

語の共起の統計情報を用いた画像と説明文の対応付け

Linking Images with Their Expository Text Based on Statistical Information of Word Co-occurrence

竹内謹治¹
Kinji Takeuchi

黄瀬浩一²
Koichi Kise

松本啓之亮²
Keinosuke Matsumoto

大阪府立大学 工学部 情報工学科¹

Department of Computer and Systems Sciences, College of Engineering, Osaka Prefecture University

大阪府立大学大学院 工学研究科 情報工学分野²

Department of Computer and Systems Sciences, Graduate School of Engineering, Osaka Prefecture University

1 はじめに

現在, Web 上には画像があふれている. これらをキーワードから検索するためには, 画像を予め, テキストによって索引付けする必要がある. 従来法では alt テキストなどを利用して索引付けを行っている. しかし, この方法ではキーワードの語数が非常に限られているため, 多様な検索要求に対処できないという問題点が指摘されている.

ところで, HTML 文書に画像が存在するとき, 単独で存在するのではなく, 本文中で説明を受けていることが多い. したがって, 本文中の説明文を画像に対応付けることができれば, 索引が充実し, 多様な要求にも対処できると考えられる. また, 説明文を読むことで画像の内容をより詳しく知ることができるという利点もある.

そこで本研究では, alt テキストを手がかりにして HTML 文書から画像の説明文を抽出する手法を提案する. 提案手法は, 統計的な指標を用いて alt テキストと文中の語の共起関係を求めることで, 画像に深く関連する文を抽出できるという特徴がある.

2 画像の説明文の抽出手法

2つの語の共起関係を用いてキーワードを抽出する方法として, χ^2 値を利用する手法 [1] がある. 提案手法では [1] の手法を応用し, 文中から画像の説明文を抽出する. ただし, 処理の対象は英文とする. 具体的な手順は以下の通りである.

- (1) 準備: HTML を解析し, 本文の抽出および, 画像の alt テキストを抽出する. そして抽出した本文に stemming を行う.
- (2) 理論確率の計算: alt テキスト中の語 $g \in G$ と本文中で共起する延べ総数 n_g と本文の語の総数 n_{total} を調べ理論確率 $p_g = n_g/n_{total}$ を求める.
- (3) χ^2 値の計算: 本文中のすべての語 w について, w と $g \in G$ の共起頻度 $freq(w, g)$ と w が出現する文の総数 n_w を求める. 次に χ^2 値を以下の式で求める.

$$\chi^2(w) = \sum_{g \in G} \frac{(freq(w, g) - n_w p_g)^2}{n_w p_g} - \max_{g \in G} \frac{(freq(w, g) - n_w p_g)^2}{n_w p_g}$$

ただし, 前置詞などの不要語 χ^2 値を 0 にする.



altテキスト:greenwich ball
説明文:This is the old royal greenwich observatory in greenwich, england, home of greenwich mean time.

Erected in 1833 by famous british architect christopher wren, the greenwich time ball in flamsteed house was one of the world's first visual time signals.

(a) 成功例



altテキスト:tower of london
説明文:Turn off the strand onto the short forecourt of the savoy hotel (known to london cabbies as the banjo, because of its shape) and you are on the only street in london where cars drive on the right side of the road.

(b) 失敗例

図 1 処理例

- (4) 説明文の抽出: 各文に含まれている単語の χ^2 値の和をとり, これを文のスコアとする. このスコアの順に文をソートし最大値の 50% 以上のすべての文を説明文として抽出する.

3 実験とまとめ

2で提案した手法の性能を評価するため画像の説明文の抽出実験を行った. 実験対象は Google のテキスト検索で「greenwich london sightseeing」を検索質問とした際の上位 40 件の Web ページである. 実験の評価には再現率 $R = |C|/|A|$, 精度 $P = |C|/|B|$, F 値 $= \frac{2RP}{R+P}$ を用いた. ここで, $|A|$ は Web ページ内で画像を説明している文の数, $|B|$ は結果として得られた文の数, $|C|$ は結果として得られた正解文の数である. この実験では再現率, 精度, F 値の平均がそれぞれ, 0.52, 0.65, 0.53 という結果を得た. 実験の成功例と失敗例を図 1 に示す. 図 1(b) のように alt テキストに “london” などの頻出語が入っている場合, 誤抽出が多かった. また, 図 1(a) のように誤抽出がない場合でも, キーワードが代名詞で表されている文は未抽出だった. そこで, 更なる F 値の向上を目指して, 今後は自然言語処理など他の手法と組み合わせて利用することを考えたい.

参考文献

- [1] 松尾豊, 石塚満: “語の共起の統計情報に基づく文書からのキーワード抽出アルゴリズム”, 人工知能学会論文誌, Vol.17, No.3, pp.217-223(2002).