5S-6

検索質問間の関係を考慮したランキング関数の学習
Learning a query dependent ranking function

吉川 幹人† 関 和広†† 上原 邦昭†††

†神戸大学大学院工学研究科 ††神戸大学自然科学系先端融合研究環

†††神戸大学大学院システム情報学研究科

我々が情報検索を行う際，一度の検索では目的の情報を発見できず，検索質問（query）を修正しながら連続して検索を行うことがある．このような「Query Chain」を利用することで，検索質問と（非）適合文書とを関連づけた学習データを効率的に自動生成する手法が提案されている．しかし，Query Chain によって作成した訓練事例を用いた検索は，学習データに出現しない検索質問に対してはうまく機能せず，一般的なウェブ検索等に用いることは困難であった．本研究では，検索質問間の類似性を考慮して訓練事例を選択的に利用することにより，この問題の解決を試みる．また，より高品質・多量の訓練事例を獲得するために Query Chain の拡張を行なう．さらに，実データを用いた評価実験によって提案手法の有効性を検証する．

1.はじめに

情報検索の分野ではさらなる精度向上のため，ランキング関数の導出に機械学習を利用する手法が多数提案されている．しかし，一般的に機械学習で利用する学習データを人手で作成することは高コストであり，現実的ではない．そこで Radlinski らは，Query Chain [1]という検索質問と（非）適合文書とを関連づけた学習データを効率的に自動生成する手法を提案した．この手法では，一度の検索では目的の情報を発見できなかった際に，検索質問を修正しながら連続して検索を行うというユーザの性質を利用することで，より質の高い学習データを生成することに成功している．（この関連したクエリの連鎖が Query Chain と呼ばれる．）しかし，この手法では学習データに存在しない検索質問に対してはうまく機能しないといった問題があり，多種多様な検索質問が想定されるウェブ検索等で用いることは困難であった．本研究では，Query Chain の枠組みを基に，学習データに存在しない検索質問にも対応できる手法を提案する．さらに，より高品質で多量の訓練事例を獲得するために Query Chain の拡張を行う．

2.提案手法

本節ではクエリログから学習データを生成し，その学習データを用いて検索を行う流れについて説明する．クエリログには，検索質問やその検索結果，それに対してユーザがどの文書をいつクリックしたかということなどが記録されている．まず，クエリログに記録された各検索質問間の類似性やその検索質問の検索結果として与えられた上位数件の文書の類似性などから，連続した検索質問が Query Chain となっているかどうかを機械学習によって同定する．この際，約 10%程度の検索質問に関しては人手で Query Chain の同定を行い，これを学習データとして残りの Query Chain を自動的に同定する．

次に，こうして得られた Query Chain を利用し，ランキング関数の学習に使用するための学習データを生成する．Radlinski らが提案した Query Chain からの学習データ生成には，例えば次のような規則がある．「検索質問 q に対する検索結果の内，検索結果の 3 番目がクリックされた場合，3 番目の文書は q に対して 1 番目と 2 番目の文書よりも適合度が高い．さらに検索質問 q の前に検索質問 q'で検索され，その 2 つが Query Chain であると判断された場合には q'に対しても同様の関係が成り立つ．」これを図で表すと，図 1 のようになる．

q’	q
1’	1
2’	2
3’	3（Click）

$3 >_q 2$　$3 >_q 1$

$3 >_{q'} 2$　$3 >_{q'} 1$

図 1.Query Chain の例

† Graduate School of Engineering, Kobe University
†† Organization of Advanced Science and Technology, Kobe University
††† Graduate School of System Informatics, Kobe University

この規則は，たとえクリックせずとも，検索結果に表示されるタイトルや要約などに基づいて，ユーザはその文書が自分の情報要求に合致するかを判断しており，クリックした文書より上位の文書は情報要求に合致しない（と判断された）という仮定に基づいている．このように，Query Chain で定義されている訓練データ生成規則は，一般的なユーザが検索を行った際に取る行動に基づいて作成されている．我々は Radlinski らの手法で用いられておらず，かつ訓練事例の生成に有用であると考えられるユーザの行動特性を新たに 2 つ同定し，Query Chain からの学習データ生成規則を拡張した．1 つ目の拡張は，「Query Chain の最後でクリックされた文書はその Query Chain 中でクリックされた他のどの文書よりも適合度が高い」という規則である．これは Query Chain の最後にクリックされた文書は，その文書の閲覧によってユーザがある情報に関する検索を終えたことを意味しており，これは他の文書に比べて適合度が高いと考えられるためである．2 つ目の拡張は，「Query Chain の最後のクリックではなく，かつそのすぐ下の文書をクリックしていない場合，クリックされた文書はそのすぐ下の文書より適合度が高い」という規則である．これは，Query Chain の最後ではないことからユーザは情報の探索を継続しており，そのすぐ下の文書の適合性も判断済みであると考えられることによる．

続いて，ユーザが検索質問を与えると，上述の処理で得られた学習データから，与えられた検索質問と意味的に類似した訓練事例を抽出する．これにより，学習データに出現しない検索質問に対しても効果的な学習が期待できる．具体的には，まず前処理として検索エンジン Lucene を用いて検索質問に関する上位 10 件の検索結果を取得する．そして，PLSI [2]を用いて，与えられた検索質問と Lucene で得た結果の関係性を潜在的な意味属性による表現に変換する．同様に，学習データ（検索質問と適合文書の組）を潜在的な意味属性空間に射影することで，所与の検索質問に類似した訓練事例を同定・抽出する．このようにして得た訓練事例を用いて SVM に基づく検索ランキング関数の 1 つである RankSVM [3]を学習することで，所与の検索質問に関する文書のランキングを決定する．

3.評価実験

評価実験には，独自に収集したクエリログを学習データ，NTCIR WEB Task で使用されたデータをテストデータとして用いた．学習データに含まれる検索質問総数は 14,168 であり，ここから生成された訓練事例総数は 11,970 件であった．比較のため，Radlinski らの Query Chain をベースラインとし，評価尺度には MAP 値を用いた．また，PLSI によって抽出する訓練事例数は 50 とした．表 1 に結果を示す．

表 1.ベースラインとの比較

	提案手法	ベースライン
MAP	0.0777	0.0674

提案手法では 0.01 ポイント（15.3%）の精度向上が見られた．

次に PLSI に基づいて抽出する訓練事例数と検索精度の関係について調べた．表 2 は，抽出する訓練事例を変化させた際の検索精度の変化を表したものである．

表 2.訓練事例数と検索精度の関係

抽出数	10	50	100
MAP	0.0676	0.0777	0.0679

この結果から，抽出する訓練事例数が検索精度に影響を与えていることが分かる．抽出数が大きすぎた場合には，検索質問にあまり類似しない事例が学習に用いられるため，PLSI の効果を得ることができず，一方，小さすぎた場合には学習データの規模が非常に小さくなり満足のいく学習ができなかったものと考えられる．今後，最適な事例数決定の方法を検討する必要がある．

4.結論

Query Chain をウェブ検索に使用する際に，PLSI による訓練事例のサンプリングを行うことにより，検索精度が向上することを示した．その際，訓練事例抽出数が検索精度に影響を与えていた．最適な事例数を決定する方法の検討，より大規模な実験に基づく提案手法の評価・議論が今後の課題である．

参考文献

[1] F. Radlinski and T. Joachims. Query chains: Learning to rank from implicit feedback. In Proc. of KDD’05, pp. 239-248.
[2] T. Hofmann. Probabilistic latent semantic indexing. In Proc. of SIGIR’99, pp. 50-57.
[3] Y. Cao, J. Xu, T. Liu, H. Li, Y. Huang, and H. Hon. Adapting ranking SVM to document retrieval. In Prof. of SIGIR’06, pp. 186-193.