

## 主観的トリガー言語モデルによる意見情報検索

関 和 広<sup>†1</sup> 上 原 邦 昭<sup>‡2</sup>

本研究では、遠方にある語の関連性をモデル化するために考案されたトリガー言語モデルを用いて、意見情報検索を行うための枠組みを提案する。このため、まず主観的意見は主たる2つの構成要素を持つことに着目する。1つ目の構成要素は主観的意見を持つ主体、あるいは批評の対象であり、もう1つは意見を表す表現である。これをトリガーモデルにあてはめると、前者はトリガー、後者は非トリガーと考えることができる。これらの語の対（トリガー対）は、意見を記述した文章から統計的に発見することができる。本研究では、このように発見されたトリガー対を基に構築した主観的トリガー言語モデルを情報検索システムの再順位付けに利用することで、意見情報検索を実現する。実データを用いた評価実験により、本提案モデルによって意見情報検索の性能を大きく向上させることが可能であることを示す。また、トリガーモデルをクエリに動的に適応させることで、通常は向上が難しい種類のクエリについても検索性能の向上が可能であることを示す。

### Subjective Trigger Models for Opinion Retrieval

KAZUHIRO SEKI<sup>†1</sup> and KUNIYUKI UEHARA<sup>‡2</sup>

This paper presents a novel application of a statistical language model, in particular, trigger language models, to opinionated weblog (blog) post retrieval. Trigger models were originally proposed to incorporate distant word dependencies that cannot be modeled by standard  $n$ -grams, and we employ the models to capture the characteristics of written personal opinions based on an assumption that there are two constituents to form a subjective opinion; One is the subject of the opinion or the object that the opinion is about, and the other is a subjective expression. The former is regarded as a triggering word and the latter as a triggered word. We automatically identify those subjective trigger patterns to build a language model from a large number of product customer reviews. Our experiments show that, when used for reranking initial search results, our proposed model significantly improves opinionated document retrieval by over 20% in MAP. In addition, we propose dynamic adaptation of the model to a given query, which is found effective for most difficult queries categorized under politics and organizations.

### 1. はじめに

ウェブが一般に普及して以来、ホームページや掲示板、ブログ等、ユーザ側が発信する様々な形態のコンテンツ（user-generated contents または UGC）が生まれてきた。雑誌や新聞等、従来のメディアと比較したとき、これらのコンテンツに特徴的な点の1つは、一般ユーザの主観的な意見が多く含まれていることにある。これらのUGCの中でも、ブログはその簡易性から、政治や趣味等に関する私的な意見・経験を表現するための方法として、多くのユーザに利用されている。また、その利用者の増加とともに、データマイニング、自然言語処理、情報検索等の研究コミュニティから、有用な情報・知識を得るための研究対象としても注目されている<sup>1)-5)</sup>。

ブログに関連した無数の研究テーマのうち、本研究では情報検索の観点から「意見情報検索」に注目する。意見情報検索では、単にユーザの検索質問（クエリ）に関連した文書を検索する従来の情報検索とは異なり、クエリに関して主観的な意見を含んだ文書を検索することを目指す。これは、米国標準技術局が主催する Text REtrieval Conference (TREC) のブログトラックで2006年から取り組まれているテーマでもあり<sup>6),7)</sup>、近年、様々な手法が提案されている。これらの手法は、大まかに、辞書に基づく手法<sup>8)-10)</sup>と分類に基づく手法<sup>11)-13)</sup>に分けることができる。前者は、人手で、あるいは自動で作成した主観的表現（たとえば「like」や「fantastic」）のリストを用い、それらの語が現れる文書は主観的な文書であると見なす。一方、後者は、正例（意見を含む文書）と負例（意見を含まない文書）を基に機械学習の手法によって分類器を作成し、意見文書を同定する。

本稿では、どちらのカテゴリにも属さない、新しい意見情報検索の手法を提案する。本提案手法は、代名詞あるいは主観的表現と批評の対象との文書中での距離を考慮することで、意見情報検索性能が向上する<sup>14)</sup>という経験的発見に部分的に着想を得たものである。より具体的には、統計的言語モデルの1つであるトリガーモデル<sup>15),16)</sup>を利用することで、主観的意見の記述に現れる前述のような特徴的なパターンをとらえる。トリガーモデルは、遠距離にある語間の依存性を扱うために提案された言語モデルである。本提案手法の主要な仮

<sup>†1</sup> 神戸大学自然科学系先端融合研究環

Organization of Advanced Science and Technology, Kobe University

<sup>‡2</sup> 神戸大学大学院工学研究科

Graduate School of Engineering, Kobe University

定は、主観的意見には2つの重要な構成要素があることである。1つは意見の主体あるいは対象であり、もう1つは主観的表現である。提案手法では、前者をトリガー、後者を非トリガーと見なし、amazon.comから収集した顧客レビューから自動的に両者の対(トリガー対)を同定する。このように主観的言語モデルを構築した後、TREC ブログトラックのテストコレクションを用いた種々の実験により、本提案モデルは意見情報検索の精度を顕著に向上させ、また、所与のクエリにモデルを動的に適応させることで、多くのクエリに関してさらなる精度の向上が可能であることを示す。

以下、2章では、主観的言語モデルの構築法について詳述する。続いて、3章では、構築された言語モデルをモデルの良さという見地から検証し、さらに主観的言語モデルを意見情報検索に利用したときの効果について議論する。4章では、関連研究についてまとめ、最後に、5章でまとめと今後の課題について述べる。

## 2. 意見情報検索のための主観的言語モデル

### 2.1 動機

ある文書が主観的意見を含むかどうかを判断する最も単純で直感的な方法は、文書内に主観的表現が存在するかどうかに着目することである。このような語彙に基づく方法は、主観的意見を述べる際によく使われる語を含む文書は主観的であるという仮定に基づく。たとえば、「like」は好意的な感情を示す語であると考えられ、この語(や類似の語)を含む文章は主観的意見を含む文章である可能性があるといえる。この考えに基づき、これまでの多くの研究者が、人手あるいは自動で意見表現辞書を作成し、これに基づいて意見情報検索を行ってきた<sup>8),9),17)</sup>。これらの手法はある程度、効果的ではあるものの、1つの問題点がある。それは、前述の直感に反し、主観的表現を含む文書は必ずしも主観的文書ではないということである。たとえば、「It looks like a cat.」や「She likes singing.」は主観的意見ではなく、むしろ客観的事実を示す。このような微妙な意味的差異を理解するためには、「like」だけに注目するのではなく、より広い文脈を考慮する必要がある。

より広い文脈を考慮する1つの方法は、統計的言語モデルとして広く用いられている  $n$  グラム言語モデル<sup>18)</sup> を利用することである。 $n$  グラムモデルは局所的な文脈を考慮し、接続する  $n$  語をひとまとまりにして解析する。たとえば、上の例文「It looks like a cat.」の2グラムは、「It looks」、「looks like」、「like a」、および「a cat」であり、この場合、「like」はそれぞれ「looks」、「a」と接続の文脈で解釈される。大きな  $n$  を用いれば、より広い文脈が考慮できるものの、単純に  $n$  を増加させるとデータ過疎の問題が生じるため、実際に

は  $n$  は2~3程度で用いられることが多い。

本研究では、意見情報検索を実現することを目的とし、主観的意見に特徴的な語の依存性をモデル化するためにトリガーモデルの利用を提案する。

### 2.2 主観的トリガーモデル

その単純さにもかかわらず、 $n$  グラム言語モデルはこれまで多くの自然言語処理タスクに適用され、その有用性が示されてきた。しかしながら、 $n$  で制限される局所的な文脈を飛び越える語の依存性が存在することは明らかであり、そのような依存性を考慮するため、Lau ら<sup>15)</sup> はトリガーモデルを提案した。「トリガー」とは他の語の生起を誘発する語のことであり、トリガーと依存関係を持って生起する語を「非トリガー」と呼ぶ。また、これらの語の対を「トリガー対」と呼ぶ。このようなトリガー対の依存関係を表したトリガーモデル  $P_T(w|h)$  は、通常、ベースラインの  $n$  グラム言語モデル  $P_B(w|h)$  と線形結合して用いられる。

$$P_E(w|h) = (1 - \lambda) \cdot P_B(w|h) + \lambda \cdot P_T(w|h) \quad (1)$$

ここで、 $w$  と  $h$  は、それぞれ単語と履歴( $w$  に先行する語)を示し、 $\lambda$  は両言語モデルの影響を制御するパラメータである。なお、 $P_T(w|h)$  の詳細な定義については、2.3節の最後で述べる。

トリガーモデルを構築するためには、まず最初に(意味のある)トリガー対を同定する必要がある。あるコーパス(文書集合)が与えられたとき、コーパス中に現れるすべての語の組がトリガー対になりえ、これらをすべて考慮すると、多くの無意味なトリガー対が同定されてしまう。そこで、Tillmann ら<sup>16)</sup> は、 $n$  グラム言語モデルにおける語  $w$  の生起確率  $P(w|h)$  がある閾値  $t$  より小さいときだけ  $w$  を非トリガー候補とする基準を提案した。具体的には、

$$P(w|h) < t \quad (2)$$

を満足する語  $b$  に関して、コーパスに現れるすべての語  $a$  とのトリガー対「 $a \rightarrow b$ 」の有用性を対数尤度差  $\Delta_{a \rightarrow b}$  によって評価する。 $\Delta_{a \rightarrow b}$  は、 $n$  グラム言語モデル  $P(\cdot)$  と当該トリガー対「 $a \rightarrow b$ 」のみによって拡張したトリガーモデル  $P_{E:a \rightarrow b}(\cdot)$  の対数の差であり、入力文書を語の系列  $w_1, w_2, \dots, w_m$  で表したとき、次式で算出される。

$$\begin{aligned} \Delta_{a \rightarrow b} &= \log P_{E:a \rightarrow b}(w_1, w_2, \dots, w_m) - \log P(w_1, w_2, \dots, w_m) \\ &\approx \sum_i \log (P_{E:a \rightarrow b}(w_i|h_i) - P(w_i|h_i)) \end{aligned} \quad (3)$$

トリガー対を考慮した拡張モデルがより良い(大きい)推定値を与えるほど、対数尤度差は

大きな値となる。このようにしてすべての語の組について対数尤度差を評価したのち、差が大きい組から順に任意の数のトリガー対を採用することで、最終的なトリガーモデル  $P_T(\cdot)$  を得ることができる。この基準は「低レベルトリガー」と呼ばれる。Tillmann らの実験によれば、この基準によって、「neither→nor」や「tip→iceberg」といったトリガー対が発見されている。

本研究では、2種類の仮定を導入し、さらに後述の若干の修正を加えた低レベルトリガーを用いることで、主観的意見に特徴的な語の依存関係をとらえる。1つ目の(主要な)仮定は、主観的意見は2つの基本的な構成要素、すなわち意見の主体(たとえば「I」)あるいは意見の対象(たとえば「this movie」)と主観的表現(たとえば「like」や「best」)を持つことである。そして本提案研究では、前者をトリガー、後者を非トリガーとして扱う。2つ目の仮定は、トリガーは多くの場合、代名詞として表出することである。これらの仮定は、代名詞(たとえば「I」、「you」、「me」)、主観的表現(たとえば「like」、「feel」)と意見の対象との共起が、意見らしさ(opinionatedness)を示す効果的な手がかりであるという経験的事実によっている<sup>14),19)</sup>。以上の仮定を導入することにより、たとえば、「I really like this movie.」という意見文からは「I→really」、「I→like」といったトリガー対が抽出できる。上述の関連研究の報告から、これらのトリガー対は意見文章の同定に有用であると考えられる。

従来の研究で用いられてきたアドホックな経験的事実(上述)と比較し、本提案モデルは、意見らしさを示す語間の依存関係をより理論的な枠組みによってモデルに取り込むことができる。また、少数の代名詞のみを潜在的なトリガーとして扱うことで、より効果的および効率的に主観的意見に特化した言語モデルを構築することができる。以下では、このような主観的トリガーモデルによって拡張された提案言語モデル  $P_E(\cdot)$  を「主観的言語モデル」と呼ぶ。

### 2.3 主観的言語モデルの構築

前節で概説した手順と仮定によって、本節では主観的トリガーモデルを以下のように構築する。まず、主観的意見に特徴的なトリガー対を同定するため、主観的意見を含むコーパスを構築する。本研究では、amazon.com から5,000の顧客レビューを自動的に収集し、実験に利用した。レビューを構成する文の平均長(語数)は19.5語であった。これらのレビューは、アマゾンで販売される様々な製品、たとえば本、DVD、電気製品、玩具等を対象に書かれたものである。なお、個々のレビューには、批評文のほかに1から5の評価値が与えられているが、本研究では肯定・否定等の意見の極性について関知しないため、利用

表1 対数尤度差が最大の低レベルトリガー対  
Table 1 Most prominent low level triggers.

トリガー (a)	非トリガー (b)	$\Delta_{a \rightarrow b}$
this	→ the	7.079
it	→ the	7.079
i	→ the	7.079
i	→ to	6.526
this	→ to	6.525
my	→ and	6.502
i	→ and	6.501
this	→ and	6.498
it	→ and	6.497
this	→ a	6.381
	...	

していない。また、式(2)の  $P(w|h)$  の推定には(平滑化していない)3グラムモデルを用いた。潜在的なトリガーとして、本研究では実験的に次の14の代名詞を用いた。

I, my, you, it, its, he, his, she, her, we, our, they, their, this

なお、これらの代名詞の選択は著者の恣意的な判断によるため、最適な選択方法に関しては検討の余地がある。

これらの設定のもと、低レベルトリガーを用いて(最大)10,000のトリガー対を自動的に同定した。なお、モデル構築の際、履歴  $h$  は同一文内の先行語とした。表1に、対数尤度差が最大のトリガー対を一部示す。

表1が示すように、非トリガーは冠詞や前置詞等の機能語が多数を占め、意見を含む文書の特徴付けるために有効だと思われるトリガー対はほとんど同定されなかった。この問題は、機能語は多くの文脈に現れるため、ときに  $w$  の生起確率  $P(w|h)$  が閾値  $t$  を下回る可能性があるために生じる(式(2)参照)。もう1つの問題は、低レベルトリガーは履歴  $h$  の頻度  $\text{Freq}(h)$  を考慮していないことにある。一般的に、 $\text{Freq}(h)$  が小さい場合、推定値  $P(w|h)$  の信頼性は低い。これらの要素を考慮するため、本研究では低レベルトリガーの条件式を次のように修正する。

$$\tau \cdot P(w_i|h_i) < t \quad (4)$$

ここで、 $\tau$  は  $w_i$  と  $h_i$  の頻度の比、すなわち  $\text{Freq}(w_i)/\text{Freq}(h_i)$  と定義する。この修正によって、低頻度の履歴  $h_i$  を持つ高頻度の語  $w_i$  にはペナルティが課され、(主に)機能語が非トリガーとして同定されにくくなる。他の方法として、機能語のリストをあらかじめ作

表 2 修正後の基準を用いたときの対数尤度差が最大の低レベルトリガー対  
Table 2 Most prominent triggers identified by the modified criterion.

トリガー (a)	非トリガー (b)	$\Delta_{a \rightarrow b}$
i	→ wish	5.113
i	→ felt	5.073
i	→ loved	4.862
i	→ hope	4.739
i	→ couldn	4.680
i	→ got	4.611
i	→ cannot	4.593
this	→ an	4.578
this	→ all	4.575
i	→ liked	4.552
i	→ enjoyed	4.531
	...	

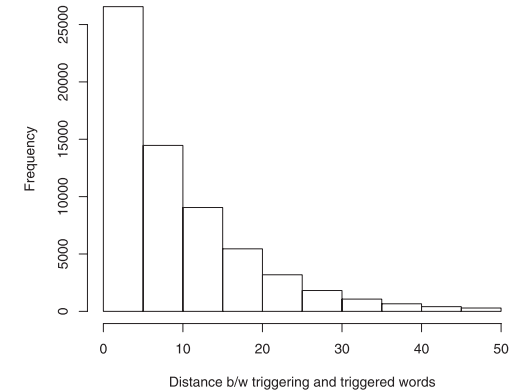


図 1 発見したトリガー対のコーパス内での距離 (語数) と頻度

Fig. 1 Histograms of the distances (the number of words) between identified trigger pairs.

成することで機能語を完全に排除することも可能である。しかしながら、後者の方法では、機能語でありながら真に意味のあるトリガー対まで排除してしまう。たとえば、機能語をすべて排除した場合、「this is the best choice」における「this → the」や「this book is all you need」における「this → all」のようなトリガー対は抽出不可能となってしまう。前者は最上級を特徴付け、後者も何かを推薦する際にしばしば用いられる慣用句を特徴付ける有用なトリガー対であると考えられる。

表 2 に、修正した基準によって同定されたトリガー対 (計 2,012 個) を対数尤度差の大きい順に示す。先ほどとは異なり、今回は主観的意見の特徴づけるようなトリガー対が同定できていることが確認できる。さらに、前述の例のような機能語を含むトリガー対も発見できていることが分かる。表には記載しきれなかったものの、同定されたトリガー対を見ると、「I → very」や「it → greatest」のように、遠距離の語の依存関係も抽出できていた。実際に、同定した 2,012 のトリガー対がコーパス内でどの程度の距離 (語数) にあったのかを集計したヒストグラムを図 1 に示す。この結果から、 $n$  グラム言語モデルではとらえにくい遠方の語間の関係が獲得できていることが分かる。

なお、構文解析器を用いても類似のトリガー対を同定することができると考えられるものの、本提案手法には 2 つの利点がある。1 つは、意見の記述に関して本研究と同様の仮定をおける場合、本手法は形態素解析・構文解析を行う必要がないため、他の言語にも容易に適用できる点である。第 2 に、本手法は係り受け関係を考慮しないため、直接係り受け関係がないながらも特徴的なトリガー対を発見することができる点である。たとえば、上述の実験

では、「I thought the writing is fantastic」等における「I → fantastic」をトリガー対として同定することができた。

続いて、同定されたそれぞれのトリガー対 ( $a \rightarrow b$ ) について、それらの関連度  $\alpha(b|a)$  を同一文脈内における語  $a$  と  $b$  の共起関係に基づいて最尤推定法により算出する。また、この関連度を基に、トリガーモデルを次のように構築する。

$$P_T(w|h) = \frac{1}{|h|} \sum_{w_j \in h} \alpha(w|w_j) \quad (5)$$

ここで、 $w_j$  は  $w$  の履歴に含まれる語である。式 (5) は、 $P_T(w|h)$  が履歴中のすべての語  $w_j$  に関する関連度  $\alpha(w|w_j)$  の平均として定義されることを意味している (トリガーモデルのさらに詳細な推定方法については、Tillmann ら<sup>16)</sup> の論文を参照)。

ベースラインの言語モデル  $P_B(w|h)$  としては、平滑化したバックオフ 3 グラムモデルを用いた。また、式 (1) における補間係数  $\lambda$  は経験的に 0.9 として、主観的言語モデルを構築した。

#### 2.4 モデル適応

主観的言語モデルはアマゾンの顧客レビューから構築されたため、このモデルは商品以外の対象 (たとえば人物) に関する意見を同定するためには必ずしも効果的ではない可能性がある。この潜在的な問題点を克服するため、本研究では、初期検索で得られたブログ記事か

ら新たにトリガー対を発見することで、トリガーモデルをクエリにより適応させる試みを行う。

この考えは擬似適合性フィードバック (pseudo-relevance feedback または PRF)<sup>20)-22)</sup> に類似している。PRF は、ユーザクエリを初期検索の結果上位の  $k$  文書から抽出した関連語によって拡張する手法である。PRF と本研究のモデル適応との重要な違いは、後者はクエリを修正するのではなく、言語モデルを更新することで検索された文書の意見らしさをより適切に推定することにある。よって、PRF とモデル適応はまったく独立の処理であり、必要であれば両者を個別に適用することができる。

以下に、本研究で提案するモデル適応の手順を示す。

- (1) 所与のクエリ  $q$  について、任意の検索モデルで初期検索を行う。
- (2) 上位  $k$  件のブログ記事を利用して、(新しい) トリガー対  $a \rightarrow b$  を動的に発見する。また、これらのトリガー対について、関連度  $\alpha'(b|a)$  を 2.2 節と同様に算出する。なお、このステップでは、2.3 節で述べた 14 の代名詞に加えて、与えられたクエリ  $q$  を潜在的トリガーとして与えることもできる。
- (3) トリガーモデル  $P_T(\cdot)$  を構築する。このとき、アマゾンの顧客レビューから得た関連度  $\alpha(b|a)$ 、あるいは前ステップで得た関連度  $\alpha'(b|a)$  のうち、より大きい方の値を用いる。すなわち、式 (5) を次式 (6) で置き換える。

$$P_T(w|h) = \frac{1}{|h|} \sum_{w_j \in h} \max [\alpha(w|w_j), \alpha'(w|w_j)] \quad (6)$$

この適応により、所与のクエリに関する初期検索で上位に現れたブログ記事中で顕著なトリガー対を発見し、クエリにより特化した表現を含む主観的言語モデルを構築することができるものと期待される。

### 3. 実験による評価

#### 3.1 データ

提案手法の有効性を評価するため、TREC ブログトラック 2006 のテストコレクション<sup>7)</sup> を用いて評価実験を行った。このコレクションは、2005 年の 12 月から 2006 年の 2 月までの 11 週間の間に収集された 3.2 百万のブログ記事から構成され、商用のブログ検索エンジンのクエリログから作成された 50 個の情報要求 (トピック) を持つ。図 2 にトピックの例を示す。図中の「title」が検索エンジンのユーザによる実際のクエリを示している。

Topic #	851
Title	March of the Penguins
Desc.	Provide opinion of the film documentary "March of the Penguins".
Narr.	Relevant documents should include opinions concerning the film documentary "March of the Penguins". Articles or comments about penguins outside the context of this film documentary are not relevant.

図 2 TREC 2006 ブログトラックで用いられたトピックの例  
Fig. 2 Example topic from the TREC 2006 Blog Track.

さらに、任意の情報検索システムの評価を可能にするため、それぞれのトピックについて、適合・非適合の記事が明示されている。適合性の評価は、次の 5 つのカテゴリに基づいて行われている。不適合 (ラベル 0)、適合かつ意見を含まない (1)、適合かつ否定的意見のみを含む (2)、適合かつ肯定的・否定的意見を含む (3)、適合かつ肯定的意見のみを含む (4)。通常の情報検索システムの評価では、ラベル 1~4 は区別されず、すべて「適合」文書として扱われる。これに対し、意見情報検索の文脈では、ラベル 2~4 だけが適合文書と見なされる。

このテストコレクションを用い、2 つの観点から提案手法を評価する。まず、3.2 節で、意見情報検索とは無関係に言語モデルの「良さ」そのものを評価する。続いて、3.3 節で、意見情報検索における提案モデルの有効性を評価し、さらにモデル適応の効果を評価する。

#### 3.2 言語モデルの評価

2.3 節で構築した主観的トリガーモデルは、アマゾンの顧客レビューから生成されたため、製品以外の多様な対象に関する意見の特徴を反映しているとはいえない。そこで、本節では、まずアマゾンのレビューから構築された主観的言語モデルが、実際の主観的なブログ記事を適切に表現することができるか否かを評価する。このため、ここでは言語モデルの評価に一般的に用いられるパープレキシティ<sup>23)</sup> という尺度を利用する。直感的には、パープレキシティとは、ある言語モデルを語の系列 (文書) の予測に用いた際、どの程度の不確実性が残るのかを示す指標である。よって、一般的に、低いパープレキシティを持つモデルほど良い (より正確な予測が行える) 言語モデルだということができる。より正確には、 $H(L, d)$  を文書  $d$  における言語モデル  $L$  のクロスエントロピーとしたとき、パープレキシティは  $2^{H(L, d)}$  と定義される。クロスエントロピーは、真の確率分布と推定された確率分布の間の情報理論的な距離であり、大きな  $m$  に関して式 (7) が成り立つ。

### 32 主観的トリガー言語モデルによる意見情報検索

表 3 パープレキシティの算出値、括弧内の数字は対応する  $P_B$  と比較したときのパープレキシティの減少率（性能向上）を示す

Table 3 Perplexity results. Figures in parentheses indicate percent decrease of perplexity as compared to corresponding  $P_B$ .

$n$	非意見 ( $d_N$ )		意見 ( $d_O$ )	
	$P_B$	$P_E$	$P_B$	$P_E$
1 グラム	9369	8946 (-4.5%)	7198	6829 (-5.1%)
2 グラム	6526	6279 (-3.8%)	4749	4546 (-4.3%)
3 グラム	5998	5762 (-3.9%)	4337	4145 (-4.4%)

$$\begin{aligned}
 H(L, d) &\approx -\frac{1}{m} \log P(w_1 \dots w_m) \\
 &\approx -\frac{1}{m} \sum_{i=1}^m \log P(w_i | h_i)
 \end{aligned}
 \tag{7}$$

この指標を用いて言語モデルの特性を評価するため、ブログトラックのテストコレクションのうち、ラベル 2~4 が付与されたブログ記事（適合かつ意見を含むすべてのブログ記事）を連結し、仮想的に長大な意見文書  $d_O$  を作成し、同様にラベル 1 が付与されたブログ記事すべてを連結することで、長大な非意見文書  $d_N$  を作成した。表 3 に、これら 2 つの仮想文書上でのベースライン言語モデル  $P_B$  と主観的言語モデル  $P_E$  のパープレキシティを異なる  $n$  ごとに算出した結果を示す。

この結果から、3 つの重要な事実が観測できる。まず、高次の  $n$  ほど言語モデルにも文書にもよらずパープレキシティが単調に低下し、少なくとも 3 までは  $n$  が高いほどより良く意見記事を表現できていることが分かる。次に、意見文書  $d_O$  は非意見文書  $d_N$  よりも低いパープレキシティを示しており、アマゾンのレビューから構築した言語モデルでも、意見を含むブログ記事の特徴をある程度反映できていることが分かる。最後に、主観的言語モデル  $P_E$  はベースライン言語モデル  $P_B$  よりも低いパープレキシティを示しており、その低下の幅は意見文書  $d_O$  でより大きい。この観測は、通常の  $n$  グラム言語モデルではとらえられない意見文書の特徴を主観的トリガー対がとらえたことを示しており、とりわけ重要である。

上記の実験は主観的言語モデルが（長大な）意見文書と非意見文書を見分けるために有効である事実を示している。続いて、この観測が個別のブログ記事においても成り立つのかを検証するため、個々の意見記事、非意見記事における  $P_E$  のクロスエントロピーを算出し、それらの分布を比較した。なお、この実験では  $n = 3$  とした。図 3 の結果より、提案モデルを用いた場合、意見記事 (opinion) のクロスエントロピーの分布は非意見記事 (non-opinion)

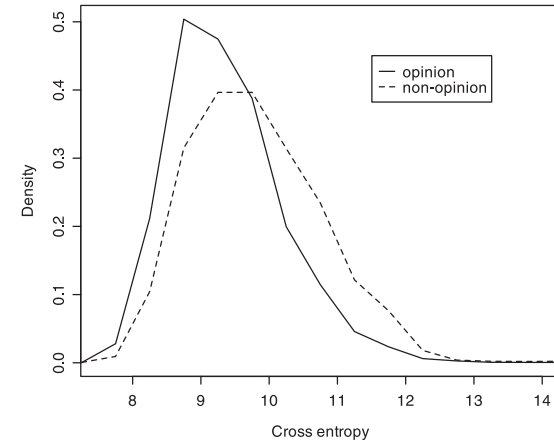


図 3 意見ブログ記事と非意見ブログ記事におけるクロスエントロピーの分布  
Fig. 3 Distributions of cross entropy for opinionated and non-opinionated blog posts.

のそれよりも、概して低い値をとることが確認できる。

次節では、主観的言語モデルを検索結果の再順位付けに用いた意見情報検索の評価実験について報告する。

### 3.3 主観的言語モデルを用いた意見情報検索

#### 3.3.1 初期検索

この実験では、一般的な検索モデルによって得られた初期検索結果を提案モデルによって再順位付けすることで、意見情報検索における提案モデルの有効性を評価する。初期検索には、TFIDF 単語重み付けを用いたベクトル空間モデル<sup>24)</sup> (VSM) と言語モデルを用いた推論ネットワークモデル<sup>25)</sup> (INM) を採用した。両者とも、索引付けの際は禁止語 (stopword) を除去し、大・小文字の区別および接辞の除去 (stemming) は行っていない。クエリとしては、トピックの title のみを用いた (図 2 参照)。また前述のように、肯定・否定の区別はせず、ラベル 2~4 の記事すべてを適合文書と見なした (3.1 節参照)。表 4 に、初期検索の結果と TREC ブログトラック 2006 参加者の公式結果を MAP (mean average precision) で示す。また、ブログトラック以降に報告された代表的な結果も参考のために示した。なお、TREC の公式結果および TREC 以降に報告された結果は、意見情報検索のための様々な機能を用いて得られた結果であるのに対し、VSM と INM はそのような機能を用いてい

表 4 一般的な情報検索モデルを用いた初期検索の結果, TREC の公式結果と TREC 以降の代表的な先行研究の結果も示す

Table 4 Initial search results using alternative IR models, where TREC 2006 official results and other post-TREC results are shown for reference.

手法		MAP
初期検索	VSM	0.1126
	INM	0.1965
TREC	Best	0.1885
	Median	0.1156
	Worst	0.0000
TREC 以降	W. Zhang ら <sup>13)</sup>	0.2726
	M. Zhang と Ye <sup>10)</sup>	0.2257

ない単なる初期検索の結果であることに注意する必要がある。

VSM と INM の性能には大きな開きがあり, INM に関しては TREC の公式結果をも上回る結果が得られた. この結果から, 以降の実験では, INM を初期検索に用いた場合のみを報告する\*1.

### 3.3.2 主観的言語モデルと検索モデルの統合

INM による初期検索では, 検索されたそれぞれのブログ記事  $d$  に関して,  $d$  がユーザのクエリ  $q$  に適合する確率  $P(q|d)$  が推定され, 検索結果の順位付けに用いられている. 所与の  $d$  が主観的意見を含むか否かは,  $d$  が  $q$  に適合するか否かと独立であると仮定すると,  $d$  が適合記事でありかつ意見を含む確率は,  $P(q|d)$  と  $P_E(d) \approx \prod_{i=1}^m P_E(w_i|h_i)$  の積で表せる. ここで  $m$  は  $d$  の長さ (単語数) である. しかしながら, 定義より, 長い文書はより小さい  $P_E(d)$  を持つ傾向にあり, また, 2つの確率分布は大きく異なる分散を持つ可能性があるため, 単純に積をとるだけでは一般的にうまく機能しない. そこで, 両者の重み付き対数和をとり, かつ  $P_E(d)$  を文書の長さ  $m$  で正規化することで, 初期検索の再順位付けに用いる最終的なスコア  $\text{Scr}(d, q)$  を得る.

$$\text{Scr}(d, q) = (1 - \beta) \log P(q|d) + \frac{\beta}{m} \sum_{i=1}^m \log P_E(w_i|h_i) \quad (8)$$

ここで  $\beta$  は主観的言語モデルの影響を制御するパラメータである. なお, 上式の第 2 項は式 (7) のクロスエントロピーに相当する.

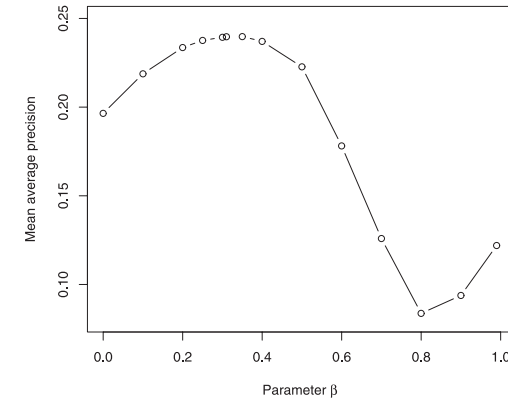


図 4 意見情報検索におけるパラメータ  $\beta$  と MAP 値の関係  
Fig. 4 Relation between parameter  $\beta$  and MAP.

式 (8) の  $\beta$  を 0 から 1 まで徐々に増加させながら, 初期検索の結果を再順位付けしたときの MAP 値の変化を観測した結果を図 4 に示す. ここで, 一番左 ( $\beta = 0$ ) の結果が初期検索に対応する.  $\beta$  を変化させることで, MAP 値は顕著に向上し,  $\beta = 0.35$  のとき 0.2398 (+22.0%) の性能が得られた. この結果は,  $\beta$  が適切に選択できれば, 式 (8) による主観的言語モデルの統合が意見を含むブログ記事の同定に効果的であることを示す. さらに, 翌年のプログトラック 2007 で用意された別の 50 のトピックについて同じ実験を行ったところ, 初期検索で得られた MAP 値 0.2508 が 0.3072 (+22.5%) まで向上し, 同様の結果が得られた. この実験でも  $\beta$  の最適値は 0.35 であり, 異なるトピック集合に関しても最適な  $\beta$  が安定しているという好ましい特性が確認できた.

なお, ここでの結果は表 4 に示した関連研究の結果とは直接比較できないことに注意を要する. これは, それぞれの関連研究ごとに初期検索の結果が異なることによる. 初期検索の性能は最終的な意見情報検索の性能に大きな影響を与えることが報告されており<sup>6),26)</sup>, 同じ検索システムを初期検索に用いない限り公平な比較を行うことはできない. これを示すための試みとして, MAP 値で 0.3022 を得たより高性能な初期検索結果に主観的言語モデルを適応した\*2. この初期検索結果に関しても, 本提案手法は MAP 値を 0.3221 (+6.6%)

\*1 この結果は必ずしも INM が VSM よりも優れていることを示すわけではない. 事実, W. Zhang ら<sup>13)</sup> らは, VSM を利用して 0.2726 の MAP 値を得ている.

\*2 この初期検索結果は, 意見情報検索のより公平な比較のため, プログトラック 2008 の参加者に配布されたものであり, 会議終了後, 一般にも公開される. 本稿執筆時点では, まだこの検索システムの詳細は明らかにされていない.

まで向上させることができた。

### 3.3.3 個々のクエリに関する結果分析

前項では、主観的言語モデルを初期検索結果の再順位付けの枠組みに統合することで、MAP 値で 22.0%の顕著な性能向上が観測された。このように、全体的な性能向上は明らかになったものの、主観的言語モデルが個々のクエリに対してどのように作用したのかは明らかではない。そこで本項では、個々のクエリごとの結果について詳細な考察を加える。まず、表 5 に個々の結果を平均精度 (average precision) で示す。ここで、「初期検索」の列は初期検索の結果を示し、「トリガー」の列は主観的言語モデルを用いて再順位付けしたときの結果を示す。パラメタ  $\beta$  は、クエリにかかわらず前項で得られた全体的な最適値 (0.35) に固定した。

表より、0.1 ポイント以上の顕著な向上が見られたトピックは、macbook pro (+0.2227), mardi gras (+0.1141), heineken (+0.1977), shimano (+0.1929), zyrtec (+0.1211), board chess (+0.1333) であった。一方、わずかに精度低下が見られたトピックは、ann coulter (-0.0099), cindy sheehan (-0.0327), sonic food industry (-0.0136), west wing (-0.0091), world trade organization (-0.0453), business intelligence resources (-0.0002) であった。最後に、jim moran (-0.1490) については比較的大きな精度低下が見られた。これらのトピック間に何らかの共通の特徴が存在するかを調査するため、それぞれのトピックについて概略 (多くは Wikipedia からの抜粋) を述べる。まず、顕著な向上が見られたトピックについてまとめる。

- MacBook Pro (#856) は、アップルが開発、販売するノートブック型 Macintosh のシリーズの 1 つ。
- Mardi Gras (#861) は、フランス語で「肥沃な火曜日」の意で、謝肉祭の最終日、灰の水曜日の前日を意味する。
- Heineken (#883) オランダのビール醸造会社およびブランド名。
- Shimano (#885) は大阪府堺市に本社を置くアウトドアスポーツ用品会社。
- Zyrtec (#893) はアレルギー疾患および皮膚疾患の治療薬 (商品名)。
- Board chess (#894) は 2 人で行うボードゲームの一種。

これらのトピックは、Mardi Gras と board chess を除いて「製品」だと考えることができる (Shimano は会社名であるものの、Shimano の製品を呼ぶ場合にもよく用いられる)。本研究で主観的言語モデルの構築に用いたアマゾンのレビューは必ずしもこれらの製品、たとえばビールや薬品を対象にしたものではないにもかかわらず、これらの意見記事検索に

表 5 個々のクエリにおける初期検索と再順位付け後の平均精度

Table 5 Individual results in average precision comparing initial retrieval and one after reranking.

トピック ID	クエリ	初期検索	トリガー	差	向上率
851	march of the penguins	0.3089	0.3704	+0.0615	19.9%
852	larry summers	0.1797	0.2241	+0.0444	24.7%
853	state of the union	0.0222	0.0286	+0.0064	28.8%
854	ann coulter	0.4690	0.4591	-0.0099	-2.1%
855	abramoff bush	0.3652	0.3783	+0.0131	3.6%
856	macbook pro	0.2327	0.4554	+0.2227	95.7%
857	jon stewart	0.3431	0.4000	+0.0569	16.6%
858	super bowl ads	0.1368	0.1757	+0.0389	28.4%
859	letting india into the club?	0.0013	0.0016	+0.0003	23.1%
860	arrested development	0.2531	0.2692	+0.0161	6.4%
861	mardi gras	0.2166	0.3307	+0.1141	52.7%
862	blackberry	0.0506	0.0715	+0.0209	41.3%
863	netflix	0.5419	0.6135	+0.0716	13.2%
864	colbert report	0.3084	0.3458	+0.0374	12.1%
865	basque	0.1361	0.2061	+0.0700	51.4%
866	whole foods	0.1596	0.1681	+0.0085	5.3%
867	cheney hunting	0.3073	0.3265	+0.0192	6.2%
868	joint strike fighter	0.1129	0.1579	+0.0450	39.9%
869	muhammad cartoon	0.1431	0.1802	+0.0371	25.9%
870	barry bonds	0.2208	0.2402	+0.0194	8.8%
871	cindy sheehan	0.4903	0.4576	-0.0327	-6.7%
872	brokeback mountain	0.3198	0.4016	+0.0818	25.6%
873	bruce bartlett	0.2412	0.3119	+0.0707	29.3%
874	coretta scott king	0.1580	0.1874	+0.0294	18.6%
875	american idol	0.2055	0.2625	+0.0570	27.7%
876	life on mars	0.0100	0.0222	+0.0122	122.0%
877	sonic food industry	0.0383	0.0380	-0.0003	-0.8%
878	jihad	0.0574	0.1182	+0.0608	105.9%
879	hybrid car	0.0972	0.1076	+0.0104	10.7%
880	natalie portman	0.2072	0.2935	+0.0863	41.7%
881	fox news report	0.0426	0.1338	+0.0912	214.1%
882	seahawks	0.0543	0.0651	+0.0108	19.9%
883	heineken	0.1799	0.3776	+0.1977	109.9%
884	qualcomm	0.3134	0.3312	+0.0178	5.7%
885	shimano	0.1182	0.3111	+0.1929	163.2%
886	west wing	0.2498	0.2407	-0.0091	-3.6%
887	world trade organization	0.1111	0.0658	-0.0453	-40.8%
888	audi	0.6640	0.7318	+0.0678	10.2%
889	scientology	0.1918	0.2294	+0.0376	19.6%
890	olympics	0.1020	0.1575	+0.0555	54.4%
891	intel	0.0535	0.0578	+0.0043	8.0%
892	jim moran	0.6218	0.4728	-0.1490	-24.0%
893	zyrtec	0.0976	0.2187	+0.1211	124.1%
894	board chess	0.1256	0.2589	+0.1333	106.1%
895	oprah	0.2127	0.2541	+0.0414	19.5%
896	global warming	0.1142	0.1427	+0.0285	25.0%
897	ariel sharon	0.0818	0.1177	+0.0359	43.9%
898	business intelligence resources	0.0008	0.0006	-0.0002	-25.0%
899	cholesterol	0.0418	0.0746	+0.0328	78.5%
900	mcdonalds	0.1137	0.1431	+0.0294	25.9%
全体		0.1965	0.2398	+0.0433	22.0%



も有用であることが示された。この結果は、これらのレビューから獲得した言語モデルが Shimano のようなブランド名を含む製品全般、および Mardi Gras, board chess といったある種の非製品に関しても一般化可能であることを示唆する。

続いて、システム性能が少しでも低下したトピックに関してまとめる。

- Ann Coulter (#854) はアメリカの評論家。コーネル大学卒、ミシガン大学法科卒で弁護士資格を持つ。保守系のコメンテーターでテレビに多数出演していることで知られる。
- Cindy Sheehan (#871) はアメリカの反戦活動家。息子をイラク戦争でなくした。
- West Wing (#886) はアメリカの放送局 NBC で 1999 年から 2006 年にかけて放送されたドラマで、ホワイトハウスを舞台に民主党のバートレットアメリカ合衆国大統領とその側近達を中心に描いた政治ドラマである。
- Jim Moran (#892) は民主党に所属するバージニアの下院議員である。
- Sonic food industry (#877) はアメリカのファーストフードレストランのチェーン店。
- World Trade Organization (#887) は自由貿易促進を主たる目的として作られた国際機関。
- Business intelligence resources (#898) は何らかの固有表現ではなく、ビジネスインテリジェンスに有用な資源のこと。

これらの概略から判断すると、最初の 4 つは「政治」に関連しており、次の 2 つは「組織」に関連していることが分かる。Jim Moran 以外の精度の低下はわずかであるものの、これらのカテゴリに属するトピックに関する意見は、本研究で提案した主観的言語モデルでは適切に表現できていないものと考えられる。なお、最後のクエリは非常に曖昧であり、初期検索でも平均精度が 0.0008 ときわめて低い。いい換えると、このクエリに関しては、そもそも初期検索の時点で意見文書が検索結果にほとんど含まれていないため、主観的言語モデルの有用性検証にはあまりふさわしくない。

次に、表 5 の 50 のトピックを「製品」、「政治」、「組織」という 3 つのカテゴリに分類し、それぞれのカテゴリごとに精度の変化を調査した。なお、いずれのカテゴリにも属さないトピックは考慮していない。その結果、製品では 46.8% の向上、政治では 7.6% の向上、組織では 3.1% の低下が見られた。以上の結果を総合すると、政治や組織等に関して主観的意見を表現するために用いられる語彙は、製品に関するそれとは異なるものと考えられる。なお、これらの比較的「難しい」カテゴリに属するトピックであっても、本研究で構築した主観的言語モデルによって意見情報検索精度向上が見られた事例が多くあったことも注目すべきで

表 6 初期検索と主観的言語モデルによる再順位付け後の結果とモデル適応後の結果との比較

Table 6 Comparison of initial search result and those after reranking by the subjective trigger model with/without adaptation.

条件	MAP	初期検索からの向上
初期検索	0.1965	—
再順位付け後	0.2398	22.0%
1) クエリのみ	0.2430	23.6%
モデル適応 2) 代名詞のみ	0.2456*	25.0%
3) クエリ+代名詞	0.2452*	24.8%

ある。これらはたとえば、Abramoff Bush (+3.6%), Colbert Report (+12.1%), Whole Foods (+5.3%), Muhammad cartoon (+25.9%), Bruce Bartlett (+29.3%), Qualcomm (+5.7%), Ariel Sharon (+43.9%) 等である。

次項では、2.4 節で述べたモデル適応を用いることで、特に上述の難しいトピックに関して性能向上が得られるか否かを検証する。

### 3.3.4 モデル適応

2.4 節で述べたモデル適応の手順に従って、意見情報検索に関する追加実験を行った。ここで、モデル適応に用いる文書数  $k$  は実験的に 50 とした。表 6 に、前出の結果とモデル適応を行った場合の結果との比較を示す。なお、モデル適応の際は、次の 3 種類の潜在的トリガー、すなわち、1) クエリ (トピックの title) のみ、2) 14 の代名詞のみ、3) クエリと代名詞の両方を考慮した。表中のアスタリスクは、主観的言語モデルによる再順位付けの結果と比較したとき、符号検定によって  $p < 0.01$  の有意な性能向上が見られた試行を示す。

潜在的なトリガーの選択にかかわらず、トリガーモデルを入力クエリに適応することで、いずれの場合も若干の精度向上が見られた。また、差はわずかであるものの、潜在的なトリガーとして代名詞のみを使った場合の向上が最も大きかった。個別のクエリについてさらに精査してみると、最も顕著な性能向上が得られたのは「Zyrtec」(薬品)であり、平均精度が 0.2187 から 0.3230 (+47.7%) へと向上していた。一方、性能が低下した事例もあり、最も顕著な例では、「Basque」(地名)の平均精度が 0.2061 から 0.1673 (-18.8%) へと低下した。これらの 2 つの事例の違いを対比するため、モデル適応によって新たに同定されたトリガー対のうち、最も影響が大きかった (初期検索で得られた文書集合に最も多く現れた) ものを表 7 に示す。

まず Zyrtec について見ると、「i → sure」や「it → perfect」、「i → bet」、「i → extreme」等、意見の表出を暗示するような有用なトリガー対が多く発見できていることが

表7 「Zyrtec」と「Basque」に関して新たに同定された最も影響の強かったトリガー対。括弧内の数字は、初期検索で得られた文書集合中のそれぞれのトリガー対の頻度を示す

Table 7 Newly identified, most influential trigger pairs for the topics “Zyrtec” and “Basque”, where the numbers in parentheses indicate the frequencies of the respective trigger pairs in the retrieved documents.

Zyrtec			Basque		
you	→ year	(744)	this	→ spanish	(224)
i	→ case	(697)	you	→ come	(161)
it	→ case	(576)	i	→ spanish	(138)
you	→ d	(525)	i	→ told	(108)
i	→ sure	(516)	it	→ last	(97)
this	→ case	(495)	i	→ simply	(85)
it	→ perfect	(478)	this	→ city	(84)
i	→ bet	(456)	it	→ spanish	(78)
you	→ come	(418)	my	→ city	(70)
it	→ kind	(400)	this	→ road	(66)
my	→ year	(353)	i	→ city	(66)
i	→ extreme	(339)	i	→ south	(60)

表8 モデル適応を行ったときの難易度の高いトピックの平均精度の変化。「トリガー」の列は表5より

Table 8 Performance (average precision) change for difficult topics before/after model adaptation.

ID	トピック	トリガー	適応	向上率
854	ann coulter	0.4591	0.4838	+2.5%
871	cindy sheehan	0.4576	0.4640	+0.6%
877	sonic food industry	0.0380	0.0453	+0.7%
886	west wing	0.2407	0.2410	+0.0%
887	world trade organization	0.0658	0.0653	-0.1%
892	jim moran	0.4728	0.4891	+1.6%

分かる。一方、Basqueについては、そのようなトリガー対がほとんど見当たらず、「this → spanish」や「i → south」といったスペインのバスク自治区に関する地理的・一般的な表現を示すトリガー対が多く同定されている。これに対し、TRECが想定するトピックはバスク過激派による分離独立運動であり、クエリのみからはユーザの情報要求を汲み取ることが困難である。このことから、ユーザの情報要求がクエリからは明らかでないような場合、主観的言語モデルの適応が逆効果であると考えられる。

次に、前項で議論した難易度の高いトピックに関して、モデル適応の効果を検証する。表8に、それらのトピックごとの主観的言語モデルを用いた再順位付け後とモデル適応も行ったときの平均精度を示す。この結果が示すように、その効果は非常に限定的ではあるものの、

モデル適応によって多くのトピックで肯定的な影響が見られた。この結果を受け、今後、意見情報検索に対するさらに有効なトリガー対の適応方法・利用方法を検討していく計画である。

#### 4. 関連研究

ブログ等のUGCに対する興味の高まりを反映し、意見情報検索、意見マイニング等、多くの研究が行われている。これらのうち、意見情報検索はTRECブログトラックが2006年に始まって以来、情報検索コミュニティでさかんに研究されている比較的新しい研究分野である。このブログトラックでは、意見情報検索は複数ある共通タスクの1つであり、他のタスクには、極性まで考慮した意見情報検索、フィード検索<sup>27)</sup>等がある。本研究で扱った意見情報検索に関しては、ブログトラックの多くの参加者が、本提案研究と同様に2段階の処理、すなわち、1) 汎用的な検索モデルによる初期検索、2) 初期検索で得られた記事中から意見記事を何らかの方法で同定、という枠組みを採用している。後者の意見記事同定の手法は、大まかに2種類に分けることができる。

最初の手法は語彙に基づく方法であり、自動的あるいは手動で作成した主観的表現のリストを用い、所与のブログ記事の意見らしさを推定する<sup>8)-10),17)</sup>。たとえば、Hannahら<sup>28)</sup>は、種々の言語学的資源から包括的な英単語のリストを作成し、それぞれの単語について、ブログトラック2006のテストコレクションに含まれる適合性判定結果(各ブログ記事に付与されている適合性を示すラベル)を基に、意見弁別性(opinionated discriminability)というスコアを算出した。続いて、このスコアを付与した単語リストを検索システムのクエリとして用いることで、各ブログ記事との適合性を算出し、得られたスコアを各記事の意見らしさを表すスコアとした。この種の手法と比較し、本提案手法は適合性判定結果がまったく不要であり、単に意見を含むような大量の文章があればよいという利点がある。なお、語彙に基づく手法では、一・二人称代名詞、主観的表現とクエリとの距離を考慮した研究<sup>14),29)</sup>もあり、これが本提案研究の着想の1つとなっている。

2番目の手法は、教師付きの学習によって作成した分類器を用いて意見文書を同定するものである。ここでは、この方法でこれまで最も良い結果を得ているZhangら<sup>12),13)</sup>の手法をまとめる。まずZhangらは、訓練データとして大量の意見文書と非意見文書をウェブ上から収集した。具体的には、前者はretails.com、後者はWikipediaから収集した。分類器作成に利用した素性は、単語の1グラムと2グラムである。この分類器をブログ記事の中でクエリを含む文に適用し、それぞれの分類結果を足し合わせることでdの全体的な意

見らしさを評価した。ブログトラック 2006 のテストコレクションを用いて行った実験では、この手法で得られた最良の MAP 値は 0.2726 であった (表 4 参照)。この結果は本提案手法で得られた結果よりも高いものの、3.3.2 項で述べたように、異なる初期検索を用いた場合は本提案手法でもさらに高い MAP 値 (0.3221) が得られている。なお、訓練データから分類器あるいは言語モデルを学習するという観点からは、分類器を用いた手法と本提案手法はいずれも機械学習に基づく手法と見なすことができる。しかしながら、分類手法とは対照的に、本手法では負例 (非意見文章) が必要ないという利点がある。

本提案手法は、上述のどちらの類型とも異なる統計的言語モデルに基づく手法である。著者らが知る限り、これまで意見文の特徴を代名詞と主観的表現の組によるトリガー対としてとらえ、かつ意見情報検索に効果的に適用した例はない。

## 5. おわりに

本稿では、主観的意見表現に焦点を当てたトリガー言語モデルの意見情報検索への適用を議論した。TREC ブログトラック 2006 のテストコレクションを用いた実験では、主観的言語モデルを利用することで、MAP 値で 22.0% の検索精度向上が見られた。また、より詳細な結果の分析によると、同定された主観的トリガー対は、通常の  $n$  グラム言語モデルではとらえられない意見表現の特徴をとらえており、主観的言語モデルの質の向上に貢献していることが確認できた。さらに、個々のクエリを見た場合、特定のカテゴリ、すなわち政治と組織に関するクエリについては、現在の主観的言語モデル (あるいはこのモデルを構築するために利用したアマゾンのレビュー) では、意見情報検索の精度向上が難しい事例があることが明らかになった。この問題に対処するため、クエリへのモデルの動的な適応を行ったところ、多くのクエリではわずかながら精度向上の効果が得られた。

今後の展開としては、より大規模なアマゾンの顧客レビューやブログトラックのテストコレクション等、他のテキスト資源の利用を検討している。また、現在の枠組みでは、ブログ記事を語の系列として扱っているものの、実際には所与のトピックとは無関係な語も多く含まれていると考えられる。これに対処するため、フレーズ抽出等の処理を導入し、文書のより良い表現方法を検討していく。

## 参 考 文 献

1) Adar, E. and Adamic, L.: Tracking information epidemics in blogspace, *Proc. 2005 IEEE/WIC/ACM International Conference on Web Intelligence*, pp.207–214

(2005).

2) Agarwal, N., Liu, H., Tang, L. and Yu, P.S.: Identifying the influential bloggers in a community, *Proc. international conference on Web search and web data mining*, pp.207–218 (2008).

3) Ding, X., Liu, B. and Yu, P.S.: A holistic lexicon-based approach to opinion mining, *Proc. international conference on Web search and web data mining*, pp.231–240 (2008).

4) Esuli, A. and Sebastiani, F.: PageRanking WordNet Synsets: An Application to Opinion Mining, *Proc. 45th Annual Meeting of the Association for Computational Linguistics* (2007).

5) Mei, Q., Ling, X., Wondra, M., Su, H. and Zhai, C.: Topic Sentiment Mixture: Modeling Facets and Opinions in Weblogs, *Proc. 16th International World Wide Web Conference* (2007).

6) Macdonald, C., Ounis, I. and Soboroff, I.: Overview of the TREC-2007 Blog Track, *Proc. 16th Text Retrieval Conference* (2007).

7) Ounis, I., de Rijke, M., Macdonald, C., Mishne, G. and Soboroff, I.: Overview of the TREC-2006 Blog Track, *Proc. 15th Text Retrieval Conference* (2006).

8) Mishne, G.: Multiple Ranking Strategies for Opinion Retrieval in Blogs, *Proc. 15th Text Retrieval Conference* (2006).

9) Oard, D., Elsayed, T., Wang, J., Wu, Y., Zhang, P., Abels, E., Lin, J. and Soergel, D.: TREC-2006 at Maryland: Blog, Enterprise, Legal and QA Tracks, *Proc. 15th Text Retrieval Conference* (2006).

10) Zhang, M. and Ye, X.: A generation model to unify topic relevance and lexicon-based sentiment for opinion retrieval, *Proc. 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pp.411–418 (2008).

11) Seki, K., Kino, Y., Sato, S. and Uehara, K.: TREC 2007 Blog Track Experiments at Kobe University, *Proc. 16th Text Retrieval Conference* (2007).

12) Zhang, W. and Yu, C.: UIC at TREC 2006 Blog Track, *Proc. 15th Text Retrieval Conference* (2006).

13) Zhang, W., Yu, C. and Meng, W.: Opinion retrieval from blogs, *Proc. 16th ACM conference on Conference on information and knowledge management*, pp.831–840 (2007).

14) Zhou, G., Joshi, H. and Bayrak, C.: Topic Categorization for Relevancy and Opinion Detection, *Proc. 16th Text Retrieval Conference* (2007).

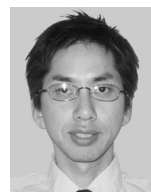
15) Lau, R., Rosenfeld, R. and Roukos, S.: Trigger-based language models: A maximum entropy approach, *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol.2, pp.45–48 (1993).

- 16) Tillmann, C. and Ney, H.: Selection criteria for word trigger pairs in language modeling, *Proc. 3rd International Colloquium on Grammatical Inference*, pp.95–106 (1996).
- 17) Yang, K., Yu, N., Valerio, A. and Zhang, H.: WIDIT in TREC-2006 Blog track, *Proc. 15th Text Retrieval Conference* (2006).
- 18) Manning, C.D. and Schütze, H.: *Foundations of statistical natural language processing*, MIT Press (1999).
- 19) Yang, K., Yu, N. and Zhang, H.: WIDIT in TREC 2007 Blog Track: Combining Lexicon-Based Methods to Detect Opinionated Blogs, *Proc. 16th Text Retrieval Conference* (2007).
- 20) Lavrenko, V. and Croft, B.: Relevance based language models, *Proc. 24th annual international ACM SIGIR conference on Research and development in information retrieval*, pp.120–127 (2001).
- 21) Sakai, T., Manabe, T. and Koyama, M.: Flexible pseudo-relevance feedback via selective sampling, *ACM Trans. Asian Language Information Processing*, Vol.4, No.2, pp.111–135 (2005).
- 22) Tao, T. and Zhai, C.: Regularized estimation of mixture models for robust pseudo-relevance feedback, *Proc. 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pp.162–169 (2006).
- 23) Jurafsky, D. and Martin, J.H.: *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*, Prentice Hall (2000).
- 24) Salton, G. and McGill, M.J.: *Introduction to Modern Information Retrieval*, McGraw-Hill, Inc. (1983).
- 25) Metzler, D. and Croft, W.: Combining the Language Model and Inference Network Approaches to Retrieval, *Information Processing and Management, Special Issue on Bayesian Networks and Information Retrieval*, Vol.40, No.5, pp.735–750 (2004).
- 26) Macdonald, C., He, B., Ounis, I. and Soboroff, I.: Limits of opinion-finding baseline systems, *Proc. 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pp.747–748 (2008).
- 27) Elsas, J.L., Arguello, J., Callan, J. and Carbonell, J.G.: Retrieval and feedback models for blog feed search, *Proc. 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pp.347–354 (2008).
- 28) Hannah, D., Macdonald, C., Peng, J., He, B. and Ounis, I.: University of Glasgow at TREC 2007: Experiments in Blog and Enterprise Tracks with Terrier, *Proc. 16th Text Retrieval Conference* (2007).
- 29) Vechtomova, O.: Using Subjective Adjectives in Opinion Retrieval from Blogs, *Proc. 16th Text Retrieval Conference* (2007).

(平成 20 年 11 月 20 日受付)

(平成 21 年 1 月 7 日再受付)

(平成 21 年 1 月 29 日採録)



関 和広

平成 14 年図書館情報大学情報メディア研究科修士課程修了。平成 18 年インディアナ大学図書館情報学研究科博士課程修了。同年より神戸大学助手(現、助教)。情報検索、自然言語処理、機械学習の研究に従事。Ph.D. 電子情報通信学会、自然言語処理学会、ACM SIGIR 各会員。



上原 邦昭 (正会員)

昭和 53 年大阪大学基礎工学部情報工学科卒業。昭和 58 年同大学院博士後期課程単位取得退学。同産業科学研究所助手、講師、神戸大学工学部情報知能工学科助教授、同都市安全研究センター教授を経て、現在、同大学院工学研究科教授。工学博士。人工知能、特に機械学習、マルチメディア処理の研究に従事。人工知能学会、電子情報通信学会、計量国語学会、日本ソフトウェア科学会、AAAI 各会員。