ブログを対象とした統計的意見情報検索

佐藤翔平† 関和広‡ 上原邦昭†
†神戸大学工学研究科 ‡神戸大学自然科学系先端融合研究環
sato-shohei@ai.cs.kobe-u.ac.jp

1 はじめに

ホームページや掲示板・ブログなど，個人発信型のウェブコンテンツの増加により，テレビや雑誌等の従来のメディアからは得ることが難しかった個人の主観的意見の収集が可能になってきている．特に，ブログはその簡便性から利用者の増加が著しく，主観的な意見情報を収集するための資源として言語処理，情報検索等の研究コミュニティで近年注目されている [1, 2, 3, 4]．特に，情報検索関連の重要な会議である TREC と NTCIR では，それぞれ 2006 年と 2007 年からブログを対象にした検索タスクが設定されている [5, 6]． これらの研究により，即時的，かつ大規模に所与の対象に関する主観的意見を収集することができれば，企業のマーケティング，個人の意思決定等に有用であると考えられる．

本研究では，所与の対象に関して著者の主観的な記述を含むブログ記事をより高精度に検索する手法を提案する．具体的には，主観的表現はしばしば意見を述べる主体，あるいは批評される対象と対で現れることに着目し，代名詞と主観的表現の組をトリガー対と仮定する．そして，言語モデルの一つであるトリガーモデルを構築，意見情報検索に利用する．モデルの構築には，Amazon.com から取得した大規模なレビュー集合を用いる．TREC 2006 ブログトラックのテストコレクションを用いた評価実験を通して，提案モデルの有効性を示す．

2 提案手法

2.1 概要

ある文章が主観的意見を含むかどうかを判断する単純な方法は，意見を述べる際によく使われる語（主観的表現と呼ぶ），例えば「good」等が文章内に現れているかどうかに着目することである．しかし，主観的表現が文内に現れたとしても，その文が必ずしも意見を表すわけではない．例えば，「This movie is good.」は書き手の意見を表していると考えられるのに対して，「I'm good.」は，一般的に，自分の体調や気分が何ら問題ないという事実あるいは主張を表す．つまり，ある文が意見であるかどうかをより正確に判断するためには，単語のみに注目するのではなく，共起語や係り受け関係にある語など，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方法として，統計的な言語処理で広く用いられている n グラム言語モデルがある．これは，連接する n 語を考慮したモデルであり，例えば，2 グラムの場合，単語ではなく，「This movie」「movie is」「is good」や「I am」「am good」のように，文を順次 2 単語ずつに分割して対象言語の特徴を捉える．これによって，独立した単語だけではなく，n で指定される一定の窓内の連接を考慮した言語のモデル化が可能になる．しかし，n があまり大きいと，モデル構築に必要なコーパスの量が膨大になり，また構築されるモデル自体が巨大になってしまうため，現実には $n = 3$ 程度で利用されることが多い．

本研究では，意見を表現する際に特に重要な語間の関連に焦点を当て，意見表現推定に特化した言語モデルの構築を目指す．このために，主観的表現がしばしば意見を述べる主体，あるいは批評される対象と対で現れることに着目する．そして，これらの表現の組をトリガー対と見なし，言語モデルの一つであるトリガーモデルを構築する．

2.2 意見表現のためのトリガーモデル

Tillmann と Ney [7] は，n グラムモデルで表現することが難しい非連接または非近傍の語間の依存関係を表すため，トリガーモデルを提案した．トリガーとは何らかの語の出現を誘発する語のことであり，誘発される語を非トリガーという．また，両者を合わせてトリガー対という．トリガーモデルは，これらトリガー

対の依存関係を表現したモデルであり，次式のようにベース言語モデル（通常 n グラム言語モデル）と線形補完して用いる．

$$P_E(w|h) = (1-\lambda) \cdot P(w|h) + \lambda \cdot P_T(w|h) \qquad (1)$$

ここで，w は語，h は w に先行する語（履歴），$P(w|h)$ はベース言語モデル，$P_T(w|h)$ はトリガーモデル，そして λ が補完係数である．

トリガーモデルを構築する際は，まずトリガー対を発見する必要がある．トリガー対の発見は，任意の語の組について，それらの語間の依存関係を考慮した場合と考慮しない場合の二通りの言語モデルを作成し，モデル間の対数尤度差に基づいて行う．このように同定したトリガー対について，最尤推定によりトリガーモデルのパラメタ推定を行う [7]．

本研究では，上述のトリガーモデルを基に，主観的表現は意見を述べる主体，あるいは批評される対象と対で現れることが多い点に着目し，前者を非トリガー，後者をトリガーと仮定する．例えば，「I enjoyed it.」において「I」はトリガーであり，「enjoyed」は非トリガーであると考えられる．また，トリガーは文章中で代名詞として表出することが多いという観測に基づいて，あらかじめ定めた代名詞だけをトリガーとして扱う．

2.3 トリガーモデルの構築

トリガー対の発見，パラメタ推定等，提案モデルの構築に用いるコーパスは主観的意見を含む文章でなければならない．ここでは，`Amazon.com` から自動的に収集した 50 万件（ファイルサイズ 433MB，延べ語数約 7,900 万語，異なり語数約 98 万語）のユーザーレビューをモデル構築に用いた．なお，レビューの対象商品は，`Amazon.com` で扱っているすべての商品（書籍，DVD，家電品，玩具など）である．

トリガーとしては，実験的に次の 14 種の代名詞，I, my, you, it, its, he, his, she, her, we, our, they, their, this を利用し，上記のレビューコーパスを利用して，50,000 のトリガー対を同定した．なお，履歴 h は同文内で先行する語すべてとした．表 1 に同定されたトリガー対の抜粋とその対数尤度差を示す．

これらのトリガー対に基づき，トリガーモデル $P_T(w|h)$ のパラメタ推定を行った．ベース言語モデル $P(w|h)$ としては，バックオフ 3 グラム言語モデルを用いた．また，ベース言語モデルとトリガーモデルの重みは同等と仮定し，補完係数 λ を 0.5 とした．

表 1: 同定されたトリガー対の一部

トリガー		非トリガー	対数尤度差
i	→	really	11.307
this	→	movie	11.159
i	→	thought	10.881
it	→	actually	9.988
it	→	quite	9.927
this	→	!	9.905
this	→	truly	9.848
we	→	pretty	9.787
i	→	perfect	9.698

3 評価実験

3.1 実験データ

本節では，TREC 2006 ブログトラックで用いられた Blog06 コレクション [5] を実験に用いた．このコレクションは，2005 年 12 月から 2006 年 2 月の 11 週間に自動収集された約 322 万件のブログ記事から成り，情報検索システムの評価実験のための 50 件のトピック（ユーザの検索要求）とその適合性評価の情報も含んでいる[1]．適合性の評価は，5 つのラベル，「不適合」「適合」「肯定的」「否定的」「混在」を用いて行われている．不適合はトピックに関連しない記事，適合はトピックに関連する記事（意見は含まない），肯定的はトピックに関連し，かつ肯定的な意見が含まれる記事，否定的はトピックに関連し，かつ否定的な意見が含まれる記事，混在はトピックに関連し，かつ肯定的および否定的な意見が含まれる記事に付与されている．

3.2 言語モデルの評価

2.3 節で構築したトリガーモデルの特性を調べるため，Blog06 で不適合以外のラベルが付与されている記事のそれぞれについて，下式のクロスエントロピーを算出した．

$$\begin{aligned} C_E(d) &\approx -\frac{1}{n} \log P_E(w_1 \ldots w_n) \\ &\approx -\frac{1}{n} \log \sum_{i=0}^{n} P_E(w_i|h_i) \end{aligned} \qquad (2)$$

クロスエントロピーは言語モデルの評価に用いられ，その値が小さいほど対象とする言語をより正確に表現していると考えられる．前節で構築したトリガーモデルは，主観的意見に特徴的なトリガーを考慮したモデルであるため，非意見記事（適合のラベルが付与された記事）ではその値が高く，意見記事（肯定的，否定的，または混在のラベルが付与された記事）ではその値が低

[1] 本研究では，2007 年度に追加されたトピックと適合性評価データは用いていない．

くなることが望ましい．クロスエントロピーの値の分布を意見記事（opinion），非意見記事（non-opinion）ごとに示したグラフを図 1 に示す．

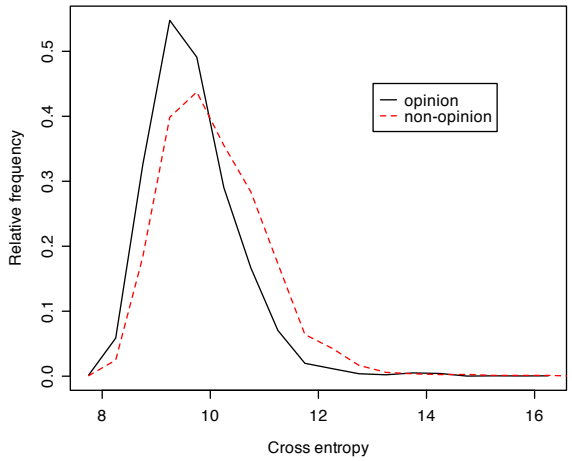

図 1: トリガー対を組み込んだ言語モデル $P_E(w|h)$ による意見記事・非意見記事のクロスエントロピーの分布

この結果から，意見記事の分布は非意見記事と比較してやや左寄りであり，提案モデルが意見記事をより良く表現していることが確認できる．次節では，このエントロピーの値を用いて，一般的な情報検索モデルによる初期検索結果を意見情報検索の観点から改善する．

3.3 検索実験

本節では，既存の一般的な情報検索モデルによって得られた初期検索結果を，トリガーモデルによる文書ごとのクロスエントロピー値を考慮することで再順位付けし，意見情報検索を実現する．なお，初期検索は次の条件で行った．検索モデルにはにはベクトル空間モデル，単語重み付けには TFIDF，類似度の計算にはコサインを用いた．索引付け・検索の際は，大・小文字を区別をせず，ストップワードは除去している．また，接辞除去は行っていない．検索質問としては，トピックのタイトルのみ（すなわちキーワード）を用いた．表 2 に，TREC 2006 ブログトラックの公式結果（トピックのタイトルを検索質問に用いた結果のみ）と初期検索結果（initial）を Mean average precision（MAP）で示す．なお，ここでは意見の極性とは無関係に，肯定，否定，混在のいずれかのラベルが付与されている記事が適合文書（意見記事）として扱われている．

表 2 から，初期検索結果は，公式結果の中間値と同等であることが分かる．

初期検索では，検索された各記事 d に対して検索質問との間のコサイン類似度 $Sim(d)$ が与えられており，

表 2: TREC 2006 公式結果と初期検索結果の比較

		MAP
TREC	Best	0.1885
	Worst	0.0000
	Median	0.1156
Initial		0.1126

この類似度に基づいて記事は順位付けされている．（この時点では，意見記事を検索するための処理は一切行っていない．）続いて，トリガー対を利用した提案言語モデルによって推定される意見記事らしさを検索結果に反映する．ここでは実験的に，コサイン類似度に記事 d のクロスエントロピー $C_E(d)$ の逆数を重み付けして加えることで，初期検索結果の再順位付けを行った．

$$Scr(d) = Sim(d) + \frac{\alpha}{C_E(d)} \tag{3}$$

式 (3) における第 2 項の適切な重みを検討するため，α の値を 0 から徐々に増やしながら意見記事検索精度の推移を観測した．図 2 に結果を示す．

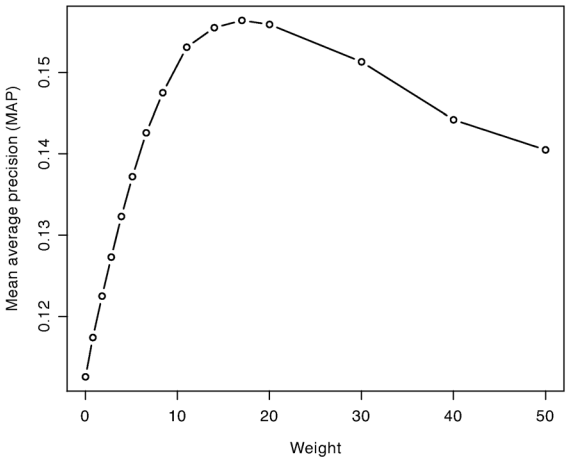

図 2: 重み α と MAP の関係

図 2 の左端（$\alpha = 0$）が初期検索の結果に相当する．重み α を増やすにしたがって MAP が上昇し，$\alpha = 17$ で初期検索と比較して 38.9%の精度向上を示した（MAP=0.1564）．この結果は，商品のカスタマーレビューから構築した言語モデルが一般的な意見記事検索にも有用であることを示している．

3.4 ベース言語モデルとの比較

前節の実験により，意見記事検索における提案モデルの有用性が示された．しかし，検索性能向上におけるトリガー対の寄与は，トリガー対を利用しないベース言語モデルとの比較を行わない限り明らかではない．そこで，提案モデル $P_E(w|h)$ とベース言語モデル $P(w|h)$

による各記事 d のクロスエントロピーの差，すなわち $C_E(d) - C(E)$ の度数分布を意見記事と非意見記事について調査した（式 (2) 参照）．一般に，クロスエントロピーが小さいほど良いモデルであることを意味するため，意見記事に関しては差が負数になる記事が多いほど，また非意見記事に関しては差が正になる記事が多いほど提案モデルが意見記事の推定に有効であることを示す．図 3 にそれぞれのヒストグラムを示す．

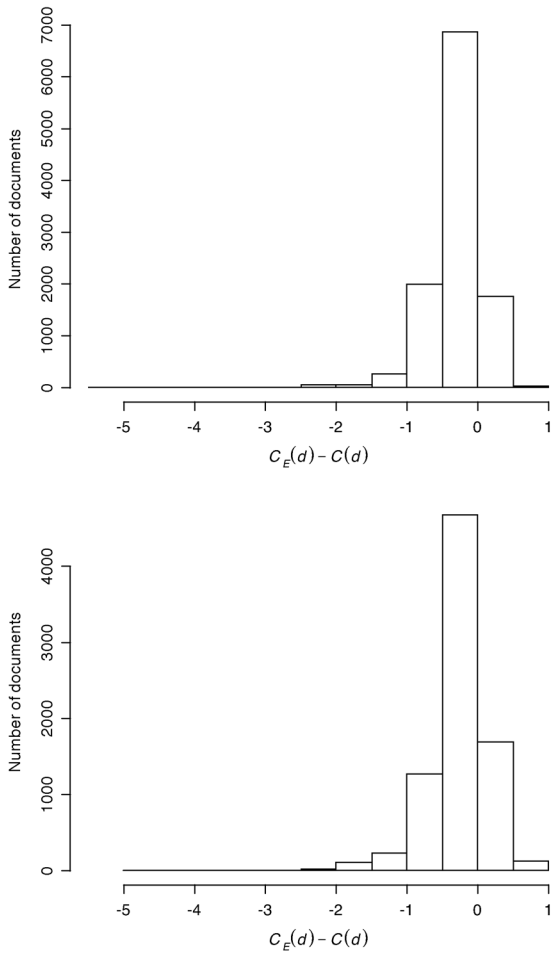

図 3: 提案モデルと 3 グラム言語モデルによるクロスエントロピーの差.（上）意見記事,（下）非意見記事

二つのヒストグラムとも負の度数が多く，意見記事，非意見記事に関わらず，トリガー対を取り入れることでモデルのパラメタ推定が向上していることが分かる．しかし，-1.0 から -0.5，0 から 0.5 の区間に注目すると，意見記事と非意見記事では逆の傾向が現れている．すなわち，意見記事では負の度数が多く，非意見記事で正の度数が多い．この違いが，トリガー対を利用したことによる意見記事・非意見記事に関するモデルの弁別性向上を示していると考えられる．

4 おわりに

本研究では，代名詞表現と主観的表現に着目し，意見文に特徴的なトリガー対を自動的に同定した．また，これらのトリガー対を組み込んだ言語モデルが意見情報検索に有用であることを示した．ただし，`Amazon.com` から収集したユーザーレビューは商品に関する批評であるため，必ずしも現実の情報要求に適していない可能性がある．例えば，商品ではなく，国や個人に関する意見を探す場合，これらの批評に使われる表現・語彙は，書籍等の商品に用いられる表現・語彙とは異なると考えられる．このような情報要求へ対応するためには，トリガー対，言語モデルを拡張・更新方法を検討する必要がある．

参考文献

[1] Andrea Esuli and Fabrizio Sebastiani. PageRanking WordNet synsets: An application to opinion mining. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, 2007.

[2] Nozomi Kobayashi, Kentaro Inui, and Yuji Matsumoto. Extracting aspect-evaluation and aspect-of relations in opinion mining. In *Proceedings of the Joint Meeting of the Conference on Empirical Methods in Natural Language Processing / the Conference on Computational Natural Language Learning*, 2007.

[3] Ryan McDonald, Kerry Hannan, Tyler Neylon, Mike Wells, and Jeff Reynar. Structured models for fine-to-coarse sentiment analysis. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, 2007.

[4] Qiaozhu Mei, Xu Ling, Matthew Wondra, Hang Su, and ChengXiang Zhai. Topic sentiment mixture: Modeling facets and opinions in weblogs. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, 2007.

[5] Iadh Ounis, Maarten de Rijke, Craig Macdonald, Gilad Mishne, and Ian Soboroff. Overview of the TREC-2006 blog track. In *Proceesings of the 15th Text Retrieval Conference*, 2006.

[6] Yohei Seki, David Kirk Evans, Lun-Wei Ku, Hsin-His Chen, and Noriko Kando. Overview of opinion analysis pilot task at NTCIR-6. In *Proceedings of the 6th NTCIR Workshop*, 2007.

[7] Christoph Tillmann and Hermann Ney. *Grammatical Interference: Learning Syntax from Sentences*, chapter Selection criteria for word trigger pairs in language modeling, pp. 95–106. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 1996.