction of Presentation Slides

大山真司[†]
Shinji Oyama

山崎俊彦[†]
Toshihiko Yamasaki

相澤 清晴[†]
Kiyoharu Aizawa

1. はじめに

プレゼンテーション (以下プレゼン) は意見や情報を伝える一般的な手段である。多くの場合、発表者は事前にスライド資料を作成し、補助に用いてプレゼンを行う。良いプレゼンを行うためには、話す内容や話し方 (話す速度やジェスチャー) や表情などに気を配るだけでなく、分かりやすい (あるいは聴衆の関心を引く魅力的な) スライドを作ることが重要である。しかし、自力で分かりやすいスライドを作るのは困難である。スライド作成を支援する書籍やウェブページは数多く存在するものの、作成したスライドに対する客観的な評価やフィードバックが得られないため改善が難しい。そのため多くの場合、作成したスライドを身近な人間に見せることで改善のためのフィードバックを得る。この方法であればスライドの改善は可能であるが、手間である。

そこで本研究は、分かりやすいプレゼンスライドを自力で作成するための支援システムの構築を最終目標に見据え、要素技術の一つとして、スライドの客観的印象予測と、スライドの印象に影響を与える要因解析を目指す。本稿ではまず、同じような内容を表現する場合でも表現方法によってスライドの印象が変化すること、一般的に「分かりやすいスライド」が存在することを実験的に示すことで、本研究の意義を確認する。さらに、スライドの画像特徴、構造特徴、内容特徴を組み合わせることで、スライドが分かりやすいか分かりにくいかの 2 クラス分類問題を 74% の精度で行うことができた。

2. 関連研究

文書やテキストデータから (半) 自動でプレゼンスライドを生成する手法が提案されている [1, 2, 3, 4, 5, 6, 7, 8, 9]。一般的にスライドの自動生成は元文書の要約問題として扱われ、箇条書きのスライドを生成する。自動生成されたスライドは第一稿作成の手間を軽減するものの、実際にプレゼンに用いるためには事前に人手による修正が必要であるため、自力でスライドを改善する難しさは依然として解消されていない。また、数式や図表に対しては重要度を計算できないため、スライドに自動で掲載することができない問題もある。

スライドの構造解析を行った研究として、桜木ら [10] は、Office Open XML [11] の情報を用いてスライド中のオブジェクト同士の意味関係を平均 70% の精度で推定した。また、Saleh ら [12] は、インフォグラフィック同士の類似度をメトリックラーニング [13] で学習する手法を提案した。一般的な物体認識で成功している種々の画像特徴量の比較を行い、Color histogram とセ

ルサイズ 16 で計算した HoG-16 [14] を組み合わせた特徴量が、インフォグラフィックのスタイル表現において最も高性能であるという結果を示した。

プレゼンの印象評価を行う研究として、福島ら [15] は、大規模な講演会である TED* のウェブサイト公開されているプレゼン動画を用いて、視聴者が抱く 14 個の印象 (beautiful, confusing, etc.) の予測を行った。内容や音声から抽出した各種の特徴量を用いて予測を行った結果、各印象の投稿率が上位・下位 10% のプレゼンに対し、各印象を抱くか抱かないかの識別を平均で 91.1% の精度で行えることを示した。また、Luzardo ら [16] は、スライドと音声を用いて、学生のプレゼンテーションが指導教員から受ける評価が高いか低いかの 2 クラス間の識別を行った。その結果、スライドを用いた場合は 65%、音声を用いた場合は 69% の精度でプレゼンテーションの良し悪しを判定可能であることを示した。さらに、スライドの場合は単語数、図表の量、最大フォントサイズが識別に与える影響が大きく、音声の場合は filled pauses (“Uh...”, “um...” など) や話すピッチの平均・変化が識別に与える影響が大きいことを明らかにした。スライドに直接評価を与える研究として、栗原ら [17] はテキスト情報からスライドの良し悪しを定量的に評価する SlideChecker を提案したが、日本語特有の性質を用いた特徴を使用しているため日本語以外で書かれたスライドには対応していない。

本稿では、PowerPoint で作られたスライド 1 ページに対し、スライドが分かりやすいか分かりにくいかの 2 クラス分類を行う。スライドの詳細情報が記述されている XML から得られる構造特徴や内容特徴を画像特徴と組み合わせることで、スライドの印象予測を高精度で行う。

3. スライド表現が印象に及ぼす影響

本章では、同じトピックの内容を伝えるスライドでも、表現の違いで分かりやすさに差が出ることを実験的に確認する。同じトピックを元に作られた複数のスライドを分かりやすい順に順位付けした時、異なる人間が与えた順位同士に相関があるかどうかを調査する。

本実験では、専門的でない以下の 5 つのトピックを設定し、それぞれに関する文章 (または質問) を用意した。トピック (A), (C), (D) の文章を例として図 1 に示す。

- A) Introduction (Albert Einstein)
- B) Grand Slam (tennis)
- C) Eyeglasses & Contact Lenses
- D) Rock-paper-scissors

[†]東京大学

*<https://www.ted.com/>



図 1: スライド生成に用いるトピックごとの文章例

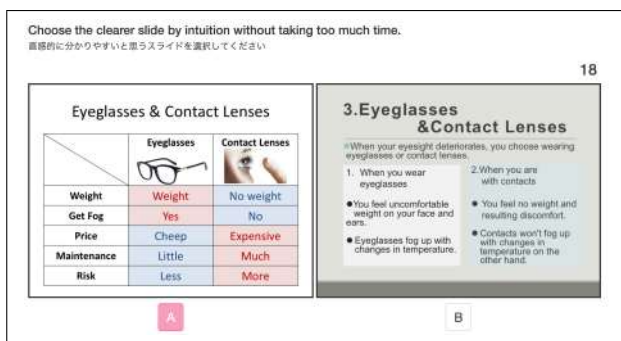


図 2: スライド比較画面

E) Windows vs Macintosh (open question)

これらの文章・質問に対するスライドの作成を作業者に依頼し、各トピックにつき 30 枚のスライドを得た。作業者は CrowdWorks [18] のワーカー 15 名と東京大学の大学院生 15 名で、トピックごとにスライドを 1 ページずつ作成するように依頼した。

次に、収集したスライドに対し、各トピックの 30 枚ごとに分かりやすい順に順位付けを行った。1 枚のスライドに対して絶対的な評価を与えることは一般的に難しいが、相対的な印象評価であれば比較的容易であろうと考え、一対比較法により全体の順位を計算した。スライドの分かりやすさに関する得点分布のパラメータを動的にアップデートすることで、通常より少ない比較回数で順位を収束させることができる、Chen ら [19] の手法を用いて順位を算出した。比較は、図 2 に示す画面でスライドを 2 枚ずつ東京大学の大学院生に提示し、直感的に分かりやすいと感じた方のスライドを選択するように指示した。1 つのトピックのスライド 30 枚に対し、2 人の作業者が 250 ペアずつ比較を行うことで順位を収束させた。各トピックごとに異なる 4 人が比較を行うことで、2 回ずつ順位を算出した。

各トピックごとに 2 回ずつ求めた順位同士について計算した Spearman の順位相関係数を表 1 に示す。無相関検定における有意水準 (両側確率) 1% の限界値は 0.467 なので、全トピックのスライドに対して順位相関

の有意性が認められた。この結果から、同じトピックを伝えるスライドを作成する場合でも、表現方法次第でスライドの分かりやすさが変化すること、スライドが分かりやすいかどうかは個人の主観に依存するだけのものではなく、一般的に分かりやすいと感じるスライドと分かりにくいと感じるスライドが存在することが確認された。

30 人のスライド作成者の中のある 6 人 (a, b, c, d, e, f) が作ったスライドの、トピックごとの平均順位を表 2 に示す。表 2 を見ると、あるトピックで順位の高いスライドを作成した作業者のスライドは他のトピックでも評価が高く、あるトピックで順位の低いスライドを作成した作業者のスライドは他のトピックでも評価が低いことが分かる。同様の傾向は作業者全体に対しても認められた。このことから、トピックに関わらず分かりやすいスライドを作ることができる人とできない人が存在することが分かった。分かりやすいスライドの特徴を学習することで、スライド作りが苦手な人の支援が可能であると考えられる。

表 1: 各トピックのスライドに対する Spearman の順位相関係数 (r_s)

Topic	r_s
A	0.717
B	0.622
C	0.653
D	0.610
E	0.652

トピック (A), (C), (D) について、分かりやすさの順位 Top-3 と Worst-3 のスライドを図 6 に示す。評価の高いスライドは文章量が少なく、表や図を用いて視覚的に分かりやすいように工夫されている一方で、評価の低いスライドは文章量が多く、フォントサイズや背景とのコントラストが小さいために視覚的に分かりにくいスライドであることが直感的に分かる。

表 2: 各作業者のスライドのトピックごとの平均順位

	topic A	topic B	topic C	topic D	topic E
worker a	1	3.5	2.5	4	2.5
worker b	2.5	2	3	9	2.5
worker c	8	6	4	3.5	2
worker d	22.5	27.5	26	25	21
worker e	24.5	20.5	26.5	24.5	26.5
worker f	27.5	28	30	29	29

4. スライドの特徴抽出

スライドから特徴を抽出する先行研究 [16] では、プレゼンに用いるスライド全ページをまとめて扱い、総単語数や最大フォントサイズ等の大域的な特徴を用いていた。しかし、身近な人間からフィードバックを得てスライドを修正する場合、直感的な印象・分かりやすさの面での修正は 1 ページ単位で行われることが多い。そのため、スライド全ページに対して大域的な特徴を用いて印象予測を行った場合、たとえ高精度な予測が実現しても、改善のためのフィードバックを個人に与えることが難しい。そこで本研究では、スライド 1 ページ内の特徴を利用し、スライド 1 ページの分かりやすさを予測する問題に取り組む。

本稿では、スライドの印象予測に用いる特徴を以下の 3 種に大別する。

- 画像特徴
- 構造特徴
- 内容特徴

本章では、これら 3 種の特徴ごとに、本研究でスライドの印象予測に用いる特徴量の詳細を述べる。本研究では、実験データとしてスライドの PowerPoint ファイルを収集しており、以下の特徴も PowerPoint で作成されたスライドから得ることを前提としている。

4.1. 画像特徴

PowerPoint で作成されたスライドを画像出力し、スライドの画像特徴を得る。画像特徴は、ハンドクラフトなアプローチと、畳み込みニューラルネットワークを用いたアプローチにより得る。

4.1.1. ハンドクラフト特徴

Saleh ら [12] は、インフォグラフィックの類似度学習に対し、画像認識に用いられる一般的な特徴量を利用し、その性能を比較した。その結果、Color histogram とセルサイズ 16 で計算した HoG を組み合わせた特徴

量 (Color histogram + HoG-16) を用いた場合が最も高性能であることを示した。

インフォグラフィックで高い性能を発揮した Color histogram + HoG-16 特徴は、スライドの直感的な分かりやすさを予測する場合でも有効と考えられる。Saleh ら [12] に倣い、Color histogram, HoG-16 それぞれを主成分分析 (PCA) により次元削減してから結合した 230 次元のベクトルを、本稿で扱うハンドクラフト特徴とする。ただし、印象評価を与えたスライドは本稿執筆時点で 150 枚であったため、未評価のスライドを加えた 1000 枚に対して特徴抽出をした後、PCA を行った。以下、PCA による次元削減は同様の方法とする。

4.1.2. 深層特徴

近年、深層学習は機械学習の様々なタスクでその有用性を示している。特に画像認識分野においては、畳み込みニューラルネットワーク (CNN) [20] の登場と、ImageNet [21] などの大規模教師付きデータセットがいち早く整備されたことにより、多くのタスクで驚異的な性能を発揮し、注目を集めるようになった。また、これらの大規模データセットを用いて適切に学習させた CNN のパラメータは汎用性が高く、高性能な特徴抽出器として他のタスクに転用可能であることが知られている。物体認識の精度を競うコンテストである ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [22] で 2015 年に優勝した Deep Residual Network (ResNet) [23] は、既存モデルの中で高い性能を発揮している。

本稿では、公開されている 50 層の ResNet (ResNet-50) [24] を特徴抽出器として用いた、以下の 4 種の特徴を扱う。

- Pretrained ResNet-50 [2048 dim]
- Pretrained ResNet-50 [256 dim]
- Fine-tuned ResNet-50 [2048 dim]
- Fine-tuned ResNet-50 [256 dim]

Pretrained ResNet-50 [2048 dim] は、公開されている ImageNet2014 による学習済みモデルの pool5 層の 2048 次元出力を特徴としたものであり、Pretrained ResNet-50 [256 dim] は、それを PCA で 256 次元にしたものである。また、Fine-tuned ResNet-50 [2048 dim] は、スライドが分かりやすいか分かりにくいかの 2 クラス分類タスクで学習済みの ResNet-50 を fine-tuning したモデルの pool5 層の 2048 次元出力を特徴としたものであり、Fine-tuned ResNet-50 [256 dim] は、それを PCA で 256 次元にしたものである。fine-tuning するタスクの詳細については、5.2 章で述べる。

4.2. 構造特徴

PowerPoint で作成されたスライドの情報は、全て Office Open XML [11] の PresentationML で記述されている。PresentationML を解析すると、図表、テキストボックスの位置と数、アニメーションの有無から文字のタイプフェイスやフォントサイズまで、スライドに関する詳細な情報を得ることができる。

PresentationML では、スライド中のオブジェクトを `<p:shapeTree>` タグの子要素として記述しており、主

要なオブジェクトを `<p:cxnSp>`, `<p:graphicFrame>`, `<p:pic>`, `<p:sp>` の 4 通りに分類している。`<p:cxnSp>` は `<p:sp>` を結ぶコネクションオブジェクト (矢印など), `<p:graphicFrame>` は表オブジェクト, `<p:pic>` は画像オブジェクト, `<p:sp>` は主にテキストオブジェクトを表す。本研究では, 図 3 に示すようにスライドを 3×3 の 9 点でサンプリングし, 各点が属する要素を表す特徴を, 上述した 4 通りのオブジェクトタイプに背景を加えた以下の 5 次元 one-hot ベクトル (1 つの要素だけが 1 であとは 0) で表す。

- bg (背景)
- cxnSp (コネクションオブジェクト)
- pic (画像オブジェクト)
- graphicFrame (表オブジェクト)
- sp (テキストオブジェクト)

サンプリング点の数は, 6×8 の 48 点で特徴抽出した場合と比較した結果, 5 章で述べる印象予測の精度が変わらなかったため 9 点でのサンプリングとした。こうして得られた 9×5 次元の特徴ベクトルを, スライドのレイアウト情報として用いる。さらに, スライド 1 ページ全体に対して以下の 6 次元特徴を用いる。

- is Bullet point (BP)
- Number of objects (NO)
- Number of words (NW)
- Maximum font size (MAXFS)
- Minimum font size (MINFS)
- Number of colors (NC)

BP は, スライドが箇条書きかどうかの 2 値である。色情報はカラーコードを取得することも可能だが, 本稿で提案する構造特徴は色情報の扱いに長けた他の画像特徴と組み合わせるに留めた。以上を統合した 51 次元のベクトルを, スライド 1 ページの構造特徴 (Structure) とする。

4.3. 内容特徴

スライドの内容を表現する特徴として, Mikolov らの手法 [25] を実装した word2vec による特徴量を用いる。Mikolov らのモデルは, ある単語の学習の際に文書内での単語の共起ではなく前後の数個の単語のみを用いるため, 文法情報が保持され, 文書全体の情報が入りにくいという特徴がある。単語周辺の文脈から単語を推定する CBOW (Continuous Bag-of-Words) モデルと, 単語から文脈中の単語を推定する Skip-gram モデルが存在するが, 本手法では後者を用いる。Wikipedia 英語版の記事群を用いて学習させた, 単語ベクトルが 500 次元, window サイズが 10 の word2vec を用いる。

PresentationML を解析することでスライド中のテキスト内容を取得し, 出現する全単語のベクトル和をスライドの内容特徴とする。スライド中に出現する全単語のベクトル和を計算し長さを 1 に正規化した特徴と比較した結果, 正規化をしない方が性能が良かったので, 本稿では正規化は行わずに用いる。

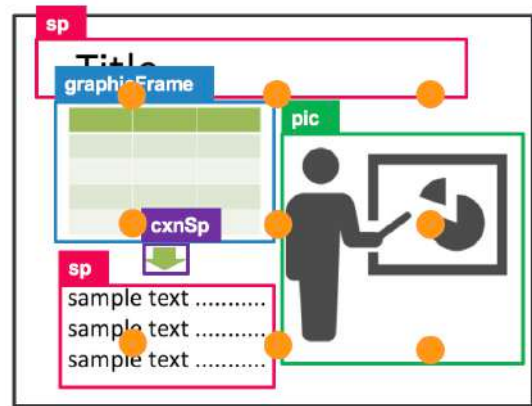


図 3: Structure 特徴

5. 印象予測実験

5.1. 実験概要

PowerPoint で作成された 1 ページのスライド (pptx, jpeg) を入力として, スライドが分かりやすいか分かりにくいかの 2 クラス間の識別を行う。実験データとしては, 3 章の実験で用意した 5 トピック * 30 枚のスライドを用いる。スライドには各トピックごとに 30 位までの順位が与えられているので, トピックごとに 2 分割し, 上位の 15 枚を Good クラス, 下位の 15 枚を Bad クラスとした。全体として, Good クラス, Bad クラスそれぞれ 75 枚である。

5.2. ディープラーニングによる印象予測

一般的に, ニューラルネットワークの学習には大量のデータが必要である。高い識別性能を狙うためには, さらにスライドデータセットを充実させる必要があると考えられるが, 本稿では予備実験として, ImageNet2014 データセットで学習済みの ResNet-50 [24] モデルを基盤に用いてスライドの分かりやすさの予測を行った。本稿では, 以下の 4 パターンで実験を行った。

- ResNet-50
- ResNet-50 + word2vec
- ResNet-50 + Structure
- ResNet-50 + word2vec + Structure

PowerPoint で作成されたスライドを画像出力し, 224×224 にリサイズして ResNet-50 の入力とした。ResNet-50 は, ResNet-50 の pool5 層の後を全結合して 2 次元にし, ソフトマックスの交差エントロピーを損失関数として学習を行った。ResNet-50 + word2vec は, ResNet-50 の pool5 層の 2048 次元の出力に, 4 章で述べた構造特徴である 500 次元の word2vec を加えた 2548 次元のユニットを, 2000 次元の中間層 1 層, 2 次元の出力層を持つ MLP につなげ, 同様に学習させた。同様に, ResNet-50 + Structure と ResNet-50 + word2vec + Structure も, ResNet-50 の pool5 層の 2048 次元出力に追加の特徴を加え, MLP につなげたネットワークアーキテクチャとした。

5 分割交差検定を行った結果を表 3 に示す。単純にスライド画像を入力とした ResNet-50 では、70% の精度で判別できた。一方で、word2vec や Structure を加えた場合は精度が下がってしまった。特徴を組み合わせても思うように精度が向上しなかった理由の一つとして、ResNet-50 の pool5 層の出力に単純に word2vec や Structure 特徴を加えることで、単に画像特徴のノイズになってしまった可能性が考えられる。お互いの特徴の良さを活かすためには、今後ネットワークアーキテクチャを工夫していく必要がある。構造特徴のレイアウト特徴については、テキストボックスや図表のレイアウトマスク画像が作れるので、これらを ResNet-50 の入力として一緒に畳み込む手法が考えられる。特徴の組み合わせで精度が向上しなかったもう一つの理由としては、当初の想定通り、スライドデータの不足により学習が上手くいかなかったことが考えられる。

表 3: ディープニューラルネットワークによるスライドの印象予測結果

Feature	Accuracy
ResNet-50	0.70
ResNet-50 + word2vec	0.67
ResNet-50 + Structure	0.67
ResNet-50 + word2vec + Structure	0.63

5.3.SVM による印象予測

4 章で述べた以下に示す特徴と、それらを組み合わせた特徴を用いて、Support Vector Machine (SVM) によるスライドの分かりやすさの 2 クラス識別を行った。

- 画像特徴
 - Color histogram + HoG-16 [230 dim]
 - Pretrained ResNet-50 [2048 dim]
 - Pretrained ResNet-50 [256 dim]
 - Fine-tuned ResNet-50 [2048 dim]
 - Fine-tuned ResNet-50 [256 dim]
- 構造特徴
 - Structure [51 dim]
- 内容特徴
 - word2vec [500 dim]

識別には RBF カーネルの SVM を用い、5 分割交差検定の精度による評価を行った。また、実験には LibSVM [26] を用いた。

5 分割交差検定の精度を図 4 に示す。上記の特徴を単体で用いた場合、Fine-tuned ResNet-50 [2048 dim] が 71% で最も高精度であった。画像特徴同士を比較すると、インフォグラフィックの類似度学習で高性能である Color histogram + HoG-16 [230 dim] は 62% であり、ResNet-50 の 2048 次元の特徴には 6% 以上及ばな

い結果であった。次元削減した ResNet-50 の特徴と同等以上の精度は出ているものの、スライドを表現する画像特徴としては、ハンドクラフトな特徴よりも深層特徴を利用する方が好ましいと言える。本実験の結果では、Fine-tuned ResNet 特徴が Pretrained ResNet 特徴に対して優位に働く全体的な傾向は掴めなかった。データが少なかつたために fine-tuning が十分になされなかつた可能性がある。

word2vec [500 dim] は、スライドのテキスト情報のみを用いているにも関わらず、単体でも 67% と比較的高精度であった。図 6 に示すように、分かりやすいスライドは文章が短く、分かりにくいスライドは文章が長い傾向が見て取れる。word2vec は、短いキーワードで簡潔にまとめたスライドと長い文章で冗長にまとめたスライドを区別できるため、高精度で識別できたと考えられる。

Structure [51 dim] は、単体での精度は 61% であるが、他の特徴と組み合わせることで高い精度を発揮した。word2vec + Structure は、71% であり、次元削減により精度が落ちた ResNet-50 特徴の精度を 8% 以上高めた。本実験で選択したスライドの Structure 特徴は、他の特徴では正確に取得できない詳細な情報を含むため、他の特徴と組み合わせるの有用である。

全体で最も精度が高かつたのは、Pretrained ResNet-50 [2048 dim] + word2vec で 74% であつたが、全体の傾向としては、ResNet-50 + word2vec + Structure が最も高精度であり、各 ResNet-50 特徴を用いた場合の平均で、72% の精度であつた。

本実験は、データ数、バリエーションに乏しい小規模なデータセットを用いて行ったものであつたので、大量のスライドとその印象情報があれば、さらに精度を向上可能であると期待できる。さらに、本実験ではデータが少なかつたためスライドを 2 分割して上位を Good クラス、下位を Bad クラスに割り当てたが、分割の境界付近のスライドは良し悪しが曖昧であるため、上位 10% のスライドを Good クラス、下位 10% のスライドを Bad クラスのように割り当てて実験をした方が印象予測の精度が向上すると考えられる。また、Structure 特徴は設計により改善の余地があるので、よりスライドの性質を表現できる特徴ベクトルを、今後実験的に見つけていく必要がある。

5.4. スライド特徴空間の可視化

本稿で提案したスライド特徴が、どの程度スライドを表現できているかどうかを直感的に知るため、多次元特徴空間を PCA で次元削減することにより、二次元平面上に描画した。この結果を図 5 に示す。青い点は Good クラスのスライド、赤い点は Bad クラスのスライドを表している。図 5 を見ると、Color histogram + HoG-16 と word2vec は、特に両クラスが混在している一方、Structure と Pretrained ResNet-50 [256 dim] は、比較的両クラスを分離できているように見える。また、word2vec は Pretrained ResNet-50 [256 dim] と組み合わせることで、両クラスの特徴をより良く表現できている。

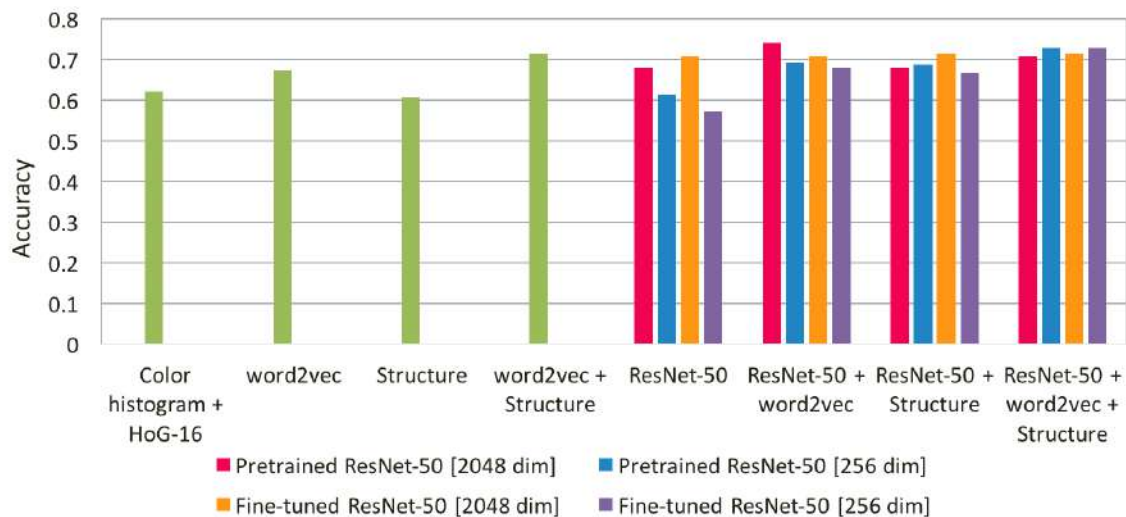


図 4: SVM によるスライドの印象予測結果

6. おわりに

本稿は、同じトピックを伝えるスライドでも表現方法によって分かりやすさが変わり、一般的に分かりやすいスライドと分かりにくいスライドが存在することを実験的に示すことで、スライドの表現スタイルの改善支援の重要性を確認した。実験のため、5トピックごとに PowerPoint スライドを 30 枚収集し、一対比較法で順位付けを行うことにより、分かりやすさに対する主観評価のラベルを与えたスライドデータセットを構築した。また、スライドを表現する特徴として画像特徴・構造特徴・内容特徴を提案し、それらの特徴を用いてスライドが分かりやすいかどうかの印象予測を行った。ディープラーニングによるアプローチでは、スライド画像を入力とした 2 クラス分類で、70% の精度でスライドの分かりやすさを判別できたが、構造特徴や内容特徴を単純に組み合わせただけではそれ以上の精度向上とはならなかった。SVM による 2 クラス分類では、提案した特徴を組み合わせることで単体で用いた場合よりも精度を向上させ、74% の精度で識別を行った。

今後はスライドデータセットをさらに充実させ、さらに高精度での印象予測と、印象に影響を与える特徴の解析を行っていく。また、本稿ではスライドの印象を直感的な分かりやすさと定義したが、面白さや知的好奇心など、多軸での印象評価・要因解析を行う予定である。

謝辞 本研究の一部は科研費 (26700008), JST-CREST (JPMJCR1686), MS CORE13 の支援を受けている。

参考文献

- [1] SM. Al. Masum, M. Ishizuka and Md. T. Islam. 'auto-presentation': a multi-agent system for building automatic multi-modal presentation of a topic from world wide web information. In *IEEE/WIC/ACM International Conference on Intelligent Agent Technology*, pp. 246–249. IEEE, 2005.
- [2] SM. Al. Masum and M. Ishizuka. Making topic-specific report and multimodal presentation automatically by mining the web resources. In *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence*, pp. 240–246. IEEE Computer Society, 2006.
- [3] U. Masao and H. Kōiti. Automatic slide presentation from semantically annotated documents. In *Proceedings of the Workshop on Coreference and its Applications*, pp. 25–30. Association for Computational Linguistics, 1999.
- [4] Y. Yoshiaki, T. Masashi and N. Katsumi. A support system for making presentation slides. *Trans. Japanese Soc. Artif. Intell.*, Vol. 18, pp. 212–220, 2003.
- [5] M. Sravanthi, CR. Chowdary and PS. Kumar. Slidesgen: Automatic generation of presentation slides for a technical paper using summarization. In *FLAIRS Conference*, 2009.
- [6] S. Tomohide and K. Sadao. Automatic slide generation based on discourse structure analysis. In *International Conference on Natural Language Processing*, pp. 754–766. Springer, 2005.
- [7] PK. Gokul, H. Mathivanan, M. Jayaprakasam and TV. Geetha. Document summarization and information extraction for generation of presentation slides. *Advances in Recent Technologies in Communication and Computing*, 2009.
- [8] Y. Hu and X. Wan. Ppsgen: Learning-based presentation slides generation for academic papers. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 27, No. 4, pp. 1085–1097, 2015.
- [9] S. Wang, X. Wan and S. Du. Phrase-based presentation slides generation for academic papers. In *AAAI*, pp. 196–202, 2017.
- [10] S. Yuki, A. Atsushi, K. Fuminori and M. Akira. A method for estimating meanings for groups of shapes in presentation slides. *International Journal of Computer Theory and Engineering*, Vol. 8, No. 1, p. 74, 2016.
- [11] office open xml. <http://officeopenxml.com>.
- [12] B. Saleh, M. Dontcheva, A. Hertzmann and Z. Liu. Learning style similarity for searching infographics.

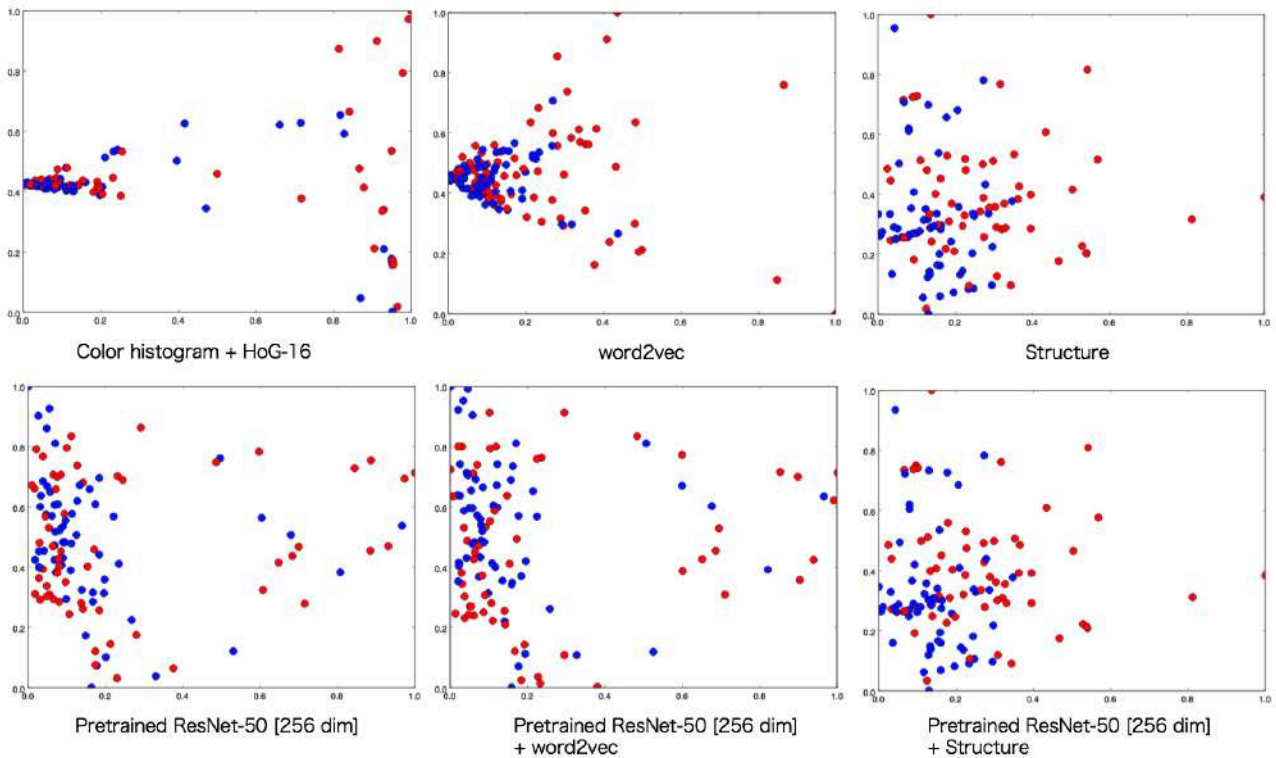


図 5: PCA によるスライド特徴空間の可視化

- In *Proceedings of the 41st Graphics Interface Conference*, pp. 59–64. Canadian Information Processing Society, 2015.
- [13] B. Kulis. Metric learning: A survey. *Foundations and Trends in Machine Learning*, Vol. 5, No. 4, pp. 287–364, 2012.
- [14] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Vol. 1, pp. 886–893. IEEE, 2005.
- [15] 福島悠介, 山崎俊彦, 相澤清晴. 文書と音声解析に基づくプレゼンテーション動画の印象予測. *電子情報通信学会論文誌 D*, Vol. 99, No. 8, pp. 699–708, 2016.
- [16] G. Luzardo, B. Guamán, K. Chiluiza, J. Castells and X. Ochoa. Estimation of presentations skills based on slides and audio features. In *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge*, pp. 37–44. ACM, 2014.
- [17] 栗原一貴, 加藤公一, 大浦弘樹. Slidechecker: プレゼンテーション資料の基礎的な定量的自動評価手法. *WISS 第17回インタラクティブシステムとソフトウェアに関するワークショップ論文集*, pp. 89–94, 2009.
- [18] crowdworks. <https://crowdworks.jp>.
- [19] X. Chen, PN. Bennett, K. Collins-Thompson and E. Horvitz. Pairwise ranking aggregation in a crowd-sourced setting. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pp. 193–202. ACM, 2013.
- [20] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, Vol. 86, No. 11, pp. 2278–2324, 1998.
- [21] J. Deng, W. Dong, R. Socher, LJ. Li, K. Li and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 248–255. IEEE, 2009.
- [22] Imagenet large scale visual recognition challenge (ilsvrc). <http://www.image-net.org/challenges/LSVRC/>.
- [23] K. He, X. Zhang, S. Ren and Sun J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.
- [24] Deep residual networks. <https://github.com/KaimingHe/deep-residual-networks>.
- [25] T. Mikolov, I. Sutskever, K. Chen, GS. Corrado and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pp. 3111–3119, 2013.
- [26] CC. Chang and CJ. Lin. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, Vol. 2, No. 3, p. 27, 2011.


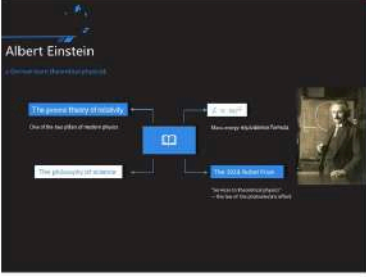



(A) Top-3	<h3>Albert Einstein</h3>  <p>Born : 14 March 1879, German Died : 18 April 1955 Field : Theoretical Physics Known for : The theory of relativity Prize : 1921 Nobel Prize in Physics</p> <p>Einstein Formula $E = mc^2$</p>	<h3>Albert Einstein</h3> <ul style="list-style-type: none"> Theoretical physicist Achievements: Mass-energy equivalence formula $E = mc^2$ Won the Nobel Prize in 1921 – The law of the photoelectric effect 	<h3>Albert Einstein</h3> <p>Profile</p> <ul style="list-style-type: none"> German-born theoretical physicist Known for: $E = mc^2$ <p>Achievements</p> <ul style="list-style-type: none"> Discovered general theory of relativity – One of the two pillars of modern physics Discovered law of photoelectric effect – Received 1921 Nobel Prize in Physics 																																														
(A) Worst-3		<h3>Albert Einstein</h3> <ul style="list-style-type: none"> He was a German-born theoretical physicist. He developed the general theory of relativity. His mass-energy equivalence formula $E = mc^2$ (which has been dubbed "the world's most famous equation"). He received the 1921 Nobel Prize in Physics for his discovery of the law of the photoelectric effect. 	<h3>Introduction</h3> <p>Albert Einstein was a German-born theoretical physicist. He developed the general theory of relativity, one of the two pillars of modern physics (alongside quantum mechanics).</p> <p>Einstein's work is also known for its influence on the philosophy of science. Einstein is best known to popular culture for his mass-energy equivalence formula $E = mc^2$ (which has been dubbed "the world's most famous equation").</p> <p>He received the 1921 Nobel Prize in Physics for his "services to theoretical physics", in particular his discovery of the law of the photoelectric effect, a general step in the evolution of quantum theory.</p>																																														
(C) Top-3	<h3>Eyeglasses & Contact Lenses</h3> <table border="1"> <tr> <td>heavy</td> <td>light</td> </tr> <tr> <td>easily fog up</td> <td>always clear</td> </tr> <tr> <td>cheap</td> <td>expensive</td> </tr> <tr> <td>little maintenance</td> <td>much maintenance</td> </tr> <tr> <td>safe for infections</td> <td>risky for infections</td> </tr> </table>	heavy	light	easily fog up	always clear	cheap	expensive	little maintenance	much maintenance	safe for infections	risky for infections	<h3>Eyeglasses & Contact Lenses</h3> <table border="1"> <thead> <tr> <th></th> <th>Eyeglasses</th> <th>Contact Lenses</th> </tr> </thead> <tbody> <tr> <td>Weight</td> <td>Heavy</td> <td>Light</td> </tr> <tr> <td>Fog up</td> <td>Fog up</td> <td>Don't Fog up</td> </tr> <tr> <td>Price</td> <td>Inexpensive</td> <td>Expensive</td> </tr> <tr> <td>Maintenance</td> <td>Easy to maintain</td> <td>Should buy a lot</td> </tr> <tr> <td>Risk</td> <td>Low risk for eye pain</td> <td>High risk for eye pain</td> </tr> </tbody> </table>		Eyeglasses	Contact Lenses	Weight	Heavy	Light	Fog up	Fog up	Don't Fog up	Price	Inexpensive	Expensive	Maintenance	Easy to maintain	Should buy a lot	Risk	Low risk for eye pain	High risk for eye pain	<h3>Eyeglasses & Contact Lenses</h3> <table border="1"> <thead> <tr> <th></th> <th>Eyeglasses</th> <th>Contact Lenses</th> </tr> </thead> <tbody> <tr> <td>Weight</td> <td>Weight</td> <td>No weight</td> </tr> <tr> <td>Get Fog</td> <td>Yes</td> <td>No</td> </tr> <tr> <td>Price</td> <td>Cheap</td> <td>Expensive</td> </tr> <tr> <td>Maintenance</td> <td>Little</td> <td>Much</td> </tr> <tr> <td>Risk</td> <td>Less</td> <td>More</td> </tr> </tbody> </table>		Eyeglasses	Contact Lenses	Weight	Weight	No weight	Get Fog	Yes	No	Price	Cheap	Expensive	Maintenance	Little	Much	Risk	Less	More
heavy	light																																																
easily fog up	always clear																																																
cheap	expensive																																																
little maintenance	much maintenance																																																
safe for infections	risky for infections																																																
	Eyeglasses	Contact Lenses																																															
Weight	Heavy	Light																																															
Fog up	Fog up	Don't Fog up																																															
Price	Inexpensive	Expensive																																															
Maintenance	Easy to maintain	Should buy a lot																																															
Risk	Low risk for eye pain	High risk for eye pain																																															
	Eyeglasses	Contact Lenses																																															
Weight	Weight	No weight																																															
Get Fog	Yes	No																																															
Price	Cheap	Expensive																																															
Maintenance	Little	Much																																															
Risk	Less	More																																															
(C) Worst-3	<h3>Eyeglasses & Contact Lenses</h3> <p>When you wear eyeglasses, you feel uncomfortable weight on your face and ears, while you cannot see fog up and reading discomfort. Also, eyeglasses fog up with change in temperature, while contacts don't.</p> <p>On the other hand, eyeglasses also offer many benefits over contacts. For example, eyeglasses are cheaper than contacts in the long run and require very little maintenance. Also, you don't need to touch your eyes to wear eyeglasses, so you have less risk of eye infections than wearing contacts.</p>	<h3>Eyeglasses & Contact Lenses</h3> <p>eyesight deteriorates! → wearing eyeglasses or contact lenses</p> <p>MERIT & BENEFIT OF EYEGGLASSES</p> <ul style="list-style-type: none"> *you feel uncomfortable weight on your face and ears *fog up with changes in temperature *cheaper *less risks for eye infections 	<h3>Eyeglasses and Contact Lenses</h3> <p>Eyeglasses and contact lenses are deteriorated eyesight aid tools that have advantages and disadvantages with each.</p>																																														
(D) Top-3	<h3>Rock-paper-scissors</h3> 	<h3>Rock-Paper-Scissors</h3> 	<h3>Rock-paper-scissors</h3> <p>• Zero-sum hand game</p> 																																														
(D) Worst-3	<h3>Rock-paper-scissors</h3> <p>Rock-paper-scissors or Scissors-Paper-Rock, is a non-zero-sum game usually played between two people, in which each player simultaneously forms one of three shapes with an outstretched hand.</p> <p>These shapes are "rock" (a simple fist), "paper" (a flat hand), and "scissors" (a flat with the index and middle fingers together forming a V).</p> <p>The game has only three possible outcomes other than a tie: if a player who decides to play rock will beat another player who has chosen scissors ("rock crushes scissors"), but will lose to one who has played paper ("paper covers rock"); a player of paper will lose to a player of scissors ("scissors cut paper").</p> <p>If both players choose the same shape, the game is tied and is usually immediately replayed to break the tie.</p> <p>In the other version for the game in the English-speaking world, include rock-paper and other meanings of the three shapes, associated with "rock" being called "stone".</p>	<h3>Rock-paper-scissors</h3> <p>"rock" (a simple fist) "paper" (a flat hand) "scissors" (a flat with the index and middle fingers together forming a V)</p> <p>The game has only three possible outcomes other than a tie: if both players choose the same shape, the game is tied and is usually immediately replayed to break the tie.</p>	<h3>Rock-paper-scissors</h3> <ul style="list-style-type: none"> Each player simultaneously forms one of three shapes which are "rock" (a simple fist), "paper" (a flat hand), and "scissors" (the index and middle fingers forming a V). Rock will beat Scissors (it crushes scissors) but lose to Paper (it covers rock). Paper lose to Scissors (it cuts paper). Both players choose the same shape, the game is tied and replayed to break the tie. In English-speaking countries it is called Roshambo. 																																														

図 6: 高評価スライド・低評価スライドの例