

# 検索連動広告における広告選択手法の提案

龍田 賢治†

上原子 正利†

小柳 滋†

立命館大学情報理工学部†

## 1 はじめに

検索連動広告は検索エンジンを運営する企業にとって重要な収入源であり、広告主にとっての費用対効果も高い。現在 Yahoo!, Google, Microsoft などの提供する検索連動広告では、広告主が広告に関係の深いキーワードを登録 (入札) し、ユーザの入力したクエリにキーワードが現れたとき、広告が表示される。このため、キーワードによるユーザの目的に合わせたターゲティングが可能となっている。また、キーワードベースによるターゲティングは、表示回数やクリック率が調整しやすく、広告主が管理しやすいという利点も持つ。

しかし、あらゆるクエリに有効なわけではない。一般的な語や、単語が多く組み合わせられたクエリ (レアクエリ) では、有効な広告があってもほとんど表示されない場合がある。この問題を解決するため、本稿ではクエリ拡張を使った広告選択手法を提案する。

## 2 関連研究

Broder ら [1] の研究では、キーワードによらず検索結果に現れる単語、フレーズ、カテゴリによって広告を検索する。この手法は有効だが、広告の文章によって広告の表示回数が大きく変化し、広告の効果や費用が予測、管理しづらい欠点がある。

Zhang ら [2] の研究では、クエリ拡張を広告の拡張に応用している。クエリ拡張とは、ユーザから受け取ったクエリと関連の深い別の単語、フレーズの検索結果も元のクエリの検索結果に含めて拡張する手法である。これを広告に応用し、関連の深いクエリ、入札キーワードに対応する広告を追加することで、元の広告表示を拡張している。

クエリ拡張などのクエリ操作は、一般に「クエリリライトイング」と呼ばれ、多くの研究がなされている。また、Malekian ら [3], Radlinski ら [4] により、クエリ拡張における広告収入の最適化が研究されている。しかし、これらのクエリ拡張を用いた手法では、キーワードベースの利点を残すことができるが、複雑なクエリに対応することが難しい。

## 3 提案手法

我々は、Broder らのウェブ検索結果中の文章を広告検索にフィードバックする手法をクエリ拡張と組み合

わせ、より効果的な広告拡張を行う手法を提案する。Zhang らのクエリ拡張は、類似クエリの発見のためクエリの文章類似度、統計的特徴、共起性、検索結果に同じ URL を含む割合 (以下 URL 共通性)、共通の広告入札者を持つ割合に注目し、能動学習により精度を高めている。これに対し、我々の手法ではクエリに含まれる単語、Web 検索結果の頻出単語から拡張クエリ候補を取得し、URL 共通性と Web 検索結果に現れる文書の cosine 類似度を比較することで、より有効な類似クエリを発見する。

以下に手法の詳細を説明する。ユーザからのクエリを受け取ったとき、まず拡張の候補となるクエリを収集する。このとき、元のクエリと部分的に一致するクエリと元のクエリの検索結果から取り出した頻出単語を含むクエリを候補とする。本来ならばクエリログから候補を収集するべきだが、ここでは Yahoo! 関連語サーチから取得したクエリで代用する。次に、収集した候補クエリと元のクエリとの類似度を 2 つの尺度で計算する。1 つは、Web 検索結果中の文書の cosine 類似度であり、文書を形態素解析し、単語の出現回数をベクトルとして次の式で比較する。

$$COSsim(q, q') = \frac{\sum_i^k q w_i \cdot q' w_i}{\sqrt{\sum_i^k q w_i^2} \cdot \sqrt{\sum_i^k q' w_i^2}} \quad (1)$$

ただし、 $q$  は元のクエリ、 $q'$  は拡張候補クエリ、 $q w_i$  はクエリ  $q$  の検索結果に含まれる単語  $i$  の出現数を表す。もう一方は URL 共通性であり、次式で表される。

$$URLsim(q, q') = \frac{\sum_i^k q u_i \cdot q' u_i}{k} \quad (2)$$

ただし  $q u_i$ ,  $q' u_i$  は URL がクエリに含まれれば 1, 含まれなければ 0 とする。つまりこの式は (クエリ  $q$  と  $q'$  で共通する URL 数) ÷ (検索結果に含まれる URL 数) であり、クエリ間での URL が共通している割合を表す。最後に 2 つの尺度にウェイトをつけて合計する。 $\alpha$ ,  $\beta$  がそれぞれのウェイトである。

$$sim(q, q') = \alpha COSsim(q, q') + \beta URLsim(q, q') \quad (3)$$

拡張クエリが決定されれば、本題の広告拡張を行う。類似度  $sim(q, q')$  が上位の候補クエリで、かつ閾値以上の値を持つクエリを拡張クエリとして採用し、拡張クエリで表示される広告も拡張広告として表示する。

## 4 実験と評価

### 4.1 動作例

以下に、提案手法による広告拡張の例を示す。ユーザが "java 継承" というクエリを検索エンジンに入力し

A new retrieval method for ad search.

†Kenji TATSUTA,

Masatoshi KAMIHARAKO, Sigeru OYANAGI.

†College of information science and engineering,

Ritsumeikan University, 2010.

表 1 "java 継承"に対する広告拡張の例

	クエリ	スコア	クエリに属する広告
通常広告	"java 継承"	-	個別指導! Java 講座...(PC 講座) Java 講座とサポート...(PC 通信講座)
拡張広告	"Java 継承 性質"	0.83	なし
	"Java 継承 キャスト"	0.80	なし
	"クラス 継承"	0.79	クラスと継承 なら...(書籍)
	"Java 継承 プログラム"	0.78	JAVA 講座を一括資料請求...(PC 講座) JavaEclipse 就職学校...(PC 講座) エンジニアスクール...(情報系専門学校) Java のプログラムは...(情報系専門学校) Java 習得なら...(学習教材) ソリューション会社...(SI 企業)

たときを想定する。まず、クエリとその Web 検索結果から特徴語を抽出する。クエリからは"java", "継承"しか抽出できないが、Web 検索結果からは"クラス", "多重", "キャスト"などの特徴語を抽出することができる。次に、特徴語から拡張クエリ候補を取得する。Yahoo! 関連語サーチにより, "クラス 継承", "Java キャスト", "java 入門" など, このクエリでは 95 個の候補を取得できた。候補が取得できれば, それらと元のクエリの類似度を調べ, より近いものから拡張クエリとする。

第 3 章で説明したように, 類似度計算には出現単語のコサイン類似度と URL 共通性を用いる。コサイン類似度の重みを大きくしたときに結果が良かったため, それぞれのウェイトを  $\alpha = 3, \beta = 1$  とした。また, 拡張の閾値として  $sim() > 1.6$  とした。類似度計算の結果, "Java 継承 性質", "Java 継承 キャスト", "クラス 継承", "Java 継承 プログラム" を上位のクエリとして決定した。表 1 に, これらの拡張クエリによって拡張された広告を示す。表 1 からわかるように, プログラミングの学習に関する広告だけでなく, クラスと継承に関する書籍など有効な広告が多く発見できた。

#### 4.2 比較実験

本手法の有効性を, 既存手法との比較実験で評価する。提案手法との比較対象として, 現行の Yahoo! スポンサーサーチ (2010 年 1 月現在), Broder ら, Zhang らの手法を再現したものを同様に評価する。ただし, Broder らの手法に使われているカテゴリが使えないため, Yahoo! ディレクトリで代用するなど実装上の相違がある。

比較法を以下に示す。比較には Broder らの評価法に基づき, 1 つの広告に対してクエリとのマッチを 6 段階の評価 (完全にマッチ (+3), マッチ (+2), ほぼマッチ (+1), 少しマッチ (0), ほぼミスマッチ (-1), ミスマッチ (-2)) で行った。今回, マッチとミスマッチの割合が分かりやすいように, それぞれ分けて集計した。

結果は図 1 のグラフのようになった。グラフ中の Onyx とは Broder らの手法名である。比較結果から分かるように, ミスマッチスコアが Broder らの手法よりやや多いものの, 多くの有効な広告を表示できていることがわかる。

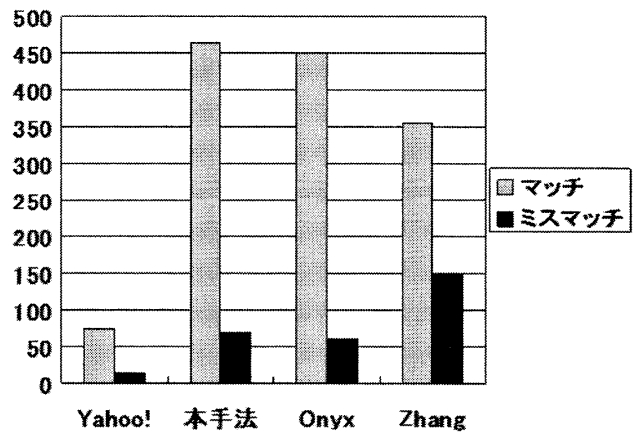


図 1 比較結果

#### 5 おわりに

本稿では, 検索連動広告における広告拡張の手法を提案し, キーワードベースの利点を残しつつ有効な広告をより多く収集できることを実験と評価によって示せた。

#### 参考文献

- [1] A. Z. Broder, P. Ciccolo, M. Fontoura, E. Gabrilovich, V. Josifovski, L. Riedel. Search advertising using web relevance feedback. In CIKM 2008:Proceeding of the 17th ACM conference on Information and knowledge management. pp.1013-1022,2008.
- [2] W. V. Zhang, X. H. Fei, B. Rey, R. Jones. Query rewriting using active learning for sponsored search. In SIGIR '07:Proceedings of 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. pp.853-854,2007.
- [3] A. Malekian, CC.Chang, R. Kumar, G. Wang. Optimizing query rewrites for keyword-based advertising. In ACM 2008:Proceedings of the 9th ACM conference on Electronic commerce.pp.10-19,2008
- [4] F. Radlinski, A. Broder, P. Ciccolo, E. Gabrilovich, V. Josifovski, L. Riedel. Optimizing relevance and revenue in ad search: a query substitution approach. In SIGIR '08:Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval. pp.403-410,2008.