

# 緩和検索におけるクエリの構造推定に基づく ユーザの意図を反映したランキング手法

金子 恭史<sup>†</sup> 中村 聡史<sup>†</sup> 田中 克己<sup>†</sup>

京都大学大学院情報学研究科<sup>†</sup>

## 1. はじめに

検索語入力型 Web 検索サービスは情報収集ツールとして必要不可欠なものとなりつつあるが、入力された検索語を含む Web ページしか得られず、ユーザの検索意図を十分に考慮できていない。我々は、ユーザの検索意図が多様であることを考慮し、ユーザが確信を持っていない検索語に対して「緩和度」という値を付加することで、その検索後を緩和した（その検索語と関連のある語も含む）検索結果を返す検索システム（緩和検索）を提案してきた[1]。本稿では、ユーザの意図はユーザの入力した検索語に対する緩和度と、入力された検索語間の関係から推定されるクエリの構造に基づくものであると仮定し、入力されたクエリの構造推定および推定された構造に基づく検索結果のランキング手法を提案する。

なお、本稿では緩和する検索語を「?」で表現することとする。例えば「京都 豆腐? 和食」は豆腐を緩和するように指定されたクエリを意味する。また、緩和される検索語やその関連語を緩和語、それらをまとめて緩和語群と呼ぶこととする。

## 2. クエリの構造推定

緩和検索では緩和度の付加された検索語に対してその語と関連のある語をうまく選ぶ必要があるが、ユーザが求める関連語はユーザの想定している検索語間の関係性によって異なる。例えば「京都 豆腐? 和食」というクエリが与えられた時、ユーザの意図が「豆腐とか京都らしい食材を」であるときは京都と豆腐が強く結び付いているため、湯葉や京野菜といった関連語が候補として考えられる。一方、ユーザは和食料理の一つとして豆腐を挙げており、京都は場所を示すためだけに入力したのならば、刺身や焼き魚といった和食と関係性のある語を望んでいると考えられる。

求められる関連語を推定するには、ユーザが緩和語とその他の検索語との間にそれぞれのどのような関係性を想定しているかを考慮する必要がある。本稿では、このような緩和語と検索語の関係性の組を構造と呼ぶ。一つのクエリに対して考えられる構造は一通りとは限らない。そこで、緩和度の付加された検索語に対してまず緩和語群を取

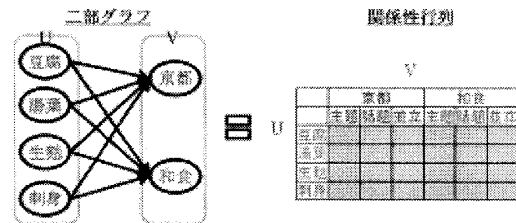


図 1: 二部グラフと関係性行列

得し、各緩和語と他の検索語との関係性を取得する。そして、得られた関係性を基に複数の構造とその構造に適合する緩和語を推定する。

### 2.1. 緩和語群と他の検索語間の関係性の取得

本稿では、緩和語と他の検索語間の関係性をまとめて取り扱うために、二部グラフを構成する。ノード集合  $U$ ,  $V$  はそれぞれ緩和語と検索語からなり、関係性を表すエッジが緩和語群から他の検索語へと張られる。 $u \in U$  と  $v \in V$  間のエッジの値は  $u$  に対する  $v$  の関係性の強さを表す。本稿では、2 語間に成り立つ関係性として以下の 3 つの関係性を利用する。

- $u$  に対する  $v$  の主題らしさ
- $u$  に対する  $v$  の話題らしさ
- $u$  と  $v$  が並立に扱われている度合い

各関係性の強さを Web 上から取得し、それらの値を基にグラフを生成する。「京都 豆腐? 和食」の場合の例を図 1 に示す。図 1 の右側に示したように、二分グラフは行列と見なすことができる。本稿ではこの行列を関係性行列と呼ぶこととする。ここで、この二部グラフは一つのエッジが 3 種類の関係性の強さを持つため、図 1 のように関係性の種類ごとに各  $v$  を分割して扱うこととする。

### 2.2. クエリの構造の推定

関係性行列から各緩和語が持つ関係性の複数の傾向を抽出する。本稿では行列を  $r$  個の軸に分解する NMF[2] という手法を用いて抽出を行う ( $r$  は手動で設定)。NMF は元の  $n \times m$  行列  $M$  を  $n \times r$  重み行列  $W$  と  $r \times m$  特徴行列  $H$  に分解する手法で、 $M$  の行成分と列成分をそれぞれ  $r$  個の軸に分解することができ、 $W$  の  $i$  列目と  $H$  の  $i$  行目は同じ  $i$  番目の軸に対応している。

関係性行列に NMF を適用した場合、 $H$  の各軸はそれぞれ一組の関係性の傾向と対応するため、 $H$  の各軸はそれぞれ異なる構造を表していることに

Ranking Method for Relaxation Search based on Detected Structures

<sup>†</sup>Graduate School of Informatics, Kyoto University

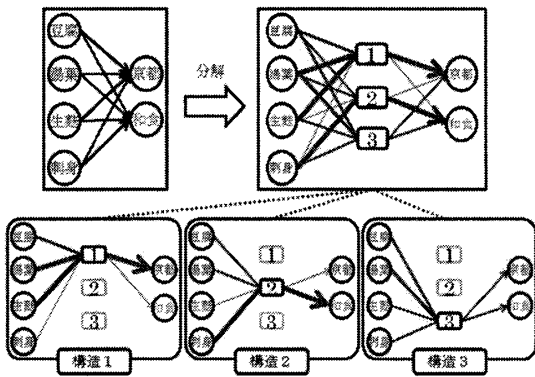


図 2: 行列の分解による構造推定

なる。また、各軸を異なる構造と見なした場合、同時に得られる  $W$  は各構造における各緩和語の重みを表していると見なすことができる。このように、NMF によって複数の構造とその構造における緩和語の重み同時を抽出することができる。抽出例をグラフで表したものを図 2 に示す。図中のエッジの太さは値の大きさを表す。

### 3. 構造に基づく検索結果の生成

始めに検索結果に反映させる構造を選択する。ここでは緩和される語 (例では豆腐) の重みの高い構造ほどユーザが想定していた構造だと仮定し、重みが高い構造から順に提示される。これを基にユーザが構造を選択することでその構造に対応する結果を得ることができる。なお、選択されなかった場合は緩和される語の重みが最も大きい構造を自動的に採用することとする。次に、緩和語を関連語で置き換えることによって複数のクエリを生成し、それらのクエリ群からページ集合を取得する。そして、ページ  $p$  のスコア  $Score(p)$  は以下のようにして算出する。

$$Score(p) = \sum_{u \in RT(p)} w(t)$$

ここで  $RT(p)$  は  $p$  のタイトルとスニペット中に含まれる緩和語、 $w(t)$  は  $t$  の重みを表す。最後にこのスコアに基づきページを並び替えることで構造ごとに異なる検索結果が生成される。

クエリ「京都 豆腐? 和食」に対する検索結果を図 3 に示す。各構造は語集合として提示し、重みの大きい語ほど大きく表示する。これによりユーザは構造を意識することなく、自分の興味のある語集合を手軽に選ぶことができ、その語を含むページを得ることができる。

### 4. まとめ

本稿では緩和検索における複数の構造の推定手法と、構造に応じたランキング手法を提案した。

緩和検索では、「京野菜や生湯葉など京都らし

図 3: システム実行例

いものに特に興味があり、他にも生麩や焼き魚にも少し興味がある」といったように、各緩和語に対するユーザの興味の度合いは異なる場合が多い。しかし、ユーザにとってこうした緩和語ごとの興味を逐一入力するのは大きな手間である。これに対して我々の検索システムでは各構造に対する各緩和語の重みが得られているため、ユーザは構造を一つ選択するだけで各緩和語の興味の度合いという複雑な意図を伝えることができ、その意図が反映された検索結果を得ることができる。

一方、提案手法はタイトルやスニペットに出現する緩和語の重みの和に基づくため、適合するページであっても緩和語がタイトルやスニペットに出現しないものは上位に現れなかった。例えば、「京都のおばんざい」というタイトルのページは京都の和食料理に関するページだが、「おばんざい」という語を緩和語として持っていなければ高いスコアは与えられず、上位には並べられない。今後はこのようなページもうまく上位に並べられるように緩和語の重み以外の要素も考慮できるスコアリング手法を考えていく予定である。

### 謝辞

本研究の一部は、京都大学グローバル COE プログラム「知識循環社会のための情報学教育研究拠点」、文部科学省科学研究費補助金特定領域研究「情報爆発時代に向けた新しい IT 基盤技術の研究」、計画研究「情報爆発に対応するコンテンツ融合と操作環境融合に関する研究」(研究代表者: 田中克己, 課題番号 1809041) によるものです。ここに記して謝意を表すものとします。

### 【文献】

- [1] Y.Kaneko, S.Nakamura, H.Ohshima, and K.Tanaka: "Query Relaxation Based on Users' unconfidences on Query Terms and Web Knowledge Extraction", ICADL 2008, pp. 71-81
- [2] D.lee, and H.Seung: "Algorithms for non-negative matrix factorization", Advances in neural information processing systems, pp. 556-562 (2001)"