WebAlert：Web 情報の印象集約を利用した閲覧ページ内容に対する反対意見提示

WebAlert：Counter Opinions Presentation to Reading Web Page using Sentiment Aggregation

山本 祐輔 ♥ 手塚 太郎 ♦
アダム ヤトフト ♠ 田中 克己 ♣

Yusuke YAMAMOTO Taro TEZUKA
Adam JATOWT Katsumi TANAKA

本論文では，閲覧中の Web ページの内容に誤りの可能性がある際に警告を促すシステム WebAlert を提案する．本システムでは，トピックに対する印象に着目し，あるトピックに関する閲覧中の Web ページのポジティブな内容 (またはネガティブな内容) に対して，そのトピックに関する他の記事の多くがネガティブな内容 (またはポジティブな内容) で書かれていた場合ページ内容が信頼できない可能性があると考える．WebAlert は Web ブラウザとして実装されており，本システムを用いることで，ユーザは意識することなく閲覧中の Web ページの内容に誤りに気付くことが可能となる．

In this paper, we propose WebAlert, the system which alerts users when the content of web pages which they watch can be wrong. Our system analyzes the web page using sentiment mining. If the page which users watch has positive contents and most of web pages about the same topic have negative contents, we think that the content of its web page can be wrong. WebAlert is implemented as web browser. Using this system, users can find wrong contents in the web page if they always worry about the trustworthiness of it.

1. はじめに

今日，インターネットの普及により Web コンテンツを閲覧・発信する機会が増えている．ユーザは必要とする情報を必要なときに膨大な Web ページから自由に取得することができる．しかしながら，誰でも容易にコンテンツをアップロードできるため，Web 上に存在するページは既存のマスメディアの情報のように信頼性は保証されていない．このように Web からの情報取得は，必要な情報を自由に取得できる反面，誤った情報を取得してしまう可能性も存在する．そのため信頼性を考慮した Web 検索閲覧環境の実現が望まれている．

ユーザが Web ページに至る経路としては，(1)Web 検索エンジンからの流入，(2) リンクナビゲーションによる流入，(3) 直接 Web ページに流入する (URL を打ち込む)，などが挙げられる．昨今，PageRank[1] に代表される高精度の検索アルゴリズムの出現により，Web 検索エンジンが隆盛を極めている．その多くはクエリと Web ページの内容適合度評価やリンク解析に基づく支持度評価に焦点を当てており，信頼性の高い Web ページを検索するには不向きである．このために Nakamura らは信頼性の観点からの Web 検索結果のリランキング機構を提案している [2]．また筆者らはフレーズの形で入力されたファクト型知識の信頼性を評価判断するための検索エンジンを提案している [3]．このように Web 検索エンジンの信頼性に関する研究が立ち上がってきた一方で，閲覧中の Web ページの内容の信頼性を評価ならびに判断支援する機構に関してはあまり提案されていない．情報の信頼性の担保が最も必要となるケースは**その情報を知らず知らずのうちに受け入れて不都合が生じるケース** であることが多い．よって閲覧ページの信頼性を評価する場合，閲覧中に背後で自動的に信頼性評価をし，信頼性の低い情報がある場合に通知するようなシステムが必要となる．

そこで本論文では，閲覧中の Web ページの内容が信頼できない可能性がある際に警告および反対意見を提示するシステム WebAlert を提案する．本システムでは，信頼できない内容であるかを評価する基準としてトピックに対するページ内容の印象に着目する．あるトピックに関する閲覧中の Web ページのポジティブな内容 (またはネガティブな内容) に対して，そのトピックに関する Web 上の記事の大半がネガティブな内容 (またはポジティブな内容) で書かれていた場合，内容が誤っている可能性があると考え，警告および反対意見を提示する．本システムを用いることで，ユーザは意識することなく閲覧中の Web ページの内容に誤りに気付くことが可能となる．

2. システムの概要

本章では，提案システム WebAlert の概要を述べる．図 1 は WebAlert のシステムフローを表している．本システムでは Web ブラウザとして実装されており，ユーザが本システムを通して Web ページを閲覧すると，バックグラウンドで内容に対する反対意見が存在するかを分析する．そのために，まずシステムは閲覧中の Web ページを適当なブロック (段落) に分割する．ブロック分割は<p>タグ，<td>タグなどを利用する．次に分割されたブロックから内容を表したクエリを生成する．次に，生成されたクエリを検索エンジンに投げ，得られた検索結果を印象分析し，

♥ 学生会員　京都大学大学院 情報学研究科 社会情報学専攻 博士課程後期課程　yamamoto@dl.kuis.kyoto-u.ac.jp
♦ 正会員　立命館大学 総合理工学院 情報理工学部 メディア情報学科　tezuka@media.ritsumei.ac.jp
♠ 非会員　京都大学大学院 情報学研究科 社会情報学専攻 adam@dl.kuis.kyoto-u.ac.jp
♣ 正会員　京都大学大学院 情報学研究科 社会情報学専攻 tanaka@dl.kuis.kyoto-u.ac.jp

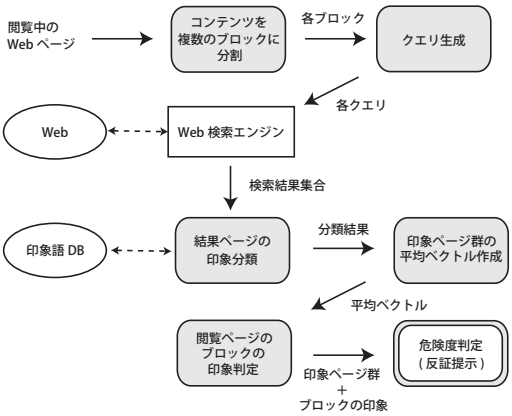

図 1　WebAlert のシステムフロー

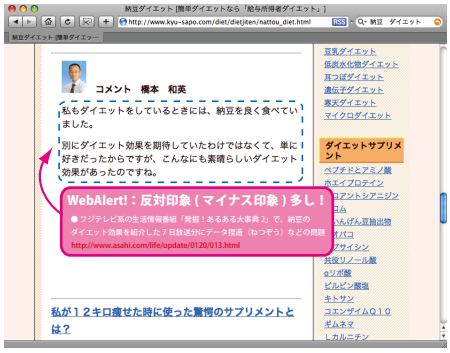

図 2　WebAlert の動作例

ポジティブな内容/ネガティブな内容に分類した後，クエリの生成元となったブロックの内容がポジティブな内容/ネガティブな内容のどちらに分類されるかを判定する．ブロックの内容が Web 上で多数派の印象内容に分類された場合は安全，少数派に分類された場合は危険と判定する．ブロックが危険と判定された場合，システムはブロックの下に警告および反対意見を挿入する．反対意見としては反対印象のページ要約と URL を提示する．ブロックの内容が危険でない，あるいは特に印象が抽出されなかった場合は警告は出されない．このような処理を分割された全てのブロックに対して行う．

図 2 は実際に WebAlert を通して Web ページを閲覧した例を示している．例として取り上げた Web ページでは，「納豆ダイエットは非常に効果的である」というテーマの基，納豆がダイエットに効果的な要因，体験談などが書かれている．閲覧中にシステムが内容をバックグラウンドで解析した結果，危険であると判定されたブロックの下に警告が挿入される．図では閲覧ページ中で「納豆はダイエットに非常に良い」とポジティブに書かれている内容に対して，Web 上ではネガティブな内容が多いという警告と「納豆ダイエットにテレビ局が捏造した話題である」という反対意見が提示されている様子が確認できる．反対意見には URL も付加されているので，反対内容を詳しく確認したい場合，URL をたどることで確認できる．この例では反対意見としてネガティブな内容が抽出されたため赤色の警告が出されたが，ポジティブな内容が反対意見として抽出された場合青色の警告が出される．このような処理をシステムがバックグラウンドで自動的に行うので，ユーザが信頼性を意識せずに Web 閲覧をしていたとしても，内容に誤りの可能性がある場合は気付くことができる．

3.　アルゴリズム

本章では，WebAlert の内部処理の詳細を述べる．本システムは既存の Web 検索エンジンを利用している．閲覧ページの内容に対して警告を出すべきかの情報を収集するために，システム内でクエリを生成する度に，リアルタイムで検索エンジンに問い合わせを行う．以下では，閲覧ページをブロックと呼ばれる段落に分ける手法，検索エンジンに問い合わせをするためのブロックからのクエリ生成手法，文章の印象分析，ならびにブロックに対する反対意見検索の手法を述べる．

3.1　閲覧ページの分割

まず閲覧ページを幾つかの段落に分ける手法を述べる．本システムは閲覧ページの内容とは逆の内容があるか無いかを Web 上の情報を集約することで判定することが狙いであるが，閲覧ページの内容の全体を評価対象とするのは粒度が荒すぎる．閲覧をしているユーザにとっては，ページのどの箇所が正しく，どの箇所が誤りであるかを部分的に指摘された方が内容の取捨選択がしやすい．そこで本システムでは，閲覧ページを「ブロック」と呼ばれる段落に分割し，各ブロックの内容の反対意見分析を行う．

文章をブロックに分割する場合，形式段落に分割することが自然である．Web ページも文章であるので形式段落が存在する．人間が文章から形式段落を認識するには段落頭を字下げ，あるいは段落間の空行を手がかりにしているが，計算機の場合，そのような視覚的な手がかりを理解することは困難である．幸いにして Web ページは HTML を用いて作成されているため，ページ作成者が構造を明確にしてページを作成した場合は DOM(Document Object Model) を基にブロックを抽出することができる [5]．DOM は Web ページの階層構造を記述するモデルであり，特に<P>，<TABLE>，<H1>~<H6>などのタグは文章中の意味的なブロックを抽出する情報として有益と考えられる．本システムでは文章内のできるだけ小さいの意味ブロックごとに反対意見分析を行うために，形式段落を意味する<P>タグに着目する．以下 Web ページからブロックを抽出する手順を記す．

1. 閲覧中の Web ページを DOM パーサに入力する
2. <P>タグを抽出する
3. 各<P>タグ内のテキストを取り出す
4. 余分なタグを除去しそれらをブロックとする

3.2　ブロックからのクエリ生成

3.1 章で提案した手法で得られたブロックの内容の反対意見が存在するかを分析するために本研究では Web 上の情報を利用する．ブロックの内容と関連のある情報を収集するためには検索エンジンに問い合わせるためのクエリを生成する必要がある．適切

なクエリとしてはブロックの特徴を表す語集合が考えられる．文書検索の分野では，文書の内容を tf/idf 法，信号/雑音比を利用して語に重み付けをした特徴ベクトルで表現することがあるが，これらの手法は文書が複数与えられたときに有効な手法であるため，本システムのようにそもそも文書集合が閲覧ページのみである状況下では使えない．ブロックを文書と見なし，ブロック集合から特徴ベクトルを生成することも考えられるが，tf/idf 法，信号/雑音比は文書間の相対的な観点から語の重みを決定するため，他のブロックと比べて特徴的なベクトルが生成されるためブロックの内容を表す適切なベクトルが作成できるとは限らない．

そこで本研究では専門用語自動抽出 Web サービス「言選 Web」を用いる[1]．言選 Web では，単名詞バイグラムを用いることにより複合名詞がどのような単名詞で構成されているかという連接情報と候補語の頻度情報を用いて文章から重要語抽出を行う．言選 Web に文章を与えると重要語とスコアのリストが得られる．検索エンジンに投げるクエリ数が多いと得られる検索結果数も少なくなるため，本システムではブロックの文章を言選 Web に与えた際，得られた重要語のうち上位 3 つをブロックを特徴付けるクエリとして採用する．以下は実際に言選 Web に文章を与えた例である．

文章例

納豆ダイエットとは、納豆に含まれる納豆菌、食物繊維、オリゴ糖、ポリアミン、ナットウキナーゼ、ジピコリン酸、イソフラボンの成分で整腸して、下っ腹をスッキリさせるダイエット方法です。

抽出されたクエリ

{納豆ダイエット=>1.86，納豆=>1.73，ダイエット方法=>1.68}

3.3 文章の印象分析

ある文章がブロックに対する反対意見が存在するかを判定するために文章の印象を利用する．このため文章の印象を判定する必要がある．本研究では文章の内容に対する印象をポジティブ/ネガティブの 2 種類に分類する．印象の判定手法に熊本らの提案手法を参考としている [4]．本手法では，文章の印象は印象尺度「ポジティブ⇔ネガティブ」に 0 1 に対応させた印象スコアとして算出される．印象スコアは予め作成された印象語辞書 DB と判定対象の文章から算出される．以下印象語辞書 DB の構築手法と文章の印象判定手法を述べる．

印象語辞書辞書は予め用意された印象語リストに存在する単語と印象スコアの評価対象となる語と共起関係から算出される．本研究ではポジティブ/ネガティブな印象語リストとして表 1 に記されるような語を各印象毎に 31 個ずつ用意した．このリストは筆者らの主観によって構成した．

次に各印象尺度に対応する印象語リスト中の語をクエリとして順次 Web 検索エンジンに投げ，その検索結果を取得する．そして各検索結果のスニペット (検索結果要約) 中に現れる単語の出現頻度をカウントする．これにより各印象尺度と抽出された単語 (名詞，動詞，形容詞，副詞) との共起関係を評価する．ここで，抽出された単語 t とポジティブ印象語リスト $PosList$ 中の語との共起確率 $P(t, PosList)$ とネガティブ印象語リスト $NegList$ 中の語との共起確率 $P(t, NegList)$ との内分比 s を求める (式 1)．

印象尺度	印象語リスト
ポジティブ	幸せな，穏和な，感謝する，役に立つ， 信頼できる，公平な，楽しい，誠実な，快適な
ネガティブ	不幸な，侮辱する，役に立たない，不快， 不公平な，つまらない，不誠実な，頼りない

表 1 印象語リストの一部

$$s = \frac{P(t, PosList)}{P(t, PosList) + P(t, NegList)} \tag{1}$$

内分比 S は 0 に近いほど単語がネガティブな印象な文脈に現れやすく，1 に近いほどポジティブな文脈で現れやすいことを意味する．この内分比を単語の印象値として用いる．今回は印象語リスト中の語をクエリとして収集された 62000 の検索結果に対して解析を行い，85776 個の単語について印象値を評価した．

次に構築された印象語辞書 DB を用いて与えられた文章の印象判定を行う手法を述べる．まず文章が入力されると，形態素解析器 MeCab[2]により文章を形態素解析し，名詞，動詞，形容詞，副詞を抽出する．そして，抽出された形態素 $t_i(i = 1, ...n)$ に対応する印象スコア s_{t_i} を印象語辞書 DB から取り出す．最終的に以下の式により文章の印象値 S を算出する．ポジティブともネガティブとも判定されなかった一般的な単語 (印象値 s の値が 0.5 付近の単語) に関するスコアの影響を緩和するために，$|2s_{t_i} - 1|$ という項を s_{t_i} に掛けるという演算を行っている．

$$S = \frac{\sum_{i=1}^{n} |2s_{t_i} - 1| s_{t_i}}{\sum_{i=1}^{n} |2s_{t_i} - 1|} \tag{2}$$

文章の印象判定は，しきい値 $\theta_{impress}$ を設定し，$S > \theta_{impress}$ の時文章は「ポジティブな内容」，$S < 1 - \theta_{impress}$ の時「ネガティブな内容」であると判定する．それ以外の時は「どちらでもない」と判定する．

3.4 反対意見検索

最後に各ブロックの内容に対する反対意見の存在を分析する手法を述べる．ここでブロックに対する反対意見が存在するとは先に定義したように「ブロックの内容がポジティブ (ネガティブ) である際に，同じトピックに対する Web 上の情報の大半がネガティブ (ポジティブ) である」ことを指す．本システムによって閲覧ページから抽出された各ブロックに対して反対意見が存在するか否かを判定する手順を以下に記す．図 3 は手順の概略を表している．

[1] 言選 Web，http://gensen.dl.itc.u-tokyo.ac.jp/gensenweb.html

[2] MeCab，http://mecab.sourceforge.net/

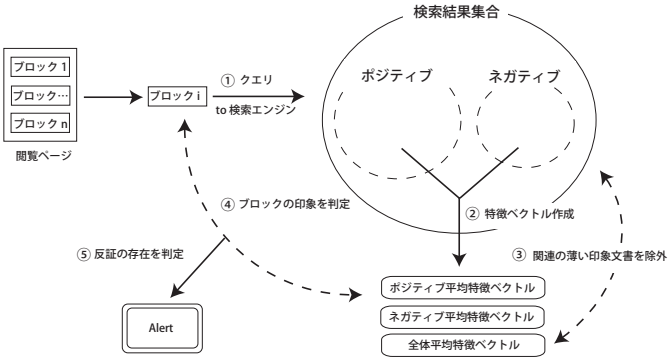

図 3　反対意見存在判定のプロセス

1. **ブロックに関連する Web ページの収集と印象分類**
   3.2 章で述べた手法により各ブロックから抽出されたクエリを Web 検索エンジンに投げることで，ブロックの内容と関連のある Web ページを収集する．この際検索結果として上位 N 件を取得する．次に 3.3 章で述べた手法により，得られた検索結果ページをポジティブな印象ページ群，ネガティブな印象ページ群に分類する．この時印象分析に用いる文章はスニペットを用いる．Web ページをダウンロードせずに検索結果要約を用いることで高速化が図れる．検索結果 N 件に対して印象を分析を行った結果，全ての検索結果がポジティブ，ネガティブのいずれでも無いと判定された場合は以降の処理は行わない．
2. **印象ページ群からの特徴ベクトルの作成**
   全ステップで判定された各印象ページ群から特徴ベクトルを生成する．ここで作られたベクトルは後のステップで述べる「ブロックの印象判定」「クエリと関連のないページ群の除外」「反対意見が存在した場合の提示文書の選択」に用いられる．まずブロックの文章と 1 ステップで得られた検索結果要約集合からそれぞれを表現する特徴ベクトルを作成する．ベクトルの生成には tf/idf 法を用いる．今，ブロック文章の特徴ベクトルを$\boldsymbol{v_{b_i}}$，各検索結果 $r_j(j = 1, ...n)$ のスニペットの特徴ベクトルを$\boldsymbol{v_{r_j}}$とする．次に，後のステップのためにポジティブな印象ページ群を代表するベクトル，ネガティブな印象ページ群を代表するベクトル，ポジティブ/ネガティブな印象ページ群の全体を表すベクトルを作成する．各ページ群の代表するベクトルを求めるには各ページ群の平均ベクトルを用いることが考えられる．そこでそれらを以下のように定義する．

$$\boldsymbol{p_{avg}} = \frac{1}{N_p} \sum^{Positive} \boldsymbol{v_{r_j}} \tag{3}$$

$$\boldsymbol{n_{avg}} = \frac{1}{N_n} \sum^{Negative} \boldsymbol{v_{r_j}} \tag{4}$$

$$\boldsymbol{e_{avg}} = \frac{1}{N_p + N_n} \sum^{Positive+Negative} \boldsymbol{v_{r_j}} \tag{5}$$

式中の N_p はポジティブと判定された印象ページの数，$Positive$ はポジティブと判定された印象ページ群，N_n はネガティブと判定された印象ページの数，$Negative$ はネガティブと判定された印象ページ群を意味する．

3. **ブロックとの関連性の低い印象ページの除外**
   ステップ 1 で判定された印象ページ群は印象語辞書 DB から判定されているため，クエリが表す内容に対してポジティブ (またはネガティブ) な内容が書かれていない場合でも，一般的にポジティブ (またはネガティブ) な文脈で使われる単語が多く含まれるとポジティブ (またはネガティブ) な文章候補として判定されてしまっている可能性がある．よってクエリの表す内容と関係のないページを除外する必要がある．
   ブロックの内容に関連があり，かつポジティブないしはネガティブな内容をもつ文章は「比較軸となる共通要素」と「ポジティブ (またはネガティブ) と判定される要因」から構成されると考えられる．ステップ 1 で判定された両印象のページ群の要素がブロックの内容に関連するためには少なくとも「比較軸」は文章中に含まれていなければならない．比較軸はポジティブな印象ページ群とネガティブな印象ページ群の共通要素であるので，ステップ 2 で定義した特徴ベクトル$\boldsymbol{e_{avg}}$が共通要素を特徴付けるベクトルと考えられる．ブロックとの関連性の低い印象ページを除外するために，ステップ 2 で作成した各スニペットの特徴ベクトル$\boldsymbol{r_j}(j = 1, ...n)$と$\boldsymbol{e_{avg}}$とのコサイン類似度を計算し，しきい値 θ_r を下回った印象ページを除外する．
4. **ブロックの印象の判定**
   ブロックの内容に印象の観点から反対意見があるかを判定するためには，ブロックがポジティブな内容なのかネガティブな内容なのかを判定する必要がある．そのためにステップ 2 で作成したブロックの特徴ベクトル$\boldsymbol{v_{b_i}}$とポジティブ/ネガティブな印象ページ群の代表ベクトル$\boldsymbol{p_{avg}}$と$\boldsymbol{n_{avg}}$間のコサイン類似度を評価する．v_b とのコサイン類似度が大きい方の印象をブロックの印象と判定する．
5. **反対意見の提示**
   便宜上ステップ 4 により判定されたブロックの印象を A とし，それと対立する印象を B とする (A はポジティブかネガティブかのいずれかを取る)．ブロックの内容に対して反対意見が存在するかの判断は，ブロックと反対の印象を持つ内容がブロックの内容と比べて世の中で主流であるか否かが重要な要素となる．これを判定するための要素を考える．まず要素として印象 A を持つページ群と反対の印象 B を持つページ群の割合を考慮する．また 2 つの印象ページ群間に相違があるほど内容が対立，相違がそれほどなければ内容は対立的ではないと考えられる．そこで印象 A のページ群の平均ベクトルと印象 B のページ群の平均ベクトルとの非類似度 $DisSim$ を考慮に入れる．最終的に反対意見を提示すべきかどうかの基準である危険度 $Danger$ を以下のように定義する．

$$Danger = \frac{N'_B + 1}{N'_A + 1} DisSim(\boldsymbol{p_{avg}}, \boldsymbol{n_{avg}}) \quad (6)$$

$$= \frac{N'_B + 1}{N'_A + 1}(1 - Sim(\boldsymbol{p_{avg}}, \boldsymbol{n_{avg}})) \quad (7)$$

ここで N'_A，N'_B はステップ 3 によってブロックと関連の無いページを除外した印象 A，印象 B をもつページ群の大きさを表している．類似度の計算にはコサイン類似度を用いた．式 (6)(7) の右辺第一項の分母，分子に 1 が加算されているのは分母が 0 になることを防ぐためである．この危険度 $Danger$ を閲覧ページから抽出された全てのブロックに対し計算し，$Danger$ の値がしきい値 θ_d を超えたときブロックに対する反対意見が存在する旨の警告を発し，内容に対する反対意見を提示する．

反対意見として提示する内容はブロックの印象 A としたとき，印象 B を持つページ群の中に存在するが，全てを提示するのは適切ではない．すなわち反対意見として提示するに適切なページを選択する必要がある．反対意見として適切な文章とは反対意見内容を多く含む文章である．ステップ 3 で述べたように各印象ページは比較軸と印象を決定する要因とで構成されている．そこで印象を決定する要素を抽出し，それらを最も多く含むページを見つけることができれば，それを反対意見として提示することができる．印象を決定する要因は印象平均ベクトルと全体平均ベクトルの差分で求められる．今印象 B を決定づける要因を表す印象 B 差分ベクトル $\boldsymbol{diff_B}$ は以下の式で定義される．

$$\boldsymbol{diff_B} = \boldsymbol{B_{avg}} - \boldsymbol{e_{avg}} \quad (8)$$

B はポジティブ/ネガティブのいずれかの印象，$\boldsymbol{B_{avg}}$ は $\boldsymbol{p_{avg}}$，$\boldsymbol{n_{avg}}$ のいずれかを意味する．ここで印象差分ベクトル $\boldsymbol{diff_B}$ の各語の重みが印象を決定づける要因の大きさと考える．$\boldsymbol{diff_B}$ を構成する単語 t_i の重みを I_{t_i} とした時，印象 B を持つ検索結果ページ P の反対要因度 $Counter_P$ を以下のように定義する．

$$Counter_P = \sum_{t_i \in P} I_{t_i} \quad (9)$$

$t_i \in P$ はページ P 内にある単語でかつ $\boldsymbol{diff_B}$ を構成する単語である．式 (9) により印象を決定づける大きな要因となる語を多く含んでいるページを抽出する．$Counter_P$ の値を全ての印象 B をもつページ群に対して計算し，上位 N 個のページのスニペットと URL を反対意見として提示する．

4. ケーススタディ

本システムによって反対意見の存在が警告された例をいくつか示す．なお反対意見分析ではしきい値の設定として，$\theta_r = 0.3$, $\theta_{impress} = 0.7$，$\theta_d = 0.35$ を用いた．反対意見が存在した場合，$Counter_p$ の値が大きい上位 3 件のスニペットを反対意見として提示した．

一つ目のケーススタディとして納豆ダイエットを取り扱った Web ページ[3]に対して分析を行った例を示す．以下はこの Web ページ中の反対意見が存在すると実際に判定されたブロックと反対意見である．

ケーススタディ1-1

これらの納豆の成分はダイエットと美肌に、強い味方になってくれるでしょう。

本システムは上記文章の内容をポジティブと判定した．危険度 $Danger$ は 0.518 となった．この文章に対する反対意見としては以下が提示された．

- **僕は『納豆ダイエット』に対する疑惑を書きました。 それから数日後。 … いっても、それは納豆ダイエットへの疑いだけから出た思いではありません。 … 納豆いいよねー．女性の味方イソフラボンだしねー．大量に注文来て急にこんな状態になって …**
- **なんでも、某テレビ番組で、納豆がダイエット食材として紹介されたからしいけど、一人暮らしの味方納豆が、スーパーから消えるのは痛い。 それが、今日に限ってダイエット効果を紹介した情報番組で実験データの捏造発覚!!**

反対意見を見ると分かるように，納豆がダイエットに効果的である，という放送が捏造であったという文章が提示されており，「納豆の成分はダイエットと美肌に、強い味方になってくれる」という内容を再考するきっかけを与えてくる．ただし，美肌に対する反対意見にはなっていない．

ケーススタディ1-2

納豆ダイエットの納豆に含まれる菌を効率よく使うには、納豆を摂取するタイミングが大事です。

上記文章の内容はネガティブと判定された．危険度 $Danger$ は 0.460 となった．反対意見としては以下が提示された．

- **くめ納豆の「クール納豆茶漬け」 threepine (07/28) マクロビオティック … 簡単ダイエット、納豆ダイエットなど簡単なダイエット方法 … 病院が医学的にダイエットを効率よく行わせてくれるのです。 今まで …**
- **酢全般のダイエット効果について、ピンキーが解説します。 … またアミノ酸を摂取して、さらに運動をすると、脂肪の燃焼効率がアップします。 … キムチ納豆ダイエット．しょうが紅茶ダイエット．味噌汁ダイエット．酢ダイエット特集**

この例では提示された反対意見は適切ではない．この例では $Danger$ の値が低い．今回の実験では理想的なしきい値 θ_d を調べていないこともあり，適切に反対意見存在判定ができる値を探

[3] http://www.kyu-sapo.com/diet/dietjiten/nattou_diet.html

す必要がある．

二つ目のケーススタディとして「瞬間接着剤での傷口の治療」を取り扱った Web ページ[4]に対して分析を行った例を示す．以下はこの Web ページ中の反対意見が存在すると実際に判定されたブロックと反対意見である．

ケーススタディ2-1

傷口に瞬間接着剤っていいんですか？

本システムは上記文章の内容をネガティブと判定した．危険度 *Danger* は:0.356 となった．この文章に対する反対意見としては以下が提示された．

- **こんにちは、mi_ppi です。私の兄貴が、指の切り傷にアロンアルファ(瞬間接着剤) を塗って傷口を塞いでいました。 … アロンアルファは元々は医療用接着剤として作られた物だから … 接着剤系の仕事をしてるので、コメントします。**
- **さあ。そういえば、手術用のアロンアルファってあるよ。 何か違うんかなぁ。 … 瞬間接着剤が傷口に付いたときの毒性を教えてください。 さあ。 そういえば、手術用のアロンアルファってあるよ。 何か違うんかなぁ。**

この例では文章に対する印象判定に失敗したため，反対意見として提示されたものが逆の印象になっていない．これは入力となった文章長が短いため，コサイン類似度を用いた印象判定に失敗していると考えられる．印象判定方法も検討が必要である．

5. おわりに

本論文では，閲覧中の Web ページの内容に対する反対意見の存在判定，およびその提示に関する手法について提案した．反対意見の存在判定には文章の印象に着目し，判定対象がポジティブ (ネガティブ) な内容であり，かつそれに関連する Web 上の情報の大半がネガティブ (ポジティブ) な内容である場合に反対意見が存在するとし，自動的に警告及び反対意見の提示するシステム「WebAlert」を実装した．実験では本システムの根幹となる印象判定能力の評価，およびケーススタディとして幾つかの例文に対する反対意見存在の検証を行った．今後はより評価実験でも問題に挙がった印象判定能力の向上，および閲覧ページからのクエリの生成手法，印象分析以外の観点からの反対意見の存在判定手法について検討する．

[謝辞]

本研究は一部，グローバル COE 拠点形成プログラム「知識循環社会のための情報学教育研究拠点」，文科省研究委託事業「知的資産の電子的な保存・活用を支援するソフトウェア技術基盤の構築」，科研費：計画研究「情報爆発時代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」(課題番号 18049041)，NICT 委託研究「電気通信サービスにおける情報信憑性検証技術に関する研究開発」，若手研究 (B) 「ウェブ活用のための情報統合による信頼性判断支援」(課題番号：18700086)，および若手研究 (B) 「情報検索とウェブアーカイブにおけるマイニング」(課題番号：18700111) 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] L.Page, S.Brin, R.Motwani and T.Winograd: “The pagerank citation ranking: Bringing order to the web”, Technical report, Stanford Digital Library Technologies Project (1998).

[2] S.Nakamura, S.Konishi, A.Jatowt, H.Ohshima, H.Kondo, T.Tezuka, S.Oyama and K.Tanaka: “Trustworthiness analysis of web search results”, ECDL2007, pp. 38–49 (2007).

[3] Y.Yamamoto, T.Tezuka, A.Jatowt and K.Tanaka: “Honto? search: Estimating trustworthiness of web information by search results aggregation and temporal analysis”, APWeb/WAIM2007, pp. 253–264 (2007).

[4] 熊本忠彦, 灘本明代, 田中克己：“記事の印象を伝達するニュース番組生成システム wee の設計と評価”, 電子情報通信学会論文誌, **J90-D**, 2, pp. 185–195 (2007).

[5] D.Cai, S.Yu, J.-R. Wen and W.-Y. Ma: “Block-based web search”, SIGIR ’04: Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, New York, NY, USA, ACM, pp. 456–463 (2004).

山本 祐輔 Yusuke YAMAMOTO
京都大学大学院情報学研究科博士後期程在学中．2008 年京都大学大学院情報学研究科修士課程終了．情報の信頼性，Web マイニングの研究・開発に従事．情報処理学会，電子情報通信学会，日本データベース学会学生会員．

手塚 太郎 Taro TEZUKA
立命館大学総合理工学院情報理工学部メディア情報学科講師．2005 年京都大学大学院情報学研究科社会情報学専攻博士後期課程修了．博士（情報学）．主に地域情報検索システム，ウェブからの知識発見，検索システムの教育への応用の研究に従事．情報処理学会，電子情報通信学会，日本データベース学会各会員．

アダム ヤトフト Adam JATOWT
京都大学大学院情報学研究科社会情報学専攻助教．2005 年東京大学大学院情報理工学系研究科電子情報学博士後期課程修了．博士（情報学）．主にウェブ検索，ウェブアーカイブマイニングの研究に従事．ACM 会員．

田中 克己 Katsumi TANAKA
京都大学大学院情報学研究科社会情報学専攻教授．1976 年京都大学大学院修士課程修了．博士（工学）．主にデータベース，マルチメディアコンテンツ処理の研究に従事．IEEE Computer Society，ACM，人工知能学会，日本ソフトウェア科学会，情報処理学会，日本データベース学会等各会員．

[4] http://oshiete1.goo.ne.jp/qa238135.html