

統計ポスターの視覚的な要素に対するCNNを用いた 自動評価システム「Vesper」の開発

奈田 青葉[†] 渡辺 博芳[†] 林 宏樹[‡]

[†]電気通信大学 [‡]雲雀丘学園中学校・高等学校

1 はじめに

高等学校の「情報 I」では、「データの活用」と「情報デザイン」を連携させた学習活動の成果物として、統計ポスターが制作されることがある [1]。統計ポスターを「情報デザイン」の観点から評価する際には、「受け手を意識して内容を構成する力」と、「伝える情報を明確に伝える力」の2つが重要視される。「伝える情報を明確に伝える力」では、レイアウトや配色といった視覚的な要素が重要な役割を果たすと思われる。こうした視覚的な要素については、伝える内容には踏み込まずに成果物の画像を解析することで、自動評価ができるのではないかと考えた。

これまでに我々は、統計ポスターの視覚的な要素に対して評価基準を設定し、ポスターがそれらの基準を満たしているかどうかを、CNN（畳み込みニューラルネットワーク）を用いて自動評価するモデルを構築してきた [2]。

本研究では、これらの研究成果を基盤として、生徒がアップロードしたポスターの視覚的な要素が、評価基準を満たしているかどうかを自動評価するシステム「Vesper」を開発することを目的とする。本システムを高等学校のポスター作成現場に導入することで、生徒が判定結果を参考にしながらポスターの改善に取り組むことが期待される。さらに、システムから提示される評価結果や各項目の説明に触れることで、生徒がデザインの原則や配色のルールに関心を持つ契機となり、主体的な学びにつながる可能性がある。

2 データセットの作成

本研究では、従来の研究 [2] と同様に、「近接」「整列」「反復」「対比」「余白」「背景色」「色数」「色の組合せ」の8つを評価項目として採用し、評価基準を設定した。

Vesper: An Automatic Evaluation System of Visual Elements in Statistical Posters Using CNNs

Aoba Nada[†], Hiroyoshi Watanabe[†] and Hiroki Hayashi[‡]

[†]The University of Electro-Communications

[‡]Hibarigaoka Gakuen Junior & High School

アノテーションは高校生が授業で作成した統計ポスター 641 枚を対象とし、第一著者に加え、研究室所属の大学院生・学部生を含む計 7 名で実施した。各ポスターは 2 名が独立して評価し、各評価者は割り当てられた 183 枚のポスターに対して、8 項目それぞれに対して 3 値（2：十分，1：部分的，0：不十分）のラベルを付与した。3 値のラベルを採用した理由は、基準を満たしているか曖昧なポスターを一旦「部分的」として分離し、その後十分/不十分の 2 値へ統合することで、2 値のみでアノテーションする場合よりも主観性に起因するぶれ（評価者の解釈の違いによって評価が分かれる現象）を抑制できると考えたためである。

全評価者のアノテーションが完了した後、2 名の評価結果を照合した。評価が一致したものはそのまま採用し、一致しなかったものについては別の評価者が再評価を行った。

最後に、部分的なクラスである「1」を十分/不十分の二値に統合した。統合の方針（「1」を「0」と「2」のどちらに統合するか）を決定するため、予備実験としてケーススタディを行った。実験では、従来の研究と同様のモデル構築手法を用いて、学習データの「1」を「0」に統合して学習させたモデルと、「2」に統合して学習させたモデルの 2 種類を構築した。これらのモデルで、学習データに含まれない 10 枚の統計ポスターをそれぞれ評価し、結果の違いを比較した。その結果、「1」を「2」に統合すると、曖昧な事例まで「十分」と判定され、評価が過度に甘くなる傾向が見られた。一方、「1」を「0」に統合した場合は、評価者の感覚により近い判定となることが確認された。したがって、本研究では「1」を「0」に統合し、2 クラスのデータセットを作成した。

3 自動評価モデルの構築

自動評価モデルの構築には Python3 と、その代表的な機械学習ライブラリである TensorFlow を活用した。本システムに搭載する最適なモデルを探索するため、

k-分割交差検証 (k = 10) を行った. 全 641 枚のデータセットを 10 分割し, 各 fold において約 577 枚の学習データを用いてモデルを学習し, 残りをテストデータとして性能評価を行った. 交差検証により得られた 10 個のモデルのうち, テストデータに対する AUC が最も高かったモデルを本システムに採用した. AUC は, 正解率よりもテストデータにおけるクラス分布の影響を受けにくく, 陽性・陰性双方の識別性能を総合的に評価できる指標である.

各評価項目で使用する CNN モデルについて, 従来の研究結果 [2] より, 評価項目によって適切な CNN モデルが異なる可能性が示唆されている. そこで本研究では, 「近接」「整列」「反復」「対比」「余白」では MobileNetV2 を, 「背景色」「色数」「色の組合せ」では DenseNet121 を採用した.

モデルの構築手順について, はじめにポスター画像の長辺を短辺に合わせて正方形に変形し, CNN の入力画像サイズである 224 × 224 px にリサイズした. 次に, 2 クラス間でデータ数に大きな偏りが見られる評価項目については, 学習データに対して画像の変換・加工によるオーバーサンプリングを行った. その後, 各 CNN には ImageNet で学習済みの重みを初期値として, 転移学習を行った.

4 システムの開発

本研究では, 前章で構築した評価モデルを組み込んだ統計ポスター自動評価システム「Vesper」を, Docker 上で動作する Web アプリケーションとして実装した. フロントエンドには HTML, CSS, JavaScript を, バックエンドには Python の FastAPI を採用し, Nginx をリバースプロキシとして配置した.

本システムでは, 生徒が画像ファイル (JPG/PNG) または PPTX ファイルをアップロードすると, サーバ側で推論用に画像へ変換し, 8 つの評価項目について自動評価を行う.

結果画面の例を図 1 に示す. 結果画面では, モデルの出力確率を信頼度として提示し, 閾値 0.5 を用いて「基準を満たす (○)」または「改善の余地あり (△)」の二値で評価する. さらに, 結果画面の下部には各項目の説明文を表示し, 生徒が評価項目の意味を理解しながら改善に取り組めるよう配慮した.

実践利用を想定し, 評価日時, 出席番号, 判定結果, 信頼度, 入力画像を自動保存するログ機能も実装した. これにより, 学習状況の把握や将来的なモデル改良に活用可能な基盤を整えている.



図 1: Vesper の結果画面

5 おわりに

本研究では, 統計ポスターを自動評価するシステム「Vesper」を開発し, Web アプリケーションとして実装した. 今後は, 本システムを高等学校の統計ポスター作成の場に導入し, 生徒による使用感や運用上の課題を確認する予定である. また, 実用性をさらに高めるため, 評価結果に基づいて改善点を提示するフィードバック機能の拡充についても検討を進めていく.

なお, 本研究は電気通信大学 人を対象とする研究に関する倫理委員会の承認を得た (承認番号: H24001). また, 本研究の一部は JSPS 科研費 JP23K02665 の助成を受けた.

参考文献

- [1] 林 宏樹, 渡辺 博芳. 高等学校「情報 I」における統計ポスターを題材として「データの活用」と「情報デザイン」を連携する授業実践. 情報処理学会論文誌, 教育とコンピュータ, Vol.10, No.3, pp. 26–36 (2024).
- [2] 奈田 青葉, 渡辺 博芳, 林 宏樹. 統計ポスターの視覚的な要素に対する CNN を用いた自動評価 — 評価の主観性に起因するぶれへの対応 —. 情報処理学会研究報告, Vol.CE-182, No.16, pp. 1–10 (2025).