社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2008－ＦＩ－ 89 (8)
2008－ＮＬ－183 (8)
2008／1／21

意見分析タスク - 多言語テキストを対象とした意見抽出技術の評価 -

関　洋平[†]　David Kirk Evans[††]　Hsin-Hsi Chen[†††]
Lun-Wei Ku[†††]　神門典子[††]

われわれは，NTCIR-6 において意見分析タスクを開催し，約 30 の共通のトピックに適合した日本語，英語，中国語の新聞記事を対象に，15,279, 8,417, 11,909 文に対して意見情報を付与した研究資源を作成した．タスクでは，6 カ国 14 チーム（2 チームは 2 言語のタスクに参加）からの参加があり，提出された 21 の結果について，意見文判定，意見ホルダ判定，適合文判定，極性判定などの抽出技術について評価を行った．本稿では，主に，参加システムの特徴と評価結果の傾向，言語ごとに採用されるアプローチの違いについて議論する．

Opinion Analysis Task - Evaluation on Opinion Extraction Technologies for Multilingual Texts -

Yohei Seki,[†] David Kirk Evans,[††] Hsin-Hsi Chen,[†††]
Lun-Wei Ku[†††] and Noriko Kando[††]

This paper describes an overview of *the Opinion Analysis Task* from 2006 to 2007 at the Sixth NTCIR Workshop. We created test collection for 30, 28, and 32 topics (15,279, 8,417, and 11,909 sentences) in Japanese, English and Chinese. Using this test collection, we conducted opinion extraction subtask. The subtask was defined from four perspectives: (a) opinionated sentence judgment, (b) opinion holder extraction, (c) relevance sentence judgment, and (d) polarity judgment. 21 run results were submitted by 14 participants. We discuss the evaluation results based on the approach the participants took and the different tendency by languages.

1. はじめに

われわれは，2006 年の NTCIR-6 ワークショップ以降，意見抽出の技術を評価するためのタスクを開催しており，日本語・英語・中国語のテキストを対象として，意見分析のための研究資源を提供している．本稿では，タスクの概要とタスクから得られた知見を中心に説明する．

自然言語処理の研究において，意見や感情の分析の研究はここ数年盛んに行われている [1)~4)]．ソーシャルネットワーク，ブログなどの，ユーザが作成するコンテンツ，あるいは，ニュースを対象として自動分析を行うことで，Web 上に表出する評判や世論の傾向をモニターする手法が，企業や政府機関の興味を集めている．このような背景から，意見性の有無 [5)]，極性の判定 [6)]，意見ホルダの抽出 [7)] といった研究が注目を集めている．また，意見の動向をテキストマイニング技術に基づき可視化したり，潜在的な悪評やスキャンダルを明らかにするための傾向を分析するといったアプリケーションに関する研究も行われている．

NTCIR-6 では，日本語，英語，中国語を対象とした意見分析のためのタスクを新たに開催した．本稿では，テストコレクションの概要，タスク設計などについて解説し，評価結果から得られた知見を明らかにする．本タスクは，意見分析に関する研究を，言語を横断して行うための唯一の機会を提供すると考えている．

† 豊橋技術科学大学
Toyohashi University of Technology
†† 国立情報学研究所
National Institute of Informatics
††† 国立台湾大学
National Taiwan University

文書は，言語横断検索タスクの適合性判定を元に選択されており，3 言語とも共通のトピックに対する適合文書を利用している．

本稿の構成は以下の通りである．2 章では，タスク概要について解説を行う．3 章では，NTCIR-6 で作成したテストコレクションについて説明する．4 章と 5 章では，参加者のシステムの傾向と，評価結果について議論する．最後に，6 章で結論を述べる．

2. タスク設計

2.1 4 つの評価属性

NTCIR-6 の意見分析タスクでは，**表 1** に示す 4 つの評価属性に対する意見抽出サブタスクを設定した．以下，4 つの意見属性について説明する．

表 1 NTCIR-6 意見分析タスクの 4 つの評価属性

属性	値	参加者への提出要求
意見文判定	有，無	必須
意見ホルダ判定	文字列	必須
適合文判定	有，無	任意
極性判定	肯定，否定，中立	任意

(1) 意見文判定
文を単位として意見性の有無を判定した．

(2) 意見ホルダ判定
意見性があると判定された文について，意見のホルダ（意見を表明または保有している者）は誰（人，組織，国など）であるかを判定した．

(3) 適合文判定
与えられたトピック記述にその文が適合しているか，すべての文について判定した．

(4) 極性判定
意見性があると判定された文について，肯定 (POS)，否定 (NEG)，中立 (NEU) のいずれかについて判定を行った．

2.2 訓練データ

NTCIR-6 の意見分析タスクでは，日本語，中国語では共通の 4 トピック分の文書を訓練データとして参加者に提供した．英語については，*MPQA* 意見コーパス[8] を訓練データとし，1 トピックのアノテーションをサンプルデータとして参加者に提供した．

2.3 評価尺度

意見文判定，意見ホルダ判定，適合文判定，極性判定のそれぞれについて，文を基本単位とした精度，再現率，F-値を計算した．判定の基準データとしては，判定者 3 名が判定した結果のうち 3 名とも一致した結果（Strict/厳密に一致）と 2 名だけが一致した結果（Lenient/緩やかに一致）の双方を評価基準とした．

2.3.1 意見文判定・適合文判定

意見文判定と適合文判定については，精度は，$\frac{\text{システムと判定者判定が一致した数}}{\text{システムの判定総数}}$，再現率は，$\frac{\text{システムと判定者判定が一致した数}}{\text{意見文（適合文）の総数}}$ として計算した．ただし，意見ホルダと極性の評価については，判定者間の判定の一致についての解釈に応じて，3 つの異なる評価戦略を採用した．以下，それぞれの評価方法について説明する．なお，異なる評価戦略を採用した場合，システム間の順位については大きな差が現れていないことが明らかになっている（詳細は 9) を参照）．

2.3.2 ホルダ抽出評価

(1) YS 評価法
YS 評価法では，意見ホルダの抽出も，意見文判定や適合文判定と同様に，2 人または 3 人の判定者間の判定が完全に一致した要素を正解として判定を行った．また，システムと正解との一致についての判断基準は，以下の 5 段階の評価のうち，3 段階目までを一致と判定して評価した．正解判定の単位は文とした．
- (a) 意味的にも文字列としてみてもほぼ完全に一致．
- (b) 意味的にはほぼ一致，文字列的には部分的に一致，固有名は特定できていない．
- (c) 意味的に一致しているが，文字列的には不一致．
- (d) 意味的には身分・立場など部分・側面的に一致しているが，固有名は特定されていない．
- (e) 一致していない．

(2) DKE 評価法
DKE 評価法では，ホルダについての判定者間の判定一致については考慮せずに，判定者 3 名が抽出したホルダの総和との一致について評価した．ホルダの有無の判定基準は，意見性についての 2 人の判定一致，3 人の判定一致を基にして判定した．正解判定の単位は文とした．

(3) LWK 評価法
LWK 評価法では，DKE 評価法同様，判定者間の判定一致については考慮せずに，判定者 3 名が抽出したホルダの総和との一致について評価した．また，ホルダとしての抽出対象は同一文中の要素に制限しており，照応解決は考慮していない．正解判定の単位は，文とホルダそれぞれについて評価した．

表 2　NTCIR-6 意見分析タスクのテストコレクションのサイズ

言語	トピック数	文書数	文の数	意見文の割合 (2 人判定一致 / 3 人判定一致)	適合文の割合 (2 人判定一致 / 3 人判定一致)
日本語	30	490	15,279	29% / 22%	64% / 49%
英語	28	439	8,417	30% / 7%	69% / 37%
中国語	32	843	11,909	62% / 25%	39% / 16%

2.3.3 極性判定

(1) YS 評価法

YS 評価法では，極性判定についても，意見文判定や適合文判定や意見ホルダと同様に，2 人または 3 人の判定者間で判定が完全に一致した要素を正解として判定を行った．母数は文の総数とした．

(2) DKE 評価法

DKE 評価法では，極性判定については，評価した判定者の数が 3 人か，2 人か，1 人かによって重みをつけた．評価尺度の基準については，意見ホルダ同様に意見性の判定の一致を基にして判定した．母数は文の総数とした．

(3) LWK 評価法

LWK 評価法では，極性判定について，2 人または 3 人の判定が完全に一致または部分的に一致（肯定と中立ならば肯定とする等）した要素を正解として判定を行った．上記 2 つの評価法と異なり，母数は意見文数とした．

3. テストコレクション

3.1 文書集合

NTCIR-6 意見分析タスクの評価のためのテストコレクションは，NTCIR-3, 4, 5 の言語横断検索タスク [10] において使用された約 30 のトピック（話題）について適合と判定された文書を対象として作成した．文書数は，日本語，英語については 1 つのトピックあたり最大 20 文書とした．文書ジャンルは，1998 年から 2001 年にかけて発行された新聞記事を対象とした．テストコレクションのサイズとしてのトピックの数，文書数，文の数，意見文の割合，適合文の割合を**表 2** に示す．また，**表 3** にテストコレクションで利用したトピックの一覧を示す．

3.2 意見情報のアノテーション

概　　要

NTCIR-6 意見分析タスクでは，意見分析の先行研究 [11)~13)] を参考にしつつ，アノテーションを行った．日本語，英語に関しては，3 人の判定者がアノテーションを行った．中国語に関しては，7 人の判定者がトピックを分担し，トピックごとに 3 人分の判定結果を提供した．日本語，英語では，判定者間での無用の判定のゆれを避けるために，ひとつのトピックを用いて，判定者間でそれぞれ 6 時間，4 時間程度の協議を行い，その結果に基づき，残りのトピックについてアノテーションを行った．なお，日英の判定者は，雑誌の編集や翻訳など言葉の取り扱いを業務としている一方で，中国語の判定者は大学院生が中心であった．判定者間の一致率について，**表 4** に示す．

表 3　意見分析タスクのトピックの見出し

番号	見出し
1	タイムワーナー，アメリカ・オンライン（AOL），合併，影響
2	ベルー大統領，アルベルト・フジモリ，スキャンダル，賄賂
3	金大中，金正日，南北首脳会談
4	米国防長官，ウィリアム・セバスチャン・コーエン，北京
5	沖縄 G8 サミット
6	ウェン・ホー・リー，機密情報，国家安全保障
7	イチロー，新人王，大リーグ
8	ジェニファー・カプリアティ，テニス
9	EP-3 偵察機，F-8 戦闘機，飛行機衝突
10	歴史教科書論争，第二次世界大戦
11	たばこ会社，告訴，賠償金
12	タイガー・ウッズ，スポーツ界のスター
13	「秋闘」（秋の闘い），要求，労働者，抗議，台湾
14	専門家，意見，国際通貨基金（IMF），アジア諸国
15	ティーンエージャー，社会問題
16	離婚，家族の不和，批判
17	中国，反応，台湾，外交関係
18	中国，駐留，兵器，台湾
19	動物クローン技術
20	セクハラ，訴訟
21	オリンピック，わいろ，疑惑
22	北朝鮮，テポドン，アジア，対応
23	WTO への加入
24	中華航空機墜落
25	台湾省の再組織化
26	欧州通貨統合の経済的影響
27	金大中大統領の対アジア政策
28	クリントンのスキャンダル
29	戦争犯罪訴訟
30	原子力に対する抗議
31	大学入試政策
32	青年のためのカウンセリング

判定者に対する指示（日本語）

日本語についての，4 つの意見属性に対する意見情報のアノテーションの戦略を以下に示す．

(1) 意見文判定

判定者によるアノテーションには，13) を参考

表 4　判定者による判定のペアの平均一致率

言語	判定属性	Kappa 平均
日本語	意見文判定	0.6740
日本語	適合文判定	0.5415
日本語	極性判定	0.6153
英語	意見文判定	0.2940
英語	適合文判定	0.3854
英語	極性判定	0.2749
中国語	意見文判定	0.2992
中国語	適合文判定	0.3859
中国語	極性判定	0.7334

に，以下の 3 つの基準を採用した．

(a)　“怖い”，“悲しい”，“つまらない” などの述語を用いた心的状態の明示的な記述

(b)　“言った”，“書いた” 等の発話/記述イベント

(c)　(a)，(b) のような明示的な手がかりを用いずに，主観を表現する要素

以下のような要素は意見性が無いと判定した．

- “特定の話者の発言ではなく，間接的な伝聞”，“一般的な見通し・予想を述べたもの” は，普遍的な事実性を有すると考え，意見性はないことにする．
  例：“すでに地中海特有の長い昼休みの習慣もビジネス界では薄れてきているという．”，“同じ車でも、欧州間で値段に大きな違いがあることは以前から指摘されていた．”，“１１カ国の一大通貨圏が誕生する見通しだ．”
- 政府・国家などの公式発表・宣言は，“単なる「計画」「予定」を述べたもの”，“組織・団体の「宣言文」のようなもの” は，「意見文」とはしない．
- 記事の見出し，小見出しについても，意見性を判定．
- 署名，図，表は意見性が無いと判定．

(2)　意見ホルダ判定

判定者によるアノテーションには，以下の 3 つのタイプを区別して判定した．

(a)　心的状態を表明した人または組織など

(b)　発話/記述イベントの主体

(c)　主観を表現する主体（書き手/話し手など）

ホルダの抽出には，以下の規則を設定した．

- 意見保有者を意味する明示的な名詞表現が同一文書内に存在するが，同一文中では代名詞の表現をとっている場合，“指示代名詞”（“名詞表現”）のように記入する．
- 意見保有者を指示する内容が同一文書内にも無いが，実世界の表現として存在し，文中では代名詞の表現をとっている場合，“指示代名詞” < 実世界表現 > のように記入する．
- 指示表現には “同”，“同誌” 等の表現も含む．
- 指示表現がゼロ代名詞の場合には，(“名詞表現”) や < 実世界表現 > のように記入する．
- 記事の著者が意見保有者となる場合，署名を利用して記入する．
- 意見保有者の直前・直後に存在する国名・所属なども合わせて抽出する（国名・所属などの立場を考慮したアプリケーションに対応するため）．
- 所属・国名への連体修飾表現は抽出しない．
- 意見保有者を直接修飾する数量句表現は抽出．

(3)　適合文判定

すべての文について，与えられたトピック記述にその文が適合しているかを，文脈は考慮せず，その文の情報だけで判定した．

(4)　極性判定

意見性があると判定された文について，極性（肯定，否定，中立）の判定を行った．文内の節で極性が異なるものについては，以下の戦略で判定を行った．

- 極性が節の単位で反転しているものについては，文単位の極性を付与した後，() で囲み，節を単位とした極性を; で区切って付与する．
  例：NEU (POS;NEG)
- 一文内の文節ごとに極性が異なる場合，主節の極性を優先する．

4. 参加者のシステムの傾向

参加者のシステムについて，サブタスクごとの戦略の概要を，**表 5** に示す．以下では，サブタスクごとに，(1) 言語に応じてとられる戦略の違いと，(2) 注目すべき特徴的なシステムについて紹介する．また，サブタスクと独立に興味深い研究についても紹介する．

4.1 意見文判定

意見文判定については，日英のほとんどの参加システムが，SVM による機械学習アプローチを用いている．一方で，中国語の参加チームは辞書をベースにした判定手法が多い．なお，中国語については，辞書の

表 5　参加者のシステム一覧

チーム	ID	言語	意見文判定	意見ホルダ抽出	極性判定
NEC	EHBN	日本語	SVM	固有表現利用+著者分類ルール	–
情報通信研究機構	NICT	日本語	SVM	固有表現利用ルール	SVM+多数決投票
豊橋技術科学大学	TUT	日本語	SVM	SVM（著者分類）+ 固有表現利用ルール	SVM+ルール
コーネル大学	Cornell	英語	SVM	CRF による情報抽出	SVM+ルール
シェフィールド大学	GATE	英語	SVM	SVM による情報抽出	–
韓国情報通信大学	ICU-IR	英語	半教師あり学習（SVM）	固有表現利用+著者分類ルール	辞書
イリノイ工科大学	IIT	英語	辞書/SVM	ヒューリスティックルール	辞書/SVM+多数決投票
国立情報学研究所	NII	英語	SVM	固有表現利用+Logistic モデル	k-NN
豊橋技術科学大学	TUT	英語	SVM	SVM（著者分類）+ 固有表現利用ルール	SVM+ルール
香港中文大学	CUHK	中国語	教師なし学習（辞書）+ SVR	固有表現利用ルール	辞書
シェフィールド大学	GATE	中国語	SVM	SVM による情報抽出	–
中国科学院ソフトウェア研究所	ISCAS	中国語	辞書	CRF による情報抽出	辞書
国立台湾大学	NTU	中国語	辞書	固有表現利用ルール	辞書
メリーランド大学	UMCP	中国語	辞書	固有表現利用ルール	辞書

リソースをオーガナイザが提供している．注目すべき特徴的なシステムとして，以下のものがある．

- イリノイ工科大学は，Appraisal Theory ☆を基盤とした辞書作成による判定結果を提出した．
- 韓国情報通信大学は，基本ルール 6 つ（86 の手がかり語を利用）により意見性があると判定された文中に現れたすべての動詞，形容詞，副詞を収集し，SVM の特徴素として意見性を判定する半教師あり学習手法を提案した．
- 香港中文大学では，トピックに適合する Web 文書を独自に収集し，教師なし学習により辞書を作成した後，Support Vector Regression モデルにより意見文を判定する手法を提案した．

4.2 意見ホルダ抽出

意見ホルダの抽出は，日本語，英語では，著者などの文外要素，照応要素などをホルダの正解として含んでいるため，ホルダとして抽出するためのルールを組み込んでいる．一方で，中国語では，ホルダの要素を文中の要素に制限しているため，参加者はそのようなルールは設定していない．アプローチは，(1) 情報抽出の問題として定義する手法と，(2) 固有表現を手がかりとして抽出する手法に大別される．以下，(1) のタイプのシステムを中心に説明する．

- コーネル大学は，文献 7) で最初にホルダ抽出を提案した Choi（参加メンバの一人）が実現した CRF ベースの抽出手法を基本とし，AutoSlog を用いて抽出されたパターンを特徴素とするシステムにより，結果を提出した．
- 中国科学院ソフトウェア研究所は，中国語において Choi らの提案を実現した．
- シェフィールド大学は，SVM を用いて，開始トークンと終了トークンをそれぞれ 2 値判定で決定するという手法を提案した．

また，著者の分類について，豊橋技術科学大学は，著者の意見文判定と非著者の意見文判定について SVM 分類器をそれぞれ独立に実現し，ホルダの抽出に利用するという手法を提案した．

4.3 極性判定

極性判定のアプローチは，(1) 複数の SVM による判定結果をルールや多数決投票で組み合わせる手法と，(2) 辞書ベースの方法に大別される．中国語はすべての参加チームが (2) の手法を採用している．英語では，韓国情報通信大学（WordNet などを用いたシードの拡張），イリノイ工科大学（Appraisal Theory に基づいた極性判定）といった辞書を丁寧に作成したチームが好成績を収めている．

4.4 その他

その他，着目すべき研究に以下のものがある．

- 複数の言語について結果を提出したのは，シェフィールド大学（英語，中国語），豊橋技術科学大学（日本語，英語）の 2 チームであり，それぞれ言語間の比較を行っている．
- シェフィールド大学は，MPQA コーパスと NTCIR-6 英語意見分析コーパスの比較を行っており，同じコーパスを訓練データとした方が精度が向上することなどを明らかにした．
- 国立情報学研究所は，WEKA において実現されている複数の分類器を比較して戦略を決定した．

☆ http://grammatics.com/appraisal

5. 評価結果

本節では，参加者の評価結果について概説する．2.3節で述べたとおり，ホルダ抽出，極性判定について，われわれは3つの評価手法を提案しているが，本節では，言語ごとに1つの評価手法の結果だけを示す．すべての評価結果は，文献9) を参照されたい．

5.1 日本語

表6 に，日本語意見分析タスクの参加システムの評価結果を示す．意見判定の基準は，L（Lenient, 判定者3人中2人の判定が一致したものが正解）と S（Strict, 判定者3人中3人の判定が一致したものが正解）の2つからなる．チームIDの後ろの数字は，そのチームの RunID を示す．P, R, F は，精度，再現率，F-値を示す．結果について以下にまとめる．

- 意見文判定については，NICT がよい精度を得た．TUT は再現率において優れていた．
- 意見ホルダ抽出については，EHBN-2 がよい精度を得た．TUT は再現率において優れていた．
- 適合文判定については，NICT-2 がよい精度を得た．NICT-1 が再現率において優れていた．
- 極性判定については，NICT がよい精度を得た．TUT は再現率において優れていた．

すなわち，EHBN は，ホルダ抽出の結果が優れていた．NICT は，全体的にバランスの取れた精度のよいシステムを実現した．TUT は再現率に焦点をおいたシステムを実現しており，F-値が優れていた．

5.2 英語

表7 に，英語意見分析タスクの参加システムの評価結果を示す．意見判定基準の L と S は，意見文判定と適合文判定についての解釈は日本語と同じであるが，意見ホルダ抽出と極性判定については，2.3節で述べたとおり，意見文の判定一致を基準としている点に注意されたい．結果について以下にまとめる．

- 意見文判定については，ICU-IR がよい精度を得た．GATE は再現率において優れていた．
- 意見ホルダ抽出については，ICU-IR が精度，再現率共に優れており，IIT，Cornell が続いた．
- 適合文判定については，NII がよい精度を得た．GATE，TUT は再現率において優れていた．
- 極性判定については，ICU-IR がよい精度を得た．IIT は再現率において優れていた．

すなわち，ICU-IR が全体的にバランスの取れた精度のよいシステムを実現した．彼らは配布したサンプル1トピックについて，よく分析を行った上で，システムを実現した．また，IIT が辞書ベースながらホルダ抽出，極性判定で優れていた．その他，意見文判定では GATE，適合文判定では TUT, NII が優れていた．

5.3 中国語

表8 に，中国語意見分析タスクの参加システムの評価結果を示す．意見判定基準の L と S の解釈は，英語と同様である．また，極性の母数は意見文数としているため，判定項目名を意見文内極性判定としている点に注意されたい．結果について以下にまとめる．

- 意見文判定については，CUHK, Gate-2 がよい精度を得た．UMCP, Gate-1 は再現率において優れていた．
- 意見ホルダ抽出については，CUHK が精度，再現率共に優れており，ISCAS が続いた．
- 適合文判定については，CUHK がよい精度を得た．NTU，UMCP-2 は再現率において優れていた．
- 極性判定については，CUHK がよい精度を得た．NTU, UMCP は再現率において優れていた．

すなわち，CUHK が全体的にバランスの取れた精度のよいシステムを実現した．彼らはトピックに適合する Web 文書を収集し，中国語の意見分析において特に重要な辞書の拡張に力を入れた．また，ISCAS は，CRF ベースの方法でホルダ抽出が優れていた．UMCP, Gate, NTU は，意見文判定が優れていた．

6. まとめ

NTCIR-6 意見分析タスクでは，これまで眺めてきたとおり，さまざまな評価結果が出ているものの，サブタスクごと，言語ごとに，ある程度の傾向が観察できた．意見文判定，適合文判定については，言語によるアプローチの違いこそあるものの，すべてのチームにおいて，一定の評価結果が得られている．一方，意見ホルダ抽出，極性判定については，言語，参加チームごとに評価結果のばらつきが見られ，意見文判定ほどは技術が成熟していないことが窺える．その中でも，辞書の作成やコーパスの分析を丁寧に行ったチームが，比較的優れた結果を出していることがわかる．

オーガナイザ側として難しい点は評価であった．特に，評価の基準となる複数の判定者間の一致率の向上ならびに，評価尺度における一致の取り扱いについては，議論の余地がある．一方で，極性について，3種類の異なる評価尺度を設定したものの，システムのランキングはそれほど変化しないこともわかっており[9]，さらなる分析を通して適切な基準を明らかにしたい．

今後の目標として，NTCIR-6 では2チームであった同一参加者による多言語のタスクの同時参加をより推進していくこと，文書ジャンルを，より意見文を

表 6 日本語意見分析タスクの YS 評価法による評価

チーム	L/S	意見文判定			ホルダ抽出 (S/A/B/C/D/OE/LE)			適合文判定			極性判定		
		P	R	F	P	R	F	P	R	F	P	R	F
EHBN-1	L	0.531	0.453	0.489	0.138	0.085	0.105 (224/46/6/34/806/880/2129)	-	-	-	-	-	-
EHBN-2	L	0.531	0.453	0.489	0.314	0.097	0.149 (236/39/41/77/321/293/2531)	-	-	-	-	-	-
NICT-1	L	0.671	0.315	0.429	0.238	0.102	0.143 (86/0/246/224/378/462/2311)	0.598	0.669	0.632	0.299	0.149	0.199
NICT-2	L	0.671	0.315	0.429	0.238	0.102	0.143 (86/0/246/224/378/462/2311)	0.644	0.417	0.506	0.299	0.149	0.199
TUT	L	0.552	0.609	0.579	0.226	0.224	0.225 (472/137/118/134/1006/1354/1378)	0.630	0.646	0.638	0.274	0.322	0.296
EHBN-1	S	0.414	0.479	0.444	0.079	0.094	0.086 (128/28/2/22/405/1411/1095)	-	-	-	-	-	-
EHBN-2	S	0.414	0.479	0.444	0.183	0.110	0.137 (130/25/29/31/166/626/1299)	-	-	-	-	-	-
NICT-1	S	0.546	0.348	0.425	0.133	0.110	0.120 (73/0/112/104/214/893/1177)	0.470	0.693	0.560	0.168	0.150	0.158
NICT-2	S	0.546	0.348	0.425	0.133	0.110	0.120 (73/0/112/104/214/893/1177)	0.525	0.446	0.482	0.168	0.150	0.158
TUT	S	0.414	0.620	0.497	0.131	0.251	0.172 (292/68/61/63/501/2236/695)	0.505	0.681	0.580	0.161	0.339	0.218

S/A/B/C/D = 5 段階評価
OE = 過剰推定
LE = 推定欠如

表 7 英語意見分析タスクの DKE 評価法による評価

チーム	L/S	意見文判定			ホルダ抽出			適合文判定			極性判定		
		P	R	F	P	R	F	P	R	F	P	R	F
IIT-1	L	0.325	0.588	0.419	0.198	0.409	0.266	—	—	—	0.120	0.287	0.169
IIT-2	L	0.259	0.854	0.397				—	—	—	0.086	0.376	0.140
TUT-1	L	0.310	0.575	0.403	0.117	0.218	0.153	0.392	0.597	0.473	0.088	0.215	0.125
TUT-2	L	0.310	0.575	0.403				0.392	0.597	0.473	0.094	0.230	0.134
Cornell†	L	0.317	0.651	0.427	0.163	0.346	0.222	—	—	—	0.073	0.197	0.107
NII	L	0.325	0.624	0.427	0.066	0.166	0.094	0.510	0.322	0.395	0.077	0.194	0.110
GATE-1	L	0.324	0.905	0.477	0.121	0.349	0.180	0.286	0.632	0.393	—	—	—
GATE-2	L	0.324	0.905	0.477				0.286	0.632	0.393	—	—	—
ICU-IR	L	0.396	0.524	0.451	0.303	0.404	0.346	0.409	0.263	0.320	0.151	0.264	0.192
IIT-1	S	0.070	0.578	0.125	0.054	0.461	0.097	—	—	—	0.027	0.322	0.049
IIT-2	S	0.056	0.840	0.105				—	—	—	0.016	0.359	0.031
TUT-1	S	0.065	0.553	0.117	0.029	0.241	0.051	0.171	0.605	0.266	0.016	0.195	0.029
TUT-2	S	0.065	0.553	0.117				0.171	0.605	0.266	0.019	0.229	0.034
Cornell†	S	0.069	0.662	0.125	0.041	0.392	0.074	—	—	—	0.010	0.135	0.018
NII	S	0.073	0.642	0.131	0.018	0.169	0.032	0.242	0.355	0.287	0.014	0.185	0.027
GATE-1	S	0.070	0.940	0.130	0.029	0.398	0.055	0.112	0.579	0.188	—	—	—
GATE-2	S	0.070	0.940	0.130				0.112	0.579	0.188	—	—	—
ICU-IR	S	0.102	0.616	0.175	0.085	0.515	0.146	0.177	0.266	0.213	0.034	0.301	0.061

表 8 中国語意見分析タスクの LWK 評価法による評価

チーム	L/S	意見文判定			意見文内ホルダ抽出 (文単位/ホルダ単位)			適合文判定			意見文内極性判定		
		P	R	F	P	R	F	P	R	F	P	R	F
CUHK	L	0.818	0.519	0.635	0.647/0.742	0.754/0.932	0.697/0.826	0.797	0.828	0.812	0.522	0.331	0.405
ISCAS	L	0.590	0.664	0.625	0.458/0.516	0.405/0.445	0.430/0.478	—	—	—	0.232	0.261	0.246
Gate-1	L	0.643	0.933	0.762	0.427/0.525	0.154/0.171	0.227/0.258	—	—	—	—	—	—
Gate-2	L	0.746	0.591	0.659	0.373/0.398	0.046/0.042	0.082/0.076	—	—	—	—	—	—
UMCP-1	L	0.645	0.974	0.776	0.241/0.297	0.410/0.429	0.303/0.351	0.683	0.516	0.588	0.292	0.441	0.351
UMCP-2	L	0.630	0.984	0.768	0.221/0.272	0.376/0.393	0.278/0.321	0.644	0.936	0.763	0.286	0.446	0.348
NTU	L	0.664	0.890	0.761	0.652/0.745	0.172/0.169	0.272/0.276	0.636	1.000	0.778	0.335	0.448	0.383
CUHK	S	0.341	0.575	0.428	0.707/0.794	0.785/0.806	0.744/0.800	0.468	0.900	0.616	0.197	0.596	0.296
ISCAS	S	0.221	0.662	0.331	0.470/0.527	0.406/0.456	0.436/0.489	—	—	—	0.059	0.314	0.099
Gate-1	S	0.253	0.979	0.402	0.419/0.517	0.156/0.175	0.227/0.262	—	—	—	—	—	—
Gate-2	S	0.330	0.696	0.448	0.368/0.397	0.052/0.048	0.091/0.086	—	—	—	—	—	—
UMCP-1	S	0.245	0.986	0.393	0.293/0.357	0.438/0.453	0.351/0.400	0.404	0.565	0.471	0.085	0.615	0.150
UMCP-2	S	0.239	0.993	0.385	0.274/0.333	0.410/0.423	0.329/0.373	0.354	0.953	0.516	0.081	0.604	0.143
NTU	S	0.258	0.921	0.404	0.661/0.760	0.177/0.175	0.279/0.284	0.343	1.000	0.511	0.104	0.662	0.180

多く含むブログ等に変更することが挙げられる．また，評価属性についても，意見の対象（target）や強さ（strength）などについて拡張していきたいと考えている．文より細かい節の単位の取り扱いについても検討していきたい．

謝　辞

NTCIR-6 意見分析タスクの参加者の方々，ならびにタスクの開催にご尽力いただいた NTCIR 事務局のスタッフの皆様に深謝します．この研究の一部は，文部科学省科学研究費補助金若手研究 (B) (課題番号 18700241) を受けて遂行された．

参 考 文 献

1) 乾孝司, 奥村学: テキストを対象とした評価情報の分析に関する研究動向, 自然言語処理, Vol. 13, No. 3, pp. 201–241 (2006).
2) Gamon, M. and Aue, A.: *Proc. of Wksp. on Sentiment and Subjectivity in Text at the 21th Int'l Conf. on Computational Linguistics / the 44th Ann. Meeting of the Assoc. for Computational Linguistics (COLING/ACL 2006)*, The Association for Computational Linguistcs, Sydney, Austraria (2006).
3) Shanahan, J. G., Qu, Y. and Wiebe, J.: *Computing Attitude and Affect in Text: Theory and Applications*, The Information Retrieval Series, Vol. 20, Springer-Verlag, New York (2005).
4) National Institute of Standars and Technology: TREC (Text REtrieval Conference) 2006-2007: BLOG Track [online], *TREC website* (2006). [cited 2007-1-26]. Available from: <http://trec.nist.gov/tracks.html>.
5) Wiebe, J. M., Wilson, T., Bruce, R. F., Bell, M. and Martin, M.: Learning Subjective Language, *Computational Linguistics*, Vol. 30, No. 3, pp. 277–308 (2004).
6) Wilson, T., Wiebe, J. and Hoffmann, P.: Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis, *Proc. of the 2005 Human Language Technology Conf. and Conf. on Empirical Methods in Natural Language Processing (HLT/EMNLP 2005)*, Vancouver, B. C. (2005).
7) Choi, Y., Cardie, C., Riloff, E. and Patwardhan, S.: Identifying Sources of Opinions with Conditional Random Fields and Extraction Patterns, *Proc. of the 2005 Human Language Technology Conf. and Conf. on Empirical Methods in Natural Language Processing (HLT/EMNLP 2005)*, Vancouver, B. C. (2005).
8) Wiebe, J. M., Breck, E., Buckley, C., Cardie, C., Davis, P., Fraser, B., Litman, D., Pierce, D., Riloff, E. and Wilson, T.: MPQA: Multi-Perspective Question Answering Opinion Corpus Version 1.2 (2006). [cited 2007-1-26]. Available from: <http://www.cs.pitt.edu/mpqa/databaserelease/>.
9) Seki, Y., Evans, D. K., Ku, L. W., Chen, H. H., Kando, N. and Lin, C. Y.: Overview of Opinion Analysis Pilot Task at NTCIR-6, *Proc. of the Sixth NTCIR Wksp on Evaluation of Information Access Technologies: Information Retrieval, Question Answering, and Cross-Lingual Information Access*, pp. 265–278 (2007).
10) National Institute of Informatics: NTCIR CLIR Task [online], *NTCIR* (2006). [cited 2007-1-26]. Available from: <http://homepage3.nifty.com/kz_401/>.
11) Ku, L. W., Wu, T. H., Lee, L. Y. and Chen, H.H.: Construction of an Evaluation Corpus for Opinion Extraction, *Proc. of the Fifth NTCIR Wksp. on Evaluation of Information Access Technologies: Information Retrieval, Question Answering, and Cross-Lingual Information Access*, pp. 513–520 (2005).
12) Seki, Y., Eguchi, K. and Kando, N.: Multi-Document Viewpoint Summarization Focused on Facts, Opinion and Knowledge, *Computing Attitude and Affect in Text: Theory and Applications* (Shanahan, J. G., Qu, Y. and Wiebe, J.(eds.)), The Information Retrieval Series, Vol. 20, Springer-Verlag, New York, chapter 24, pp. 317–336 (2005).
13) Wiebe, J., Wilson, T. and Cardie, C.: Annotating Expressions of Opinions and Emotions in Language, *Language Resources and Evaluation*, Vol. 39, No. 2-3, pp. 165–210 (2005).