

クリック文書の分散表現を用いたクエリ曖昧性の評価

矢野 友貴[†] 田頭 幸浩[†] 田島 玲[†]

[†] ヤフー株式会社 〒 107-6211 東京都港区赤坂 9-7-1 ミッドタウン・タワー

E-mail: †{yuyano,yutagami,atajima}@yahoo-corp.jp

あらまし クエリ曖昧性の評価はユーザの検索意図を理解する上で重要な要因の一つである。従来の手法では、クエリ曖昧性はクリック分布のばらつき具合によって定量化されるが、これらの手法ではクリックされた文書を独立として扱うため、文書は異なるがトピックが近いようなケースを正しく判断することができない。本項ではこの問題に対し、クリックされた文書の分散表現を活用することで、文書間のトピック差異を加味したクエリ曖昧性の評価手法を提案する。従来手法と比較した結果、従来手法では分類できなかった曖昧性の差異を検知することに成功し、また提案手法がクエリ毎のクリック率 (CTR) とよりよい相関があることが確認された。

キーワード 情報検索, 検索意図, クエリ曖昧性, e コマース

1. はじめに

検索エンジンにおいて、ユーザの検索意図の理解はよりよい検索結果を実現するために非常に重要である。検索意図を理解する上で重要となる指標にクエリ曖昧性がある。クエリ曖昧性は検索意図のばらつき具合を数値化したもので、例えば関連性と多様性のトレードオフの調整 [1] や、パーソナライズを行うかどうかの判定 [2], ランキングモデルへの活用 [3] など、検索精度向上において重要な要素の一つとなっている。クエリ曖昧性は多くの場合、クリック文書の分布のようなユーザの過去のクリック行動を元に数値化される [2], [4], [5]。

クエリ曖昧性の評価手法としては *click entropy* が広く知られている [2]。 *click entropy* では意図のばらつき具合をクリック分布のエントロピーとして表現することで、クエリ曖昧性の定量化を行っている。また、クエリの曖昧性は大きく分けて *ambiguous*, *broad*, *clear* の 3 種類に分類され [6], 図 1 のようにクリックされた文書群のトピック差異によって表現される。

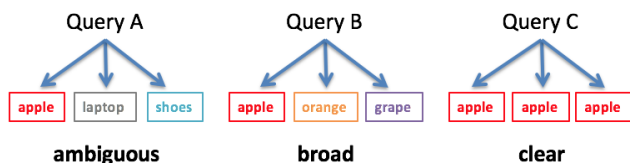


図 1 クエリ曖昧性の 3 つの分類

図 1 の 3 つを比較したとき、Query A は 3 つが完全に別トピック、Query B は果物という点は共通だが詳細が異なる、Query C は完全に同一のクリック、という違いがあるため、直感的なクエリ曖昧性は $A > B > C$ の順番となる。一方、図 1 の曖昧性評価に *click entropy* を用いた場合、Query A と Query B は同一の曖昧性と評価される。 *click entropy* ではクリックされた文書を独立として扱うため、図 1 のようなケースでは *ambiguous* と *board* をうまく分類できないという問題がある。

この問題に対し、本項では文書間の類似性を加味したクエリ

曖昧性の定量化を提案する。具体的には文書の分散表現を利用し、各文書ベクトルの向きの集約度を計算することでクエリの曖昧性を定量化する。例えば図 1 の場合、提案手法では文書の分散表現に着目することで、Query A よりも Query B の方が各文書ベクトルがより密に集まっていると判断し、2 つのクエリの曖昧性を適切に評価することが可能となる。

実験では Yahoo! ショッピングの実データを用い、各商品カテゴリの分散表現を用いて手法の評価を行った。 *click entropy* と比較した結果、 *click entropy* では適切に評価できない似通った商品群をクリックするクエリについても、提案手法ではその曖昧性の低さを認識することに成功した。また、提案手法は *click entropy* に比べ、クエリ毎のクリック率 (CTR) とより強い相関があることが確認された。

本項では以降、2 章にて提案手法の詳細を述べたのち、3 章にて Yahoo! ショッピングのデータを用いた実験と結果を、4 章にて関連研究について述べ、最後に 5 章にてまとめと今後の課題を述べる。

2. 提案手法

click entropy 及びその派生手法 [4], [5] では、各文書を独立と仮定し、それに対するクリックのばらつき具合によってクエリ曖昧性を定量化している。そのため、クリックされた文書群のトピックの近さが異なる場合でも、その大きさを曖昧性に反映することができない。提案手法では、文書群のトピックの近さをクエリ曖昧性の計算に取り込むことで、これらの問題の解決を図る。

2.1 問題設定

クエリを q 、文書を $d \in D$ とし、クエリ q において文書 d がクリックされた割合を $p(d|q)$ と定義する。本手法の目的は、クエリ q とそのクエリでクリックされた文書集合 $D_q \subset D$ が与えられたとき、そのクエリの曖昧性 $amb(q, D_q)$ を計算することである。

2.2 分散表現の導出

本手法では文書の分散表現を得るための手法に制約はなく、latent dirichlet allocation (LDA) [7] や Paragraph Vector [8]

といった任意の手法を適用することが可能である。本項では、文書自体をそのまま用いるのではなく、Yahoo!ショッピングの各商品に付与されたカテゴリ $c \in C$ を利用して文書を表現する。具体的には latent semantic indexing (LSI) [9] によってカテゴリの分散表現を導出し、それを商品の分散表現として用いた。Yahoo!ショッピングではカテゴリは木構造として保持されているが、LSI では全カテゴリを用いて分散表現を導出し、後述のクエリ曖昧性の計算には木構造の葉にあたる末端カテゴリのみを用いた。カテゴリ情報の利用は、文書数の上限をカテゴリ数に抑えることができるため、計算が容易になる利点がある。

LSI は一般的に単語-文書行列を用いて計算されるが、本項では各カテゴリが一つの巨大な文書であると仮定し、単語-カテゴリ行列を特異値分解することでカテゴリの分散表現を導出する。具体的には式 1, 2 のように tf-idf を定義し、行列の各要素を計算した。なお、 D_c はカテゴリ c に属する文書の集合を、 T_d は文書 d に含まれる単語集合^(注1) を、 $n_{t,d}$ は単語 t の文書 d での出現回数を表す。

$$tf(t, c) = \frac{\sum_{d \in D_c} n_{t,d}}{\sum_{k \in T_d} n_{k,d}} \quad (1)$$

$$idf(t) = \log \left(\frac{|D|}{|\{d | t \in T_d\}|} \right) \quad (2)$$

最終的に得られたカテゴリの分散表現を式 (3) のように文書の分散表現とする。なお、 $\mathbf{v}(x)$ は要素 x の分散表現を、 c_d は文書 d のカテゴリを表す。

$$\mathbf{v}(d) = \mathbf{v}(c_d) \quad (3)$$

2.3 分散表現に基づくクエリ曖昧性

各文書の分散表現が得られたのち、それらベクトルの向きの集約度をみてクエリ曖昧性を評価する。図 2 に図 1 を文書ベクトルで表現し直したイメージを示す。文書の分散表現を用いることで、近いトピックを持つ文書のベクトルはその向きが近くなることが期待される。そのため、各文書ベクトルの向きを比較することで、図 2 のように 3 つの分類タイプを明確に識別することが可能となる。

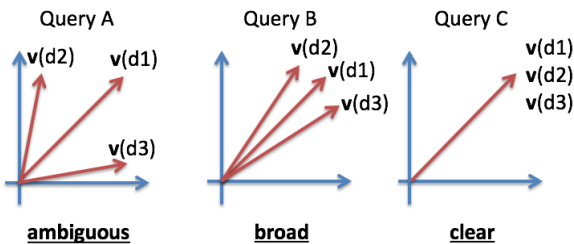


図 2 図 1 のクリックの文書ベクトル表現

本手法では、文書ベクトルの向きの集約度を計算するために、ベクトルの重心とのコサイン類似度に注目する。クエリ q でクリックされた文書の重心ベクトルを $\mathbf{g}(q, D_q)$ としたとき、重心は式 (4) のように計算される。

$$\mathbf{g}(q, D_q) = \sum_{d \in D_q} p(d|q) \frac{\mathbf{v}(d)}{|\mathbf{v}(d)|} \quad (4)$$

重心が求まったのち、クエリ曖昧性 $amb(q, D_q)$ は重心と各文書ベクトルのコサイン類似度の加重平均を用いて式 (5) のように計算する。

$$\begin{aligned} amb(q, D_q) &= 1 - \sum_{d \in D_q} p(d|q) \frac{\mathbf{g}(q, D_q) \cdot \mathbf{v}(d)}{|\mathbf{g}(q, D_q)| |\mathbf{v}(d)|} \\ &= 1 - \frac{\mathbf{g}(q, D_q)}{|\mathbf{g}(q, D_q)|} \sum_{d \in D_q} p(d|q) \frac{\mathbf{v}(d)}{|\mathbf{v}(d)|} \\ &= 1 - |\mathbf{g}(q, D_q)| \end{aligned} \quad (5)$$

なお、式 (4), (5) より $amb(q, D_q)$ は最小値が 0 であり、値が大きいほど曖昧なクエリであることを表す。

3. 実験と評価

3.1 データセット

評価には Yahoo!ショッピングの実データを用い、具体的には以下の 2 種類のデータを用いた。

- 商品データ：分散表現を導出するために利用
- 検索ログ：クエリ曖昧性の評価で利用

商品データは約 1.37 億の商品セットとなっており、26,116 カテゴリ (うち末端カテゴリは 22,990) と 3,447,459 語彙から構成される。計算上の都合から、語彙は頻度の高い上位 100,000 語彙 (全単語の出現数の 97.5%) に制限した。

検索ログは訓練用とテスト用に 2 つの期間のログを用意した。訓練用の検索ログは 2015/01/01 から 2015/03/31 までの期間で集計を行い、事前にクリック数が相対的に小さいクエリを除外し、最終的に 114,536 クエリを得た。テスト用の検索ログは 2015/04/01 から 2015/04/30 までの期間で集計を行い、訓練用データにて出現しなかったクエリを除外した。

各データセットの詳細は表 1 のようになっている。

表 1 検証で利用するデータセット

商品データ	
対象商品数	137,320,900
カテゴリ数	26,116
末端カテゴリ数 (leaf)	22,990
語彙数	3,447,459

検索ログ		
期間	01/01/2015 - 03/31/2015	04/01/2015 - 04/30/2015
対象リクエスト数	59,328,323	18,301,527
対象クエリ数	114,536	109,918

3.2 分散表現による類似性評価

実験では LSI の特異値分解として SciPy [10] を利用した。事前検証では分散表現の次元数を増やすことでよりよい結果が得られることが確認されたが、計算コストの都合から分散表現の次元数を 128 とした。

表 2 に得られた分散表現を用いて評価した類似カテゴリ top3

(注1)：本項では商品のタイトルを用いた。

表 2 各カテゴリとの類似カテゴリ top3

パソコン, 周辺機器 > デジタルカメラ
AV 機器, カメラ > カメラ > デジタルカメラ > その他
AV 機器, カメラ > カメラ > デジタル一眼レフ > その他
パソコン, 周辺機器 > デジタル一眼レフ
ファッション > レディース > バッグ > ハンドバッグ
ファッション > レディース > バッグ > トートバッグ > その他
ファッション > レディース > バッグ > ショルダーバッグ > その他
ファッション > レディース > バッグ > トートバッグ > 革
食品 > スイーツ, 洋菓子 > アイスクリーム > 詰め合わせ
食品 > スイーツ, 洋菓子 > ジェラート > 詰め合わせ
食品 > 和菓子 > くず餅
食品 > スイーツ, 洋菓子 > プリン > 詰め合わせ

の例を示す。2.2 節で述べたように、本項では商品の末端カテゴリで文書を表現しているため、末端カテゴリ間での比較を行っている。表 2 より、分散表現によって類似カテゴリがうまく取得できていることがわかり、特に「パソコン, 周辺機器 > デジタルカメラ」の例では AV 機器のカメラとの類似性を正しく認識できていることが確認できる。

3.3 クエリ曖昧性の評価

提案手法の有用性を評価するために、*click entropy* [2] との特性の差異について評価を行った。本実験では文書の末端カテゴリ情報を利用して分散表現を導出しているため、*click entropy* も式 (6) のようにクエリのカテゴリ分布を用いて計算した。

$$ent(q, D_q) = \sum_{c \in C_q} -p(c|q) \log(p(c|q)) \quad (6)$$

where $C_q = \bigcup_{d \in D_q} c_d$

以降、評価用の検索ログを用いて 2 つの手法でのクエリ曖昧性を導出し、以下の 3 つの観点からクエリ曖昧性の評価を行った。

- クエリ曖昧性の分類
- 時間変化に対する安定性
- 検索品質との相関

3.3.1 クエリ曖昧性の分類

図 3, 4 は *click entropy* (*ent*) と提案手法 (*amb*) について、訓練用クリックログの 114,536 クエリに対するクエリ曖昧性のヒストグラムである。2 つのヒストグラムを比較すると、ピーク的位置に差があることが分かる。*click entropy* に比べ提案手法ではピークがより値の小さいところに存在しているが、これは分散表現を用いて各文書の関連度を考慮したことで、全ての文書を独立に扱うよりも曖昧性を低く評価したクエリが多くなったためと考えられる。

実際のクエリ曖昧性の評価例を表 3 に示す。表 3 では各クエリについてクリック率の高かった top5 のカテゴリを記載している。初めの「テーブル」というクエリの例では、カテゴリが 265 種類と多岐にわたっており、また最もクリック率の大きいカテゴリも 0.094 と低い数値となっているため、*click entropy* ではこのクエリを 3.600、全体の 96.6 パーセントと相対的

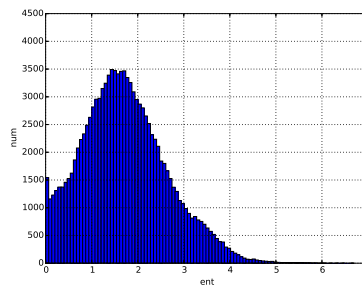


図 3 *click entropy* でのクエリ曖昧性のヒストグラム

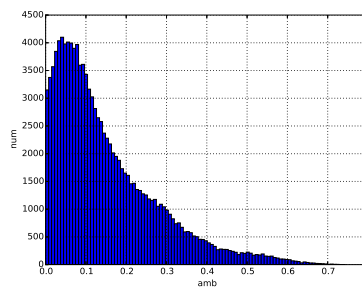


図 4 提案手法でのクエリ曖昧性のヒストグラム

に曖昧性の高いクエリと判定している。一方、提案手法では 0.117、全体の 50.3 パーセントと *click entropy* に比べ相対的に曖昧性の低いクエリと判定していることがわかる。「テーブル」というクエリの場合、一見するとクリックがばらけているように見えるが、実際にクリックされる商品はテーブルというトピックのものに偏っていることがわかり、提案手法ではこの意図の偏りをうまく数値に反映できているといえる。一方、2 つめの「バルーン」というクエリでは、先ほどの例とは逆にクリックされる商品のトピック自体が多岐にわたっていることがわかる。このようなケースでは、*click entropy*、提案手法ともにこのクエリを相対的に曖昧なクエリであると判定できている。3 つの目の「カルティエ 腕時計」のクエリは他とは傾向が異なる。このクエリの意図はクリック分布からメンズ腕時計とレディース腕時計の 2 つの主要なカテゴリで構成されていることがわかり、直感的には分布のばらつきという観点では曖昧性は低いと考えられる。しかし、提案手法ではクエリ曖昧性を 0.317、全体の 88.6 パーセントと相対的に曖昧性の高いクエリと判断してしまっている。これは、分散表現上でメンズ腕時計とレディース腕時計が近いところにマッピングされていないことに起因している。2 つのカテゴリはその関連性として、(1) 2 つは「腕時計」というトピックを共有しているため意味的に近い、(2) 「メンズ」と「レディース」は異なるトピックであり意味的に遠い、という 2 つの真逆の解釈ができ、今回のケースでは後者が強く分散表現に影響したのではないかと推測される。このような問題への対応としては、ユーザのデモグラフィック情報を組み合わせる方法が考えられる。

3.3.2 時間変化に対する安定性

クエリ曖昧性において時間変化に対する安定性は重要な特性の一つである [5]。この特性の評価を行うため、訓練データを

表 3 クエリ曖昧性の評価例

クエリ	カテゴリ ($p(c q)$)	ent (%tile)	amb (%tile)
テーブル	テーブル > センターテーブル > その他 (0.094) テーブル > センターテーブル > ガラス製 (0.089) テーブル > センターテーブル > 木製 (0.089) テーブル > ローテーブル > その他 (0.083) テーブル > 折りたたみテーブル (0.076) ... 合計 265 カテゴリ	3.600 (96.6)	0.117 (50.3)
バルーン	花, ガーデニング > フラワーアレンジメント > 一般 (0.094) ゲーム, おもちゃ > パーティグッズ > その他 (0.087) ファッション > レディース > バッグ > トートバッグ > その他 (0.086) ゲーム, おもちゃ > パーティグッズ > 室内装飾 (0.085) ファッション > レディース > スカートを > その他 (0.066) ... 合計 152 カテゴリ	3.775 (97.7)	0.624 (99.7)
カルティエ 腕時計	ファッション > アクセサリー > メンズ腕時計 > 腕時計 (0.555) ファッション > アクセサリー > レディース腕時計 > 腕時計 (0.409) ファッション > アクセサリー > 腕時計用品 > ベルト, バンド (0.017) ファッション > アクセサリー > レディース腕時計 > アンティーク (0.006) ファッション > アクセサリー > メンズ腕時計 > アンティーク (0.005) ... 合計 15 カテゴリ	0.871 (18.3)	0.317 (88.6)

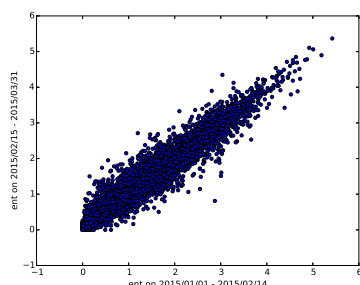


図 5 *click entropy* での 2 つの期間のクエリ曖昧性の相関図

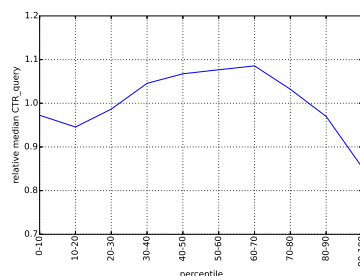


図 7 *click entropy* とクエリ毎 CTR の相対値との関係

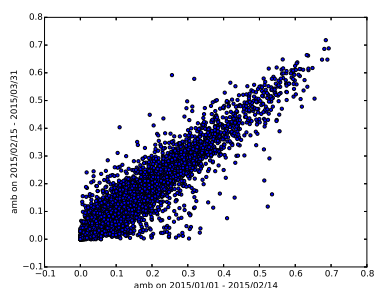


図 6 提案手法での 2 つの期間のクエリ曖昧性の相関図

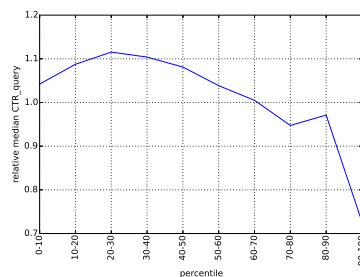


図 8 提案手法とクエリ毎 CTR の相対値との関係

2015/01/01 から 2015/02/14, 2015/02/15 から 2015/03/31 の 2 つの期間に分け, それぞれで求めたクエリ曖昧性の相関を調べた. 実験では 2 つの期間でそれぞれ 50 回以上出現するクエリの中から 5,000 クエリをランダムに選択し, 式 (7) で定義されるピアソンの相関係数を求めた.

$$r(X, Y) = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (7)$$

図 5, 6 はそれぞれ *click entropy*, 提案手法の 2 つの期間でのクエリ曖昧性の相関図を表す. *click entropy* 及び提案手法はピアソンの相関係数でそれぞれ 0.9446, 0.9275 と 2 つの期間で非

常に強い相関があり, 提案手法は *click entropy* と同等程度の時間変化に対する安定性を有しているといえる.

3.3.3 検索品質との相関

クエリ曖昧性の大小は検索の難しさと関連があると考えられるため, ここでは検索品質と求めたクエリ曖昧性との相関を見ることで, クエリ曖昧性の正しさを評価する. ここでは, 式 (8) で定義されるクエリ毎のクリック率 (CTR) を検索品質の指標として用いた.

$$CTR_{query} = \frac{\#click_{query}}{\#request_{query}} \quad (8)$$

CTR_{query} の計算には 3.1 節で述べたテスト用の検索ログを用

いた。検証で用いた検索ログは、クエリ曖昧性を考慮しないランキングモデルによって生成されているため、 CTR_{query} は純粋に検索結果とユーザの意図がどの程度一致していたかを表す指標といえる。

図 7, 8 はそれぞれ *click entropy* 及び提案手法でのクエリ曖昧性と CTR_{query} の相対値をプロットしたものである。図 7, 8 は各クエリをクエリ曖昧性に応じて 10 パーセントailずつに分割し、各グループでの CTR_{query} の中央値を計算しており、縦軸は得られた CTR_{query} の全クエリでのクエリ毎 CTR の中央値に対する相対値となっている。2つの図を比べると、提案手法の方が *click entropy* に比べてより数値が小さいグループで CTR_{query} がピークとなっていることがわかる。また、最小値に対する相対値でも、提案手法の方がより大きな値となっており、 CTR_{query} の大小をクエリ曖昧性により強く反映できているといえる。

CTR_{query} との相関を定量的に評価するために、各手法のクエリ曖昧性と CTR_{query} についてピアソンの相関係数とケンドールの順位相関係数を計算した。なお、ピアソンの相関係数は式 (7) で、ケンドールの順位相関係数は式 (9) で表される。ここで、式 (9) 中の P は順位関係が一致したペア数、 Q は順位関係が一致しなかったペア数、 T_x 及び T_y は各指標で同順位となったペアの数、 N は全てのペアの数を表す。

$$\tau(X, Y) = \frac{P - Q}{\sqrt{N - T_x} \sqrt{N - T_y}} \quad (9)$$

2つの手法と CTR_{query} との相関係数の値を表 4 に示す^(注2)。表 4 より、明確な相関が見られる程の数値ではないものの、提案手法の方が *click entropy* に比べ CTR_{query} に対してより高い相関が得られている。

表 4 各種法と CTR_{query} との相関係数

	<i>click entropy</i> (<i>ent</i>)	提案手法 (<i>amb</i>)
Pearson ($ r $)	0.0365	0.1092
Kendall ($ \tau $)	0.0191	0.0821

4. 関連研究

4.1 クエリ意図推定

クエリ意図推定には様々な情報に注目した研究が存在しており、例えばカテゴリ [3]、セッション [11]、サブクエリ [12]、潜在トピック [13] などを用いた手法が提案されている。本項は特に、カテゴリ情報及び潜在トピックを利用する手法と関連する。

Bennett 等 [3] はクリックした Web ページの分類に基づくランキングモデルを提案した。Bannett 等の手法では、Web ページのカテゴリとして open directory project (ODP) を利用し、クエリ毎に各カテゴリのクリック分布を求めることでクエリ意図を計算している。Bannett 等はそれらクエリ意図に基づく素性群をランキングモデルに追加することで、テールク

エリを含む幅広いクエリで精度改善を実現したと報告している。また Bannett 等の実験では、クエリ曖昧性を数値化した QueryClassEntropy が精度改善に大きく寄与したことが示されている。Yu 等 [13] は LDA ベースの手法で e コマースのクエリから検索意図を展開する手法を提案した。Yu 等は、カテゴリ分類によるアプローチでは粒度の不均一性やカテゴリの重複によるノイズが存在するため、クエリ意図を分類する上では不十分であると述べている。Yu 等は eBay のデータを用いて評価を行い、LDA ベースの提案手法が eBay の既存ランキングモデルに比べ、ユーザ満足度という観点から勝る結果が得られたと報告している。

4.2 クエリ曖昧性

クエリ曖昧性はクエリ意図の代表的な要素の一つであり、情報検索の改善に活用できることが知られている [1], [2], [3], [14]。Song 等 [6] はクエリ曖昧性の分類として *ambiguous*, *broad*, *clear* の 3 つを提案した。これらはクエリのトピック分布の広がりやの差異によって表現され、*ambiguous* はメイントピックが複数あるクエリ、*broad* は一つのメイントピックの中に複数のサブトピックがあるクエリ、*clear* はトピックが一意に決まるクエリ、と定義される。また、Song 等は SVM を用いて実際にクエリを 3 つのカテゴリに分類することを試しており、検証したログにおいて 16% のクエリが *ambiguous* となったと報告している。Teevan 等 [15] は曖昧なクエリの自動分類について研究を行っており、クエリ曖昧性を測る手法として *potential for personalization curve* を提案した。また、Teevan 等は実験の中でクリックのような暗黙的指標で人手ラベルによる明示的指標をうまく追従できることを示している。Hafernik 等 [16] はクエリ長と品詞情報を用いて曖昧なクエリを特定する手法を提案した。Hafernik 等はクエリを *narrow* と *general* の二種類に分類し、前述の素性がそれらの分類に効果的かどうかの調査を行っている。Luo 等 [17] は幾つかのユーザの行動に基づく素性を用いて曖昧なクエリを分類する手法を提案した。Luo 等はセッション中のクエリ列からクエリの分散表現を学習する手法として *query2vec* を提案しており、*query2vec* で得られる前後 k クエリの分散表現をセッション素性として利用している。

クエリ曖昧性の定量化では、Dou 等 [2] によって提案された *click entropy* が広く知られている。*click entropy* はクリック分布のエントロピーを計算することでクエリの曖昧性を定量化する。Duo 等は *click entropy* をパーソナライズを行うかどうかを判断する指標として利用しており、*click entropy* が小さいクエリではパーソナライズが逆に精度悪化に繋がることを示している。クエリ曖昧性の定量化には、*click entropy* を拡張した幾つかの手法が提案されている。Wang 等 [4] はユーザ毎に *click entropy* を平均化する *user entropy* を提案した。Wang 等は *user entropy* は低頻度のクエリを扱う場合に効果が高いことを示している。Duan 等 [5] は単純にクリックそのものを使うのではなく、ユーザのクリックパターンを用いる *pattern entropy* を提案した。Duan 等はクリックパターンを用いることで、複数の URL にクリックが分散するようなケースをうまく扱えると述べている。本項の手法は、*click entropy* に関連する手法群と

(注2) : 曖昧性と CTR_{query} は負の相関となるため、ここでは絶対値を提示

クエリ曖昧性の定量化を行う点は共通しているが、クリックした文書群の関連性を考慮する点が大きく異なる。

5. おわりに

本項では、文書間の類似性に注目し、文書の分散表現を活用したクエリ曖昧性の評価手法を提案した。Yahoo!ショッピングの商品データ及び検索ログを用いて提案手法を評価した結果、表面上はクリックのばらつきが大きいですが、潜在的には近いトピックの文書のクリックしているような、従来手法ではその曖昧性の低さを検知できないクエリを正しく認識することに成功した。また、手法の特性評価の結果、提案手法が従来手法と同程度の時間変化に対する安定性を有しており、また従来手法よりもクエリ毎 CTR と高い相関が得られることを示した。

今後の課題としては、オープンデータセットでの手法評価があげられる。本項では Yahoo!ショッピングという e コマースの領域での評価にとどまっているが、これをオープンデータセットを用いて同様の評価を行い、一般の問題に対しても有効であるかを示す必要がある。また、評価指標として TREC や NICIR といった ground truth なラベルの付与されたデータセットを用い、より正確な評価を行うことが望ましい。さらに、ランキングモデルや意図推定などと組み合わせて、実システムへの活用を今後検討していく必要がある。

文 献

- [1] R. L.T. Santos, C. Macdonald, and I. Ounis. Selectively diversifying web search results. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, CIKM '10, 2010.
- [2] Z. Dou, R. Song, and J.-R. Wen. A large-scale evaluation and analysis of personalized search strategies. In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, 2007.
- [3] P. N. Bennett, K. Svore, and S. T. Dumais. Classification-enhanced ranking. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, 2010.
- [4] Y. Wang and E. Agichtein. Query ambiguity revisited: clickthrough measures for distinguishing informational and ambiguous queries. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 2010.
- [5] H. Duan, E. Kiciman, and C. Zhai. Click patterns: An empirical representation of complex query intents. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, CIKM '12, 2012.
- [6] R. Song, Z. Luo, J.-R. Wen, Y. Yu, and H.-W. Hon. Identifying ambiguous queries in web search. In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, 2007.
- [7] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, March 2003.
- [8] Quoc Le and Tomas Mikolov. Distributed representations of sentences and documents. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 1188–1196, 2014.
- [9] S. Deerwester, S. T. Dumais, W. Furnas G, T. K. Landauer, and R. Harshman. Indexing by latent semantic analysis. *JOURNAL OF THE AMERICAN SOCIETY FOR*

INFORMATION SCIENCE, 41(6), 1990.

- [10] E. Jones, T. Oliphant, P. Peterson, et al. SciPy: Open source scientific tools for Python, 2001–. [Online; accessed 2015-06-23].
- [11] H. Cao, D. H. Hu, D. Shen, D. Jiang, J.-T. Sun, E. Chen, and Q. Yang. Context-aware query classification. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '09, 2009.
- [12] Y. Hu, Y. Qian, H. Li, D. Jiang, J. Pei, and Q. Zheng. Mining query subtopics from search log data. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, 2012.
- [13] J. Yu, S. Mohan, D. (P.) Putthividhya, and W.-K. Wong. Latent dirichlet allocation based diversified retrieval for e-commerce search. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, WSDM '14, 2014.
- [14] M. Sanderson. Ambiguous queries: Test collections need more sense. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, 2008.
- [15] J. Teevan, T. Dumais, S, and D. J. Liebling. To personalize or not to personalize: Modeling queries with variation in user intent. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, 2008.
- [16] C. T. Hafernik and B. J. Jansen. Understanding the specificity of web search queries. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, 2013.
- [17] C. Luo, Y. Liu, M. Zhang, and S. Ma. Query ambiguity identification based on user behavior information. In *Information Retrieval Technology*. 2014.