

# 評判検索のための対象，属性，評価の3項関係のランキング法

## Ranking Method of Triplet Reputation for Reputation Retrieval

土田正明

Masaaki TSUCHIDA

水口弘紀

Hironori MIZUGUCHI

久寿居大

Dai KUSUI

NEC 共通基盤ソフトウェア研究所

Common Platform Software Res. Labs., NEC Corp.

In this paper, we propose a ranking method for opinion retrieval. Our approach is to model the confidence score of opinion for ranking. There are two characteristics in our confidence score model. Firstly, we divide opinion as object, attribute, evaluation triplet into object-attribute pair and attribute-evaluation pair. Secondly, our confidence score model consists of syntactic and semantic confidence score. Two experiments on evaluation of the results of ranking show that our methods is effective.

### 1. はじめに

近年，インターネットの爆発的な普及により，個人や団体が自由に情報を発信できるようになった．特に最近，ウェブログ（以下，ブログと呼ぶ）の普及により，日々大量の記事が発信されている．これら記事の中から評判を自動的に収集したり分析することができれば，商品購入の際の評判検索，マーケティング，企業のリスク管理など，幅広い用途に利用できる．

筆者らは，評判情報を「ある対象物の属性（評価している注目ポイント）についての評価の情報」と定義して，ブログ記事から検索・分析するシステムを開発している [土田 08]．「対象物」は商品やサービスなどの名称，「属性」は対象物の特徴や性質や構成要素，「評価」は評価者の主観的な評価を表す表現である．

評判検索システム [土田 08] では，あらかじめ属性，評価の2項関係を抽出してインデックスを作成しておき，検索時には，ユーザの検索キーワードを対象物と考えて，対象物と十分に近い属性，評価を対応付けて3項の評判情報と見なす．

しかしながら，インデックス内には，抽出ミスの属性，評価の2項関係が含まれる．また，対象物と属性が十分に近いからといって，必ずしも対象物と属性に対応関係があるとは限らない．つまり，検索結果の評判情報を無秩序に提示すると，上位に正しくない評判情報が含まれてしまうこともある．

そこで，本稿では，評判情報と抽出ミスを区別するために，「評判情報として正しそうな度合い」を確信度として付与する方法を提案する．また，精度向上の第一歩として，確信度によるランキングで，検索結果上位の精度向上を目指す．

単純には，確信度として，機械学習による属性，評価の2項関係の評判抽出器 [土田 08] の出力値を使用できる．しかしながら，それだけでは，1) 評価表現が属性表現を評価する語として尤もらしいか，が考慮されていない．また，対象物と属性の情報を使用していないため，2) 属性表現は対象物の属性として妥当であるか，3) 対象物と属性表現が文書中で対応関係にあるか，が考慮されていない．

提案法では，抽出器の出力値に加えて，さらに前述の3項の情報を考慮することで，構文的にも意味的にも尤もらしい場合に評判情報の確信度が高くなるよう計算する．構文的な尤もら

しさは，属性，評価の2項関係の抽出器の出力値と対象物と属性の距離から計算する．意味的な尤もらしさは，大量のコーパスから推定した語の共起確率を利用して，よく共起する語の組み合わせを持つ評判情報ほど高くなるように計算する．

以降，第2節でベースとなる評判検索システム [土田 08] の概要を説明する．第3節では，提案の評判情報の確信度モデルとその具体的な計算方法を述べる．第4節で提案法を評価，第5節で考察し，第6節でまとめる．

### 2. 評判検索システム:eHyouban

本研究のベースとなる評判検索システム:eHyouban [土田 08] は，1) あらかじめ属性，評価の2項関係をインデキシングしておき，2) 検索時にユーザの検索キーワードと属性を対応付けることで，3項関係の評判情報を検索する．図1に処理の流れを示す．評価表現辞書には「いい，かわいい，悪い」など，評価に用いる約13000表現が登録されている．

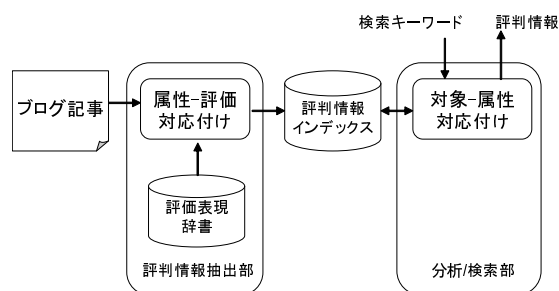


図1: 処理の流れ

1) の属性，評価のインデキシングは2ステップからなる．具体的には，i) 評価表現と1つ前の文までに含まれる全名詞を属性候補と考えて，評価表現に対応する名詞を同定して（対応同定），ii) 名詞，評価表現の対応関係が評価の情報であるか否かを判定することで（意見性判定），評価表現が名詞を評価しているペアのインデックスを作る．i) は，トーナメントモデル [飯田 05] を用いる．トーナメントモデルでは，1つの評価表現に対して，対応する可能性がある属性候補から最尤の候補を決定するために，候補間で勝ち抜き戦を行う．そのため，2つ

の候補間の勝敗を決める関数をサポートベクターマシンで学習する。ii) は、i) で同定された各ペアが評価情報であるか否かをサポートベクターマシンで判定する。

2) では、属性と検索キーワードが十分近くに出現する場合に対応づける。本システムでは、検索キーワードと属性の間が50文字以内の場合に対応付けている。これは、互いの出現位置が十分に近ければ、対応関係にある可能性が高いという仮定に基づいている。本仮定は、先行研究[土田 07]により、対象物と属性の距離が近いほど抽出精度が向上することが実験的に示されていることから、妥当と考えられる。

### 3. 確信度付与方法

#### 3.1 評判情報の確信度のモデル化

評判情報の3項関係を、1) 対象, 属性, 2) 属性, 評価の2つの2項関係に分解する。3項関係の確信度  $C_{oae}$  も、対象, 属性の確信度  $C_{oa}$  と属性, 評価の確信度  $C_{ae}$  に分解する。

$$C_{oae}(o, a, e, S) = C_{oa}(o, a, S)C_{ae}(a, e, S)$$

$o$  は対象物,  $a$  は属性,  $e$  は評価,  $S$  は3項の各要素が現れた範囲にある文の集合とする。

また、提案モデルでは、確信度を構文的確信度と意味的確信度に分けて考える。

- 構文的確信度：構文的に3項が評判情報でありそうな度合い
- 意味的確信度：意味的に3項の語が対応していそうな度合い

確信度を構文的確信度, 意味的確信度の2つに分ける狙いを具体例から説明する。

1. ラーメン屋 A にいった。美味しいラーメンだったら良いけど。
2. ラーメン屋 A にいった。ラーメンが美味しかった。
3. 秋葉原にいった。昼食が美味しかった。

1と2は、どちらも「ラーメン屋 A, ラーメン, 美味しい」の3項関係が含まれている。しかしながら、1は、仮定の話であるため、評判情報として適切でない。このように、構文的に評判情報であるか否かを評価する尺度が構文的確信度である。2と3は、どちらも「A にいった。B が C。」という構文である。しかしながら、3は「昼食」が「秋葉原」の属性ではないので、評判情報として適切でない。このように、意味的に語の3項関係が成りなっていそうかを評価する尺度が意味的確信度である。

まとめると、確信度  $C_{oae}$  のモデルは以下の通りとなる。 $C_{sem}$  は意味的確信度,  $C_{syn}$  は構文的確信度を表す。

$$\begin{aligned} C_{oae}(o, a, e, S) &= C_{oa}(o, a, S)C_{ae}(a, e, S) \\ C_{oa}(o, a, S) &= C_{sem_{oa}}(o, a)C_{syn_{oa}}(o, a, S) \\ C_{ae}(a, e, S) &= C_{sem_{ae}}(a, e)C_{syn_{ae}}(a, e, S) \end{aligned}$$

#### 3.2 確信度の計算方法

前節のモデルの確信度を計算するために、対象, 属性の2項関係の1) 意味的確信度  $C_{sem_{oa}}$  と2) 構文的確信度  $C_{syn_{ae}}$ , 属性, 評価の2項関係の3) 意味的確信度  $C_{sem_{ae}}$  と4) 構文的確信度  $C_{syn_{oa}}$  の4つの値が必要である。本研究では、それぞれ以下に基づいて計算する。

1.  $C_{sem_{oa}}$ : 検索結果における対象物と属性の共起確率
2.  $C_{syn_{oa}}$ : 対象物と属性の間の文字数
3.  $C_{sem_{ae}}$ : 大量の名詞, 評価の係り受け関係から推定した属性, 評価の共起確率
4.  $C_{syn_{ae}}$ : 属性, 評価の2項関係の抽出器の出力スコア

対象, 属性の意味的確信度:  $C_{sem_{oa}}$

対象物と属性の意味的確信度は、よく共起する語は、対象物に関連が深いため属性の可能性が高いという仮定から、対象物と属性の共起確率として計算する。具体的には、以下の式で計算する。 $A$  は、 $o$  で検索した結果に含まれる属性の集合、 $freq(x)$  は、検索結果中の  $x$  の出現頻度とする。ただし、出現頻度をそのまま使用すると、頻度の影響が強すぎると考えられるため、 $e$  を底とする対数関数で変換している。

$$C_{sem_{oa}}(o, a_i) = \frac{\ln(freq(a_i) + 1)}{\sum_k \ln(freq(a_k) + 1)}$$

$$a_i \in A, (0 \leq i \leq n)$$

検索結果から計算する理由は、事前にコーパスから共起関係を収集して、共起確率を推定しておくことが難しいためである。事前に共起確率を推定するには、全ての対象キーワードの共起関係を収集しておかなければならない。しかしながら、事前に全検索キーワードを予測することは不可能である。そのため、提案法では、任意のキーワードを対象として、属性との共起確率が計算できるように、検索結果から計算する。

対象, 属性の構文的確信度:  $C_{syn_{oa}}$

対象物と属性の構文的確信度は、対象物と属性の出現箇所が近いほど関係がある可能性が高いと仮定して、対象物と属性の間の文字数の逆数として計算する。具体的には以下の式で計算する。 $length(o, a, S)$  は、 $S$  内での対象物  $o$  と属性  $a$  の間の文字数とする。ただし、文字数をそのまま使用すると、文字数の影響が強すぎると考えられるため、 $e$  を底とする対数関数で変換してから逆数を取っている。

$$C_{syn_{oa}}(o, a, S) = \frac{1}{\ln(length(o, a, S) + 1) + 1}$$

係り受け解析や照応解析などを用いて、様々な構文情報から確信度を計算する方法も考えられる。一方、評判検索システムでは、検索時に対象, 属性, 評価の評判情報が確定するため、検索時に確信度を計算する必要がある。検索時に係り受け解析や照応解析を実行してしまうと、検索速度が遅くなってしまふ。そのため、対象物, 属性の構文的確信度は、高速に取得できる文字数のみから計算してい。

属性, 評価の意味的確信度:  $C_{sem_{ae}}$

属性と評価の意味的確信度は、属性を評価する表現として典型的であれば属性と評価が対応関係にある可能性が高いと仮定して、大量に収集した名詞と評価の係り受け関係から推定し

た属性と評価の共起確率とする．具体的には，共起確率の推定法である PLSI [Hofmann 99] を用いて計算する．

$$Csem_{ae}(a_i, e_j) = \sum_k p(z_k)p(a_i|z_k)p(e_j|z_k)$$

$$e_i \in E, (0 \leq i \leq n)$$

$$a_j \in A, (0 \leq j \leq m)$$

$n$  は収集された係り受け関係に存在する評価表現の異なり数， $m$  は収集された係り受け関係に存在する名詞の異なり数である．

PLSI は，パラメタ  $p(z)$ ,  $p(a|z)$ ,  $p(e|z)$  を推定する手法である． $p(z)$  は隠れトピックの出現確率， $P(a|z)$  は，隠れトピック  $z$  で  $a$  が出現する確率， $p(e|z)$  は隠れトピック  $z$  で  $e$  が出現する確率である．式を見て分かる通り，同じ隠れトピックで出現しやすい属性，評価のほど高い共起確率となる．

PLSI では，それぞれ属性が  $A$  に，評価が  $E$  に存在しなければ共起確率が計算できない．そのため，存在しない場合は，検索結果中で計算可能な属性，評価の共起確率の平均値を割り当てる．

属性，評価の構文的確信度： $Csyn_{ae}$

属性と評価の構文的確信度には，属性，評価の 2 項関係抽出の意見性判定のスコアを用いる．意見性判定は，サポートベクターマシンを用いて，1) 属性と評価が係り受け関係であるか，2) 属性と評価の間の文節数，3) 属性と評価が同じ文であるか，3) 属性文節の形態素情報，4) 評価文節の形態素情報，5) 属性文節の係り先の文節の形態素情報，6) 評価文節の係り先の文節の形態素情報，を素性として属性，評価の 2 項関係が評判情報であるかを判定している．文節の形態素情報は，文節内の各形態素の形態素文字列，原形，品詞である．前記 6 種の素性は，構文的情報と考えられる．そのため，意見性判定のスコアは属性，評価の構文的確信度と見なせる．

本研究では，サポートベクターマシンの出力スコア（分離超平面との距離）をシグモイド関数により変換する．シグモイド関数で変換する理由は，サポートベクターマシンの出力の確率値へのよい近似を与えることが実験的に示されている [Platt 99] ためである．

$$Csyn_{ae}(a_i, e_j, S) = \text{sigmoid}(F(f(a_i, e_j, S)))$$

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}}$$

$f(a_i, e_j, S)$  は， $S$  から  $(a_i, e_k)$  の対応関係の素性ベクトル  $x$  を返す関数， $F(x)$  はサポートベクターマシンで学習した意見性判定の関数である．

## 4. 評価実験

### 4.1 実験の目的と方法

本実験では，提案法の確信度に基づくランキングの効果を評価する．具体的には，実験 1：複数のキーワードで，提案法の確信度でランキングした上位 50 件の精度と，ランダムサンプリングで取得した 50 件の精度を比較する．また，実験 2：1) 構文的確信度 ( $Csyn$ )，2) 意味的確信度 ( $Csem$ ) のそれぞれの効果を，単独，組み合わせでランキングした場合の精度と平均適合率を比較することで評価する．

平均適合率とは，「各適合文書が検索された時点での精度の平均 [岸田 02]」であり，情報検索の評価指標に広く用いられている．平均適合率は，上位に適合文書（正解）があるほど高くなるように計算される．

表 1: ランキング上位，ランダムサンプルの 50 件の精度

キーワード	ランキング	ランダム	評判数
E O S 5 D	<b>0.30</b>	0.24	70
F 9 0 5 i	<b>0.66</b>	0.62	171
クロサギ	<b>0.78</b>	0.46	2341

### 実験データ

実験 1，2 とともに共通の実験データを用いる．検索対象の文書データは，独自に収集した 2008 年 3 月 1 日から 3 月 31 日に書かれたブログ記事である．検索キーワードは，恣意性を排除するために，1) 映画，2) 携帯電話，3) デジタルカメラ，について 2008 年 3 月 30 日までのランキング\*1 の 5 位とした．結果，1) 「クロサギ」，2) 「F 9 0 5 i」，3) 「E O S 5 D」となった．

属性，評価の意味的確信度を計算するために，ブログ 100 記事から名詞，評価の係り受け関係を収集して，PLSI で各種パラメタを推定した．低頻度の係り受け関係はノイズになると考え，2 回以上出現した係り受け関係を用いた．2 回以上出現した係り受け関係は，86119 種からなり，構成要素は名詞 16173 種，評価表現 4450 種であった．PLSI の隠れトピック数は 20 とした．隠れトピック数は，本実験とは別の検索キーワードを用いて 10, 20, 100 の 3 種を試して経験的に設定した．

### 実験 1 の結果：ランキングの効果

実験結果を表 1 に示す．表 1 の「ランダム」は検索結果からランダムに 50 件を取り出した場合の精度，「ランキング」は，提案法でランキングした上位 50 件の精度を示す．

表 1 から分かる通り，3 つのキーワードのいずれも精度が向上している．このことから，提案法のランキングにより検索結果上位の精度を向上できることが確認できた．

### 実験 2 の結果：各種確信度の効果

実験結果を表 2, 3 に示す．表 2, 表 3 の「構文」は，構文的確信度のみ，「意味」は，意味的確信度のみ，「提案」は，構文的確信度と意味的確信度の積でランキングした結果である．表 2 は上位 50 件の精度，表 3 は上位 50 件で測定した平均適合率を示す．

表 2, 表 3 を参照すると，3 つのキーワードで最も良い結果となる確信度が一貫していないことが分かる．つまり，1) 構文的確信度，2) 意味的確信度，3) 提案法（両方の積），のどれが効果的であるかは確認できない．ただし，表 1 と比較して分かる通り，3 種のいずれもランダムの精度よりは良い結果であった．

各種確信度のランキング尺度としての良さを定量評価するために，精度を正解の出現確率とした場合の平均適合率の期待値を求めて，実測値との差を計算した．期待値と実測値の差が大きいほど良いランキングができていると考えられる．平均適合率の期待値は，以下の式で求められる [岸田 02]． $N (= 50)$  は結果結果の数， $R (50 \times \text{精度})$  は検索結果中の正解数である．

$$\frac{R - 1 + N^{-1} \sum_{i=1}^N \frac{N-R}{i}}{N - 1}$$

結果を表 5 に示す．表 5 の値は，実測値 (表 3) と期待値 (表 4) の差である．プラスの値が大きいほどランキングの効果がある．

\*1 映画は，<http://movie.goo.ne.jp/ranking/boxoffice/>，携帯電話とデジタルカメラは，<http://kakaku.com/> から取得した．

高いことを示す。マイナスの場合は、ランダムに提示したほうが良かったことを示す。

表5を見ると、「F905i」の「意味」と「クロサギ」の「構文」は、実測値と期待値の差がマイナスであるため、良いランキングができていないことがわかる。これは「意味」と「構文」は、上位の精度を向上させる傾向にはあるが、ランキングのための確信度としては適切でないことがあることを意味する。

提案法は、全キーワードでランキングの効果が見られた。このことから、ランキングのための確信度としては、構文、意味の両方を考慮する提案法が優れていると言える。

表 2: 各種確信度によるランキングの精度

キーワード	構文	意味	提案
E O S 5 D	0.28	<b>0.30</b>	<b>0.30</b>
F 9 0 5 i	<b>0.72</b>	0.66	0.66
クロサギ	0.76	0.60	<b>0.78</b>

表 3: 各種確信度によるランキングの平均適合率

キーワード	構文	意味	提案
E O S 5 D	<b>0.55</b>	0.37	0.45
F 9 0 5 i	<b>0.81</b>	0.65	0.74
クロサギ	0.73	0.78	<b>0.89</b>

表 4: 各種確信度の平均適合率の期待値

キーワード	構文	意味	提案
E O S 5 D 1	0.33	0.35	0.35
F 9 0 5 i 8	0.74	0.68	0.68
クロサギ	0.77	0.62	0.79

表 5: 各種確信度の平均適合率の期待値と実測値との差

キーワード	構文	意味	提案
E O S 5 D	+0.22	+0.02	+0.10
F 9 0 5 i	+0.07	-0.03	+0.06
クロサギ	-0.04	+0.16	+0.10

## 5. 考察

評価実験では、確信度によるランキングで、検索結果上位の精度を向上できることが確認できた。一方、構文的確信度、意味的確信度、と確信度を計算するための情報を増やしても単純には効果が上がっていないことが明らかになった。そこで、本節では、提案の構文的、意味的確信度を効果的に用いる方法を考察する。

確信度の構成要素のうち、原理的に検索キーワードに依存するのは、対象、属性の意味的確信度である。対象、属性の意味的確信度は、検索結果から対象と属性の共起確率として計算しているため、検索キーワードに依存する。

対象、属性の共起確率は、検索結果が少ない場合に、適切に推定できないため、有効でないと考えられる。一般に少量のデータから推定したデータは信頼性にかけるためである。

逆に、検索件数が多い場合は、有効と考えられる。実際に、表1, 3を見ると、検索件数が多かった「クロサギ」は、意味的確信度を考慮することで、平均適合率が大幅に向上している。

以上から、構文的、意味的確信度は、検索結果の数によって重みを変えて使用すると良いと考えられる。具体的には、検索結果が少ない場合に、意味的確信度の重みを小さくすることが

考えられる。重みを考慮する計算法としては、各確信度の重み付きの線形和が考えられる。和の場合は、各確信度のスケールや分布の違いを吸収するために、Zスコア化など、何らかの正規化が必要となると考えられる。ただし、正規化には、平均や分散が必要のため、データ数が増えると検索速度が落ちる可能性がある。今後は、検索結果の数に応じて重みを変更する方法の効果や検索速度への影響を検証していく。

## 6. まとめ

本稿では、検索精度の向上を目的に、評判情報に確信度を付与してランキングする方法を提案した。具体的には、対象、属性、評価の評判情報を、対象、属性の2項、属性、評価の2項、の2つに分けて、さらにそれぞれの構文的確信度、意味的確信度を算出して、確信度を計算する方法を提案した。構文的確信度では、1) 属性、評価の意見性判定のスコア、2) 対象物、属性の間の文字数、から、3項が対応関係にあり、さらに評判情報らしい場合に高い値になるように計算する。意味的確信度は、1) 属性、評価の共起確率、2) 対象物、属性の共起確率、から、よく共起する組み合わせを持つ評判情報ほど高い値になるように計算する。

評価実験の結果、ランキングすることで、検索結果上位の精度が向上することを確認した。また、平均適合率の期待値と実測値の差から、ランキングのための確信度として提案法が優れていることを示した。

今後は、検索キーワードの数を増やして、より厳密に有効性を確認する。また、検索結果の数で意味的確信度、構文的確信度の重みを変更する方法を試す。さらに、構文的確信度、意味的確信度を考慮できる評判抽出法を開発して、抽出精度自体を向上させる方法も検討していきたい。

## 参考文献

- [Hofmann 99] Hofmann, T.: Probabilistic Latent Semantic Indexing, in *Proc of SIGIR'99*, pp. 50–57 (1999)
- [Platt 99] Platt, J.: Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods, in *Advances in Large Margin Classifiers*, pp. 61–74 (1999)
- [岸田 02] 岸田 和明: 検索実験における評価指標としての平均精度の性質, 情報処理学会論文誌 T O D ( データベース ), Vol. 43, No. SIG02, pp. 11–26 (2002)
- [土田 07] 土田 正明, 水口 弘紀, 久寿居 大: 対象-属性-評価の3項関係同定による評判情報抽出, 言語処理学会第13回年次大会論文集, pp. 412–415 (2007)
- [土田 08] 土田 正明, 水口 弘紀, 久寿居 大: ブログからの対象、属性、評価のオンデマンド評判情報分析システム:eHyouban, 言語処理学会第14回年次大会論文集, pp. 899–902 (2008)
- [飯田 05] 飯田 龍, 小林 のぞみ, 乾 健太郎, 松本 裕治, 立石 健二, 福島 俊一: 意見抽出を目的とした機械学習による属性-評価値同定, 情報処理学会研究報告 NL165-4, pp. 21–28 (2005)