

レビューにおける「理由」の分析 ～被験者実験より～

中山 記男^{†‡} 神門 典子^{‡†}

† 総合研究大学院大学 情報学専攻 〒101-8430 東京都千代田区一ツ橋 2-1-2 国立情報学研究所内

‡ 国立情報学研究所 〒101-8430 東京都千代田区一ツ橋 2-1-2

E-mail: † norio@grad.nii.ac.jp, ‡ kando@nii.ac.jp

あらまし 本研究はテキストから感情表現とその構成要素として定義した「主体」、「対象」、「理由」を自動抽出し、映画や書籍の作品検索に生かすことを目指している。従来行われてきた主観表現の抽出では、主に主観表現とその主観の対象について分析が行われてきた。それに対して我々は、感情表現に加えて「理由」が特に重要であることを示すため、8人ずつの2グループに対し10件ずつ3つの映画評を用い、映画評を見る際に重要となる部分の調査、分析を行った。その結果、我々が「理由」と定義している感情表現を導き出す理由や根拠が、まったくの他人が書いたテキストを参考とする際の手がかりになっていることがわかった。

キーワード 感情表現, 理由分析, 情報抽出, 情報検索, 主観情報

Analysis of “Reason” in Film Reviews

Norio NAKAYAMA^{†‡} Noriko KANDO^{‡†}

† Department of Informatics, The Graduate University for Advanced Studies, Tokyo, 101-8430 Japan

‡ National Institute of Informatics, Tokyo, 101-8430 Japan

E-mail: † norio@grad.nii.ac.jp, ‡ kando@nii.ac.jp

Abstract In this research, our purpose is to extract emotion expressions and their related components “Subject”, “Object”, “Reason” from text, and to apply them for movie or book retrieval. The researches on extracting subjective expression including emotion has so far focused on the analysis of the subjective statements and their objects. In contrast to them, we have focused on the emotion expressions in book reviews, and analyzed the subjects, objects and reasons for each of the emotion expressions. In this paper, we analyzed the reasons of the emotion expressed in 10 texts each of 3 film reviews and their importance for the review readers. Finally, we proved importance of the reasons of emotion when readers refer to review written by others.

Keyword Emotion Expression, Reason Analysis, Information Extraction, Information Retrieval, Subjective Statements

1. はじめに

インターネット上での商取引やブログの増加により、特定の商品や出来事についての個人の主観的な態度（以下：態度）を表明したテキストが増加している。この態度が特定の商品に対するものである時、商品へのフィードバックとして企業に注目されている。態度が特定のニュースや施策に対するものであれば、国民の反応を知る手がかりとして利用される用途なども考えられる。

我々は、このようなテキストに現れた個人の態度の表明の中でも、特に、「うれしい」「かなしい」などの個人の感情を表す感情表現に着目し、その特性を理解するためのモデルを提案し、作品検索に応用することを目的として研究をすすめている。感情とは、ある対象に対する主体の気分や心の動きであり、感情表現と

はテキストの中で感情を表明した部分である。

現在までに我々は、82件の書評および映画評に対してタグ付けを行うことで、感情表現の構成要素として表1に示す4つの上位要素と各上位要素を分類する下位要素を定義し、それぞれの働きを分析した[1]。その結果、我々が定義した構成要素の中でも特に「理由」が重要であるという提案を行った[2]。これは、作品検索をする際に利用者が重要視するのは、他人が書いたテキストに自分が共感できるかどうかであり、その鍵となる要素こそ「理由」であると考えからである。

本報告では「理由」の重要性を分析するため、ある映画に関する10件映画評を3種類用意し、8名ずつの2つのグループの参加者に参考になった個所についてのタグ付けと、それに関するグループインタビューを行った。

表 1 定義した構成要素の一覧

上位要素	上位要素の定義	下位要素	下位要素の定義	下位要素の分類
態度	態度の表明	態度記述	態度が明示的に示されている	感情, 保留, 感情以外
		ゼロ態度	態度が暗黙的に示されている	
		態度なし	態度が示されていない	
主体	態度を表明した人物, モノ	主体記述	主体が記述されている	書き手, 登場人物, 第三者
		ゼロ主体	主体が省略されている	
		主体不明	主体が不明である	
対象	態度が向けられた人物, モノ	対象記述	対象が記述されている	登場人物, 場面, 様子, ストーリー, 第三者, 発言, 心象, 事実, 行動
		ゼロ対象	対象が省略されている	
		対象不明	対象が不明である	
理由	態度を表明した原因, 理由	明確な理由の記述	明確に理由が記述されている	主観理由, 客観理由
		暗黙的な理由の記述	暗黙的に理由が記述されている	主観理由, 客観理由
		理由不明	理由が不明である	

2. 関連研究

テキストから、なんらかの主観的な態度の表明を抽出することに関連する研究は、態度の主体と着目する要素により、次の3つのグループに分類できる。1つ目のグループは、ある対象が書き手からどのような捉え方をされているかに着目し、主にテキスト中に記述されている書き手の態度とその対象となるものを扱っている。大塚[3]は自由記述型アンケートに記述された書き手の要求意図に着目している。小林ら[4], Daveら[5], Pangら[6], Turneyら[7], Nasukawaら[8]は、Web上に記述された商品の評価表現に着目している。Riloffらは新聞に記述された主観的な態度に着目している。舘野[9]は、企業のサポートセンターによせられる「お客様の声」に含まれる否定的な表現に着目し、それらを含む文を抽出している。このグループの研究では、ある対象に対する「よい」「わるい」といった主観表現そのものが対象とされており、その主観表現を導くに至る理由や根拠に対しては検討されていない。本研究では主観表現に加え理由や根拠も重要であるので、主にそれらに対して抽出を試みた。

2つ目のグループは、テキストを読んだ時、読み手がどのような感情を持つかについて着目し、テキストに書かれていない読み手の態度を扱っている。Liuら[10]は、テキストを読んだ読み手がどのような感情を感じるかに着目している。常識知を用いて本文テキ

スト中に含まれる感情を引き起こす部分を抽出し、あらかじめ決められた感情カテゴリーへ分類することで、「交通事故」から読み手は「悲しみ」を感じるといったように、記述されたある出来事から読み手が抱くであろう感情を推定している。これに対し本研究では、テキストに記述された書き手の主観表現の理由や根拠に注目し、それらを自動抽出することを目的としている。

3つ目のグループでは、テキストに書かれていないものも含めた書き手の態度に着目している。田中ら[11]は、テキスト中に記述されていないものも含めた書き手の情緒を推定するため、結合価パターンを用いている。このグループでは、「盗難にあう」は「悲しみ」といったように、出来事から書き手の情緒を推定している。本研究では記述された感情表現に対応する理由や根拠を重要としており、それらをセットで抽出することを目的としている。

これら関連研究では、最終的に感情や情緒などとその主体を抽出対象としているが、本研究では感情表現とその理由や根拠、主体、対象を抽出対象とし、それらを組み合わせることが作品検索において重要であるとしている。なかでも理由や根拠を特に重要とし、感情表現とともに自動抽出をするため、その特性や働きについて分析を行っている。

3. 実験

感情表現に対する理由や根拠の重要性を分析するため、都内の文系女子大学生 8 名の「グループ X」と、同じく都内の理系大学生 8 名の「グループ Y」を被験者として、ブログに書かれた映画評のテキストのうちの部分を参考にするかについて調査実験を行った。実験は被験者の情報とインターネット利用状況などを調べる事前アンケートと、テキストから参考になる部分に線を引くなどした本調査に分けられる。

3.1. 事前アンケート

被験者のインターネット利用状況や、映画に関する下調べの状況について、事前アンケートを行った。表 2 に質問項目を示す。

表 2 事前アンケートの質問項目

A：被験者について A1：性別 A2：年齢 A3：職業
B：PC やインターネットの利用状況に関して B1：パソコンの利用年数 B2：インターネットの利用年数 B3：週でのインターネット利用時間数（学業／趣味） B4：インターネットを主に何のために利用しますか B5：Google などの検索サイトは利用しますか B6：検索サイトをどれくらい使いますか B7：検索サイトでは何を調べる人が多いですか B8：検索サイトに不満を感じる点はなんですか
C：映画を見ると参考にする情報に関して C1：どんな映画が好きですか（ジャンルや作品名） C2：作品についての下調べをしますか C3：下調べでは何を参考にしますか C4：どんな情報が最も参考になりますか C5：ネット上の情報も参考としますか C6：ネット上ではどんなものを参考にしますか C7：よく参考にする特定のサイトがありますか C8：参考にしたいサイトはどんなものですか
D：インターネット上のテキストに関して D1：ブログや他人の日記を読みますか D2：それらを読む動機はなんですか D3：それらはなんらか行動の参考になっていますか D4：参考になっている点はなんですか

3.2. 本調査

3 つの映画どれか 1 件に関して書かれたブログのテキストを 10 件ずつ、計 30 件用意し、それを順番に提

示することで調査を行った。テキストは goo ブログ検索[12]を用いて各映画のタイトルで検索した結果の上位から、映画評とないものを 10 件選択した。テキストには A から J までの通し番号がふられており、1 つの映画についての 10 件がひとまとめとされている。実験に用いた 3 つの映画のタイトルを表 3 に示す。16 名の被験者に提示した 30 件のテキストは同じものであるが、3 つの映画および 10 件のテキストが提示された順番は異なっている。例えばある被験者には M1 のテキスト 1A, 1B…1C, M2 のテキスト 2A, 2B…2J の順で提示したのに対し、別の被験者には M3 のテキスト 3J, 3I…3A, M1 のテキスト 1J, 1I…1A の順で提示した。表 3 に各被験者に提示した順を示す。調査はグループごと 8 人まとめて行われ、グループ X とグループ Y の調査日時、調査場所は異なっている。各被験者は、3 種類それぞれの映画についての情報を探しており、その情報をもとに映画を見るかどうか決めようとしているという状況を想定して実験を行った。

表 3 調査に用いた映画タイトル

番号	映画のタイトル
M1	ALWAYS 三丁目の夕日
M2	チャーリーとチョコレート工場
M3	ハイフラワーとキルトシュー

表 4 被験者に提示した順番

グループ	被験者番号	提示順 1*	提示順 2*	提示順 3*
X	1	M1 昇順	M2 昇順	M3 昇順
	2	M1 昇順	M3 昇順	M2 昇順
	3	M2 昇順	M1 昇順	M3 昇順
	4	M2 昇順	M3 昇順	M1 昇順
	5	M3 昇順	M1 昇順	M2 昇順
	6	M3 昇順	M2 昇順	M1 昇順
	7	M1 降順	M2 降順	M3 降順
	8	M2 降順	M3 降順	M1 降順
Y	1	M1 降順	M2 降順	M3 降順
	2	M1 降順	M3 降順	M2 降順
	3	M2 降順	M1 降順	M3 降順
	4	M2 降順	M3 降順	M1 降順
	5	M3 降順	M1 降順	M2 降順
	6	M3 降順	M2 降順	M1 降順
	7	M1 昇順	M2 昇順	M3 昇順
	8	M2 昇順	M3 昇順	M1 昇順

* 昇順：A,B,C,…J の順 降順：J,I,H,…A の順

調査では、1 件のテキストの中で被験者が参考に見えると感じた部分があればそこに下線を引き、さらに下線の中で最も参考になったというものがあれば、その部分とその理由を特記してもらった。ある映画に関する 10 件のテキストについて作業が終わった時点で、その 10 件中最も参考となったテキストを選択してもらい、さらにその理由を回答してもらった。この調査を 3 つの映画に関して行った。

調査終了後、下線を引いた部分に関して、それぞれが参考とした部分や基準などについてのグループインタビューを 8 人全員で行ってもらい、意見を収めた。以下では、映画 M1, M2 についての結果について述べる。

3.3. 事前アンケートの結果

事前アンケートの質問でも数値、頻度または○か×で回答された結果を表 5 に示す。なお、各項目上段の質問 ID は表 2 のものに沿っている。

3.4. 本調査の結果

16 人の被験者が映画 M1, M2 のテキスト 10 件それぞれについて引いた下線の数と、10 件中最も参考となったテキストの ID などを表 6 に示す。また、映画 M1 の各テキストの詳細と、それぞれに対して引かれた下線の部分に対しタグ付けを行った結果を表 7 に示す。表 7 の各行は左からテキストの ID, テキストの文字数, 行数, 1 行あたりの平均文字数, 下線が引かれた部分数, 1 下線あたりの文字数, 最も参考になると選択された数, テキスト中の感情表現の数, うち線が引かれた部分の感情表現数, テキスト中の「理由」の数, うち線が引かれた部分の「理由」の数, 感情表現と「理由」のペア両方に下線が引かれていた数, 下線の部分に感情表現も「理由」もなかった数, 1 件のテキスト中で特に参考になった部分に感情表現と「理由」があった数, 同じく特に参考になった部分に感情も「理由」もなかった数である。また、表 8 に各被験者の回答した最も参考となったテキストと、その理由の一覧を示す。

表 5 事前アンケートの回答結果 1

質問 ID	被験者 番号	A1	A2 (才)	A3	B1 (年)	B2 (年)	B3		B5	B6 (頻度)	C2	C5	D1	D3
							学業/趣味 (時間)							
グループ X	X1	女	20	大学生	10	10	6/6	○	毎回	○	○	○	○	
	X2	女	21	大学生	7	7	4/30	○	毎回	×	×	○	○	
	X3	女	20	大学生	7	4	3/7	○	毎回	○	○	○	×	
	X4	女	21	大学生	4	4	7/5	○	毎回	○	○	○	×	
	X5	女	20	大学生	4	4	7/5	○	毎回	×	×	○	×	
	X6	女	21	大学生	3	3	1/5	○	毎回	○	○	×	×	
	X7	女	21	大学生	7	7	0/14	○	毎回	○	○	○	○	
	X8	女	20	大学生	11	8	1/1	○	毎回	○	○	○	○	
平均			20.5		6.63	5.88	4.0/9.1							
グループ Y	Y1	男	24	院生	10	6	15/30	○	毎回	○	○	○	○	
	Y2	男	23	院生	4.5	4.5	7/7	○	毎回	○	×	○	×	
	Y3	男	23	大学生	10	9	7/4	○	毎回	○	○	○	○	
	Y4	男	22	大学生	9	7	7/35	○	毎回	×	×	○	○	
	Y5	男	22	大学生	4	4	16/12	○	毎回	○	×	○	×	
	Y6	男	22	大学生	5	4	5/40	○	毎回	×	×	○	○	
	Y7	男	22	大学生	7	7	3/7	○	時々	×	×	×	×	
	Y8	男	21	大学生	5.5	5.5	5/50	○	毎回	×	×	○	○	
平均			22.4		6.88	5.88	8.1/23.1							
全平均			21.4		6.75	5.88	6.1/16.1							

表 6 映画 M1, M2 における各被験者の下線の数と最も参考になったテキスト

ID	1A	1B	1C	1D	1E	1F	1G	1H	1I	1J	Av1	2A	2B	2C	2D	2E	2F	2G	2H	2I	2J	Av2	Av12
X1	5	2	<u>2</u>	2	1	3	1	2	2	4	2.4	3	2	7	2	3	3	4	1	1	4	2.4	2.7
X2	0	0	2	2	0	0	1	0	0	<u>2</u>	0.7	4	2	0	1	0	2	5	0	0	3	1.7	1.2
X3	7	3	9	7	4	11	2	1	0	3	4.7	3	3	5	4	6	6	12	2	3	8	5.2	4.95
X4	5	6	4	4	3	3	2	2	2	5	3.7	0	2	3	2	0	2	6	3	3	2	2.3	3.0
X5	2	3	0	1	0	2	0	0	1	1	1.0	3	2	0	0	0	2	0	0	2	2	0.9	0.95
X6	0	2	1	1	0	1	2	0	1	<u>2</u>	1.0	0	1	1	1	0	0	2	0	0	1	0.6	0.8
X7	8	17	5	5	1	9	5	1	2	4	5.7	12	9	11	6	15	15	23	6	6	15	11.8	8.75
X8	3	8	<u>2</u>	<u>1</u>	1	4	2	1	0	5	2.7	<u>7</u>	3	6	1	6	0	3	2	2	5	3.5	3.1
AvX	4	5	3	3	1	4	2	1	1	3	2.7	4	3	4	2	4	4	7	2	2	5	3.6	3.1
Y1	6	4	0	4	1	3	1	3	2	6	3	2	5	6	1	5	4	8	3	1	11	4.6	3.8
Y2	<u>1</u>	2	0	1	2	2	0	1	2	4	1.5	3	0	3	0	0	0	2	0	2	2	1.2	1.35
Y3	3	4	2	0	2	5	1	0	1	6	2.4	1	0	2	1	1	0	11	1	2	2	2.1	2.25
Y4	4	2	2	2	3	9	0	2	2	2	2.8	0	2	<u>1</u>	0	1	1	1	1	0	6	1.3	2.05
Y5	0	0	0	3	0	4	4	0	0	5	1.6	5	0	3	0	1	0	4	0	0	1	1.4	1.5
Y6	0	1	2	2	0	0	5	1	2	5	1.8	0	0	3	0	0	0	5	1	2	4	1.5	1.65
Y7	0	2	1	2	1	2	1	1	1	<u>2</u>	1.3	3	1	1	2	0	3	1	0	1	4	1.6	1.45
Y8	4	4	1	1	2	0	1	1	0	2	1.6	1	2	4	0	2	0	3	1	1	4	1.8	1.7
AvY	2	2	1	2	1	3	2	1	1	4	2	2	1	3	1	1	1	4	1	1	4	1.9	1.97
AvA	3	4	2	2	1	4	2	1	1	4	2.4	3	2	4	1	3	2	6	1	2	5	2.8	2.58

下線太字

…最も参考になったテキスト

表 7 映画 M1 における各テキストの詳細と下線が引かれた部分の構成要素

ID	文字数	行数	文字数/行	下線	文字数/線	参考	感情	感情	感情に		理由	理由に		両方に	両方に	下線	下線	参考	参考
									下線	下線		下線	下線						
1A	438	19	23.1	16	27.4	1	16	14	6	6	6	6	6	6	5	3	3	5	2
1B	996	22	45.3	16	62.2	2	11	10	8	8	7	7	6	6	8	8	8	8	0
1C	702	17	41.3	12	58.5	1	13	13	5	5	3	3	3	3	3	1	1	3	0
1D	445	14	31.8	10	44.5	1	8	8	4	4	4	3	3	3	2	4	4	2	7
1E	634	21	30.2	9	70.4	0	12	9	2	2	1	1	1	1	0	2	2	0	1
1F	887	35	25.3	20	44.35	2	14	10	2	2	2	2	1	1	1	11	11	1	2
1G	440	11	40.0	5	88.0	1	10	6	3	3	1	1	1	1	6	1	1	6	1
1H	195	6	32.5	5	39.0	1	1	1	0	0	0	0	0	0	0	4	4	0	7
1I	564	21	26.9	7	80.6	0	11	7	6	6	4	4	4	4	5	0	0	5	0
1J	907	29	31.3	22	41.2	7	12	10	6	6	6	6	3	3	4	10	10	4	8
平均	620.8	19.5	32.76	12.2	50.9	11	8.8	4.2	4.2	4.2	2.8	2.8	4.4	4.4	3.4	2.8	2.8	3.4	2.8

表 8 映画 M1, M2 において最も参考になったテキストとその理由

ID	最も参考になったテキスト	参考になった理由の要点
Y2	1A	リターナーと同じ人が作っているとわかってよかった。
Y8	1B	原作を知った上で見た人の意見で参考になる。原作との比較が良かった。
X5		リアリティーに欠ける点が指摘されていたから。
X1	1C	思ったことがそのまま書いてありそう。CG や俳優それぞれ良かった点や豆知識が良い。
X8	1D	説明された情景が気になったので。具体的なシーンの説明がよい。
X7	1F	ストーリーの情景がよく表されている。オススメポイントや見終わった後の気分がある。
Y4		具体的な内容がかかれており、イメージしやすかった。
Y5	1G	映画の具体的な内容が書いてあってわかりやすかった。感想も簡潔でわかりやすかった。
X4	1H	短いわかりやすかった。感想が多なくて見やすかった。映画の設定が書いてあった。
X2	1J	内容がわかりやすかった。原作を絡めていてよかった。他のものに比べ感情的でなかった。
X3		対象と思われる年代が書いてあったので。
X6		感動したことが伝わってきたので。オヤジ世代が感動する点が気になったので。
Y1	1J	映画の概略が感想の前に書かれていて参考になった。簡潔でよかった。
Y3		時代背景がわかったことで、映画の雰囲気があった。映画のイメージがふくらんだ。
Y6		原作、舞台とその背景、ストーリー、俳優の演技の出来、映画の特徴が説明されていた。
Y7		映画の概要、主役とその演技、感想やコメント、特に客観的な意見が信頼に足りた。
X3	2A	シュールな笑いが好きで、そういった映画だとわかったから。批判的より好意的なものがよい。
X5		映画にシュールさやブラックジョークがある様子が面白く紹介されていたので。
X8		他の「ファンタジー」「笑える」と書いてあるレビューより冷静に紹介されている感を受けたので。
Y5		映画のおおまかな内容を紹介した上で自分の感想を述べているので。
Y2	2C	チョコレート絡めてレビューしているところが面白くてよかった。
Y4		映画の内容について書かれていないから。
X1	2E	見所が紹介されていてよかった。良いところ、悪いところ両方書いてあって信頼できる。
Y4	2G	予備知識として知っておきたいことが押さえられている。感想が押し付けがましくない。
Y6		自分のイメージと違う意見がよかった。短い文だが簡潔にまとめられたムダのない意見だった。
Y7		ストーリーの紹介と映画の感想が書かれているなど、トータルでよかった。主演や監督もわかる。
X1		映画にまつわる周辺の情報について広く書かれていてよい。感想については物足りなさも。
X3		ストーリーが書いてあったので。自分も好きなジョニーデップが好きと書いてあり、共感した。
X6		映画の舞台をわかりやすく説明している。他のものより理解しやすいレビューだった。
X2	2J	映画内容の簡単な紹介があった。他の作品と比較されている点がよかった。
Y7		映画と原作両方を比較している点が良い。映画の中身にもしっかりコメントしている。
Y8		原作や他の映画との比較がある点。テーマやメッセージ性の説明や役者、CG についても触れている。

下線が引かれた部分の集計は、実際に被験者によって線が引かれている部分が重なっている場合、同じ部分に下線が引かれているとして扱っている。以下に実例を示す。

やっぱり期待外れでありましたよ。 …①

とてもほのぼのとした良い映画。 …②

①は2つの下線を、②では3つの下線を同じ部分に引いた下線として扱った。これは、下線が引かれていなくとも、構成要素として関連性がある部分を分離して考え難かったためである。①は文末表記を加えるかどうかについて被験者ごとに揺れがあったものを同じものとして扱った例である。②は、「とてもほのぼの」にしか下線が引かれていない例が、下線を引いた被験者にとって「『とてもほのぼの』としていたのは『映画』である」という理解があった上で下線が引かれたはず

であると考えられたため、3種類の下線を同じものとして扱った例である。

3.5. グループインタビューの結果

グループディスカッションでは、下線を引いた部分を選んだ基準や、情報を参考とする基準について自由に議論してもらった。

その結果、被験者の選ぶ基準として、主に

- ・ 映画の上映場所、日時、長さがわかる …③
- ・ 登場人物がわかる …④
- ・ あらすじ、見所がわかる …⑤
- ・ 映画がどんな人に向けられているか …⑥
- ・ 映画をみてどう思ったか …⑦
- ・ 映画評の書き手の人物像がうかがえる部分…⑧
- ・ 決め付けではなく、理解しやすく共感できるような筋道立った記述がされている部分 …⑨
- ・ ~のような映画など、比較対象がわかる …⑩

の点が挙げられた。これらの点は、実際そのような部分に下線を引かなかった被験者からも支持された。

4. 分析

3節の結果をもとに、利用者の参考とするものと「理由」の関係を中心に分析を行った。

4.1. 被験者について

表5より、今回の被験者はグループXとグループYで趣味としてインターネットを使う時間が大きく違う。また、グループXがネット上の情報を参考として映画を見るのに対し、グループYはあまり下調べもネット上の情報を参考にもしない。このように両者の映画の下調べに関する性質はかなり異なっているが、表6からわかるように、どちらのグループに関して最も参考になるテキストとして選んだものは、テキスト1J、2Gで共通している。このことより、利用者のインターネット習熟度や普段の行動と、参考になるテキストを選ぶ基準というものはあまり関連性がないと考えられた。下線を引いた数の平均はグループXのほうが多い。これは表6からわかるように、グループYには全体平均の2.8本を大きく上回る本数の下線を引く被験者が2名いたためであると思われる。

4.2. テキストの長さや参考にするものの関係

表7より、参考になると回答された数が7件と最も多かったテキスト1Jについて、テキストの文字数などから考察する。文字数、1行あたりの文字数については907文字と2番目に多く、下線が引かれた部分につ

いては22箇所と一番多い。また、表6を参照すると、テキスト1Jについて各被験者が引いた平均の下線数は3.6本で10件中2番目に多いことがわかる。対して996文字とテキスト1J以上の文字数を持つテキスト1Bは、下線が引かれた部分は16箇所、最も参考になると回答された数は1件だけと、テキスト1Jに対して非常に少ない。反面、下線が引かれた平均数は3.8本とテキスト1Bのほうが勝っているが、これは被験者X7が17本もの下線を引いているためであり、被験者X7を除いて再集計すると、平均3.4本となりテキスト1Jよりも少なくなる。

さらに下線の平均数と下線が引かれた部分の数の関係に注目すると、テキスト1Jは3.6本で22箇所、テキスト1Bは3.8本で16箇所（被験者X7を除くと、3.4本で10箇所）と、大きく違っていることがわかる。これは、1Jのテキストがほぼ書き手の感情表現や感想で占められているのに対し、1Bのテキストは原作に書いてあった事実を報告している部分が多いため、あまり参考にされなかったためだと考えられる。このことより、テキストが長ければそれだけ参考になるわけではなく、内容に書き手の感情などが含まれ、共感できるようなテキストのほうが参考とされると考えられた。このことは、表8の1Jを選んだ理由に多く見られる「主役の演技への言及」「映画の雰囲気の説明」「自分が感動した点の説明」からも同様にして考えられた。なお、グループ被験者Y7の選択理由にある「客観的な意見」とは、闊雲な賞賛や批判ではない、筋道立てて説明されている意見のことである。

4.3. 参考になるテキストと感情表現、「理由」の関係

表6から、最も多く下線が引かれた、つまり最も参考になる箇所が多かったテキストが、必ずしも最も参考となるテキストとして選ばれているわけではないとわかる。例えば5箇所に下線を引いたテキスト1Aではなく、2箇所にしか下線を引かなかったテキスト1Cを選択した被験者X1の選択理由を表8で確認すると、映画の製作者が判明して参考になったと回答されている。同様に6箇所引いたテキスト1Bではなく2箇所しか引いていないテキスト1Hを選択した被験者X4の例を見ると、わかりやすかった点が理由として挙げられている。このことより、参考になる部分が多く見られることも重要であるが、被験者によっては自分が重要とするポイントまたはわかりやすさなど、内容の質について注目していることがわかる。さらに表6で回答されている理由を見ると、映画の内容に関わる内容や他のものとの比較、感想の内容やそのわかりやすさが多く指摘されていることもわかる。

表7より、下線の引かれた部分のほとんどが、感情

参考文献

表現ないしその「理由」を含んでいたことがわかる。中でもテキスト中に見られた「理由」のうち、ほとんどの部分に下線が引かれていた。1件のテキスト内でも特に参考になった部分にも、感情表現と「理由」が含まれている部分が多かったこともわかる。対して、感情表現や「理由」も含まない部分に多く下線の引かれたテキスト 1B や 1F , 1J の内容を見ると、下線が引かれている部分は映画を見に来た人の様子による事実報告や、映画のストーリー、ポイントとする映画の場面を実体験から解説している部分が多かった。

以上の結果より、被験者にとって重要なのは映画の内容を説明する部分とその感想を記述する部分に大別でき、感想を記述する部分において、感情表現と「理由」は被験者にとって参考となる情報として認識されていたと考えられた。

4.4. グループインタビューの分析

グループインタビューの内容から、被験者にとって最も参考としやすい情報は「書き手がどんな趣向でどのような性質を持つ人かわかっていること」であることがわかった。そのため、友人や知人からのクチコミのほうが、ネットに書き込まれた情報よりも信頼に足るという意見が多く出された。そんな中でも、なぜ書き手がそういった考えを表記したか理解するための足がかりになる具体的な記述、つまり「理由」があるかどうかは、書き手の性質がわからない際、ひとつの大きな判断材料となることがわかった。つまり、ブログなどにある映画評が「完全な主観的意見」である場合は信頼に足らず、対して何か少しでも明確な比較対象や具体的な説明、さらには筋道立てた内容があることで、はじめて信頼や共感が生まれると考えられた。

5. まとめ

本報告では、映画のレビューにおいて、その利用者が参考とする部分が感情表現やその「理由」であると仮定し、それを実証するために被験者実験を行い、分析を行ったものである。その結果、利用者にとって「理由」はテキストの内容を理解、信用するためのひとつの大きな要素であることがわかった。今後は感情表現とその構成要素である「主体」「対象」「理由」を自動抽出し、処理できるようなシステムの開発を目指す。

- [1] 中山記男, 江口浩二, 神門典子, 感情表現のモデル, 言語処理学会第 11 回年次大会, 発表論文集, pp:149-152, 2005 03 15.
- [2] 中山記男, 神門典子, 理由に着目した感情表現の分析, 電子情報通信学会技術研究報告 思考と言語, Vol.105, Num.291, pp.51-56, 2005. 09 16
- [3] 大塚裕子, 自由記述アンケートと回答の意図抽出および自動分類に関する研究, 神戸大学大学院自然科学研究科博士論文, 2004.3.
- [4] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一, 意見抽出のための評価表現の収集, 自然言語処理, Vol.12, No.2, pp.203--222, 2005.07
- [5] Dave, K., Lawrence, S., Pennock, D.M. Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. International World Wide Web Conference, Budapest, Hungary pp.519-528, 2003.
- [6] Pang, B., Lee, L., Vaithyanathan, S., Thumbs up? Sentiment Classification using Machine Learning Techniques, pp.79--86, Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP2002), Philadelphia, July 2002.
- [7] Turney, P., Littman, M., Measuring praise and criticism: Inference of semantic orientation from Association. ACM Transactions on Information Systems (TOIS) Vol.21, Num.4, pp.315-346, 2003.
- [8] Nasukawa, T., Yi, J., Sentiment Analysis: Capturing Favorability Using Natural Language Processing. International Conference On Knowledge Capture: Proceedings of the international conference on Knowledge capture K-CAP '03 pp.70-77, 2003.
- [9] 館野昌一, 「お客様の声」に含まれるテキスト感性表現の抽出方法, 情報処理学会研究報告 自然言語処理 Vol.2003, Num.4, pp.105-112, 2003.
- [10] Liu, H., Lieberman, H., Selker, T., A Model of Textual Affect Sensing using Real-World Knowledge. To Appear in Proceedings of IUI 2003, January 12-15 2003.
- [11] 田中努, 徳久雅人, 村上仁一, 池原悟, 結合価値パターンへの情緒生起情報の付与, 言語処理学会第 10 回年次大会発表論文, pp.345-348, 2004.
- [12] goo ブログ検索, <http://blog.goo.ne.jp/>