

## 因子分析を用いた口コミ情報における有用性指標抽出 Extraction of usefulness indicators of user reviews by factor analysis

安井 顕誠<sup>†</sup>      高橋 渉<sup>†</sup>      原田 史子<sup>‡</sup>      島川 博光<sup>‡</sup>  
Kenjo Yasui      Wataru Takahashi      Fumiko Harada      Hiromitsu Shimakawa

### 1. はじめに

近年 www の普及と同時に、オンラインショッピングの利用者が増加している。総務省の調査から、15歳以上の国民におけるオンラインショッピングの利用率は45%とみなせる [1]。オンラインショッピングは www 上の店舗のことで、実際に店舗に足を運ばなくてもショッピングができるという利点がある。しかし、オンラインショッピングでは商品を実際に見る、触るといったことができないため商品の使用感がわからない。そのため、消費者は自身にとって適切な商品を選択できずに商品の購入に対して後悔してしまうという問題点がある。

本研究では、この問題点を解消するために口コミ情報に着目する。口コミ情報は、商品を実際に使用してその商品に対する知識を得た人が書き込んだ情報のため、購入商品を選択する上で有用な情報になると考えられる。しかし、口コミ情報は商品に対する感想や評判などの主観的な内容中心の構成であるため、商品選択において有用でない情報も含まれる。ゆえに、口コミ情報から商品選択において有用な情報を抽出する指標が必要である。

本論文では、オンラインショッピングサイトから得られた口コミ情報を解析し、オンラインショッピングにおいて商品選択で有用な口コミ情報に共通する指標群を発見する。本指標群を用いると、口コミ情報から必要な部分のみを抽出、提示することができ、消費者に効率の良いオンラインショッピング環境を提供することができる。

### 2. オンラインショッピングにおける商品情報

#### 2.1 スペック表の欠点と口コミ情報の有用性

オンラインショッピングには、商品を実際に見る、触るといったことができないため商品の使い心地がわからないという欠点が存在する。この欠点を補うためにオンラインショッピングではスペック表が提供されている。このスペック表は商品の「重量」や「サイズ」などの特徴や価格が記載されている。消費者はその表の数値や名称を基に、購入を検討している商品が自身に適切かどうか判断する。しかし、スペック表では消費者は商品の使用感を誤って評価してしまう場合がある。例えば、消費者がノート PC の知識について乏しい場合、自身の要求を満たすノート PC の処理速度がどの CPU を採用すると満たせるか判断できない。同様に、消費者がノート PC の誤った知識を持っている場合、処理速度が CPU やメモリなどのスペック表のどの項目に関連するか誤って判断してしまう。すなわち、自身の要求である処理速度に対し、スペック表の項目である CPU やメモリが適切であるか正しく考慮できない。このように、スペック表に記載されている数値や名称では商品についての知識が乏しい消費者にとって適切な判断基準にならず、上手く商品選択ができない。また、商品に対する知識を持って

たとしてもその知識が誤ったものであればスペック表に記載されている数値や名称を誤って判断し、自身にとって適切でない商品選択をしてしまう。これらの誤った商品選択は、消費者にオンラインショッピングでの商品購入に対して後悔を発生させる。この後悔の発生は、消費者が購入した商品に対して満足できなかったことを表すもので、オンラインショッピングにおける問題点である。

そこで、本研究では口コミ情報に着目する。オンラインショッピングサイトにおける口コミ情報は、商品を実際に使用してその商品に対する知識を得た人が書き込んだ情報であるため、購入前の消費者の知識より正確な情報である。すなわち、消費者がノート PC の知識について乏しい、もしくはノート PC の誤った知識を持っている場合でも、「ネットサーフィン程度ならばサクサク動作し、快適でした」などという口コミ情報を提示すると消費者は商品の使用感を想像できる。したがって、口コミ情報を提示し、正しい使用感を想像させられるとその使用感が正しい商品選択の基準となり、オンラインショッピングにおいても消費者は自身にとって適切な商品を選択できると考えられる。

#### 2.2 口コミ情報の有用性を示す指標の必要性

口コミ情報のすべてが商品選択において有用ではない。口コミ情報には、「注文からの配送が早かった」などの商品とは関係のない内容も含まれる。また、「私には適さなかった」などの主観性の高い内容は商品の使用感が伝わりづらく、商品選択において有用でないと考えられる。つまり、口コミ情報すべてを提示しても消費者の商品選択において有用な情報にはなり得ない。

本研究では、口コミ情報を解析し、オンラインショッピングにおいて商品選択に影響を与える口コミ情報に共通する指標群を発見する。本研究ではこれを口コミ情報の有用性指標群と定義する。これらの指標群により、口コミ情報から商品選択において有用な情報のみを抽出、提示することで、商品の使用感を想像させ、オンラインショッピングにおける問題点を解消することができる。

### 3. 口コミ情報の有用性指標群抽出手法

#### 3.1 有用性指標群抽出手法の概要

本研究では、オンラインショッピングにおいて商品を選択する上で有用な情報に共通する指標群を抽出する。本研究にて使用する有用性指標群の抽出手法の全体図を図1に示す。有用性指標群の抽出手法は、手続きⅠから手続きⅢまでの3つの手続きを順に実行するものであり、以下の節にて説明する。

#### 3.2 手続きⅠ：評価文抽出

手続きⅠでは、口コミ情報から評価文を抽出する。本研究では、「重量が軽くて持ち運びに適している」のように商品のスペック表の各項目と形容詞、形容動詞が共起する文を評価文として口コミ情報から抽出する。これにより、商品に対する感想や評価以外も記述されている

<sup>†</sup>立命館大学大学院情報理工学研究所

<sup>‡</sup>立命館大学情報理工学部

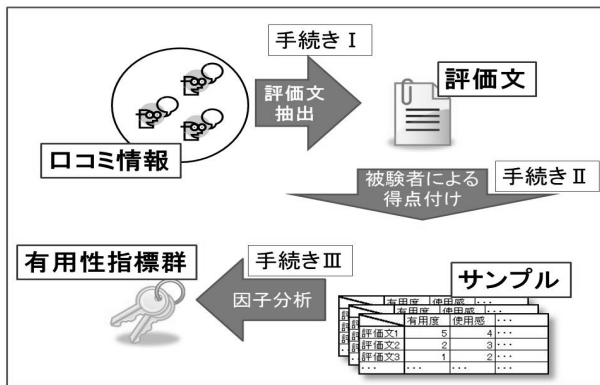


図 1: 指標抽出手法の全体図

表 1: 得点付け項目とその説明

得点付け項目	説明
有用度	文章の内容が商品の購入に影響を与えるものか
文の構成	文章の意味が理解できるように論理的に整っているか
形容詞	形容詞または形容動詞が文章の最後の文節に来ているか
主観客観	文章の内容が主観的な感覚が客観的な評価か
数値	文章に商品の性質や性能を表す具体的な数値が用いられているか
否定肯定	文章が商品に対してメリットを表している内容かデメリットを表している内容か
初知り	文章の内容を初めて知ったかどうか
比較表現	文章の内容が他の商品や自身の経験と比較している内容か
具体抽象	文章の内容が実際の経験に基づいた内容か

ロコミ情報から、感想や評価のみが抽出できると考えられる。

### 3.3 手続き II: 得点付けによるサンプル作成

手続き II では、手続き I にて抽出した評価文を被験者に得点付けしてもらいサンプルを作成する。本研究では、評価文とその評価文に対しての実験における被験者の得点付けの組をサンプルと定義する。被験者に得点付けしてもらった項目を表 1 に示す。

表 1 の得点付け項目である 9 項目は「有用度」を除き、有用性指標群に影響を及ぼすと推測されたものであり、事前実験にて設定した。事前実験では、20 代男性 3 名の被験者にノート PC をオンラインショッピングサイトで購入する場面を想定してもらった。この想定の上、楽天市場 [4] が提供する楽天データセット [5] の中で、ノート PC に対するすべてのロコミ情報から抽出された評価文 100 件から 200 件を提示し、それぞれ以下の項目に回答してもらった。

**項目 1:** 評価文を提示されるとノート PC の使用感が想像できるか (5 段階)

**項目 2:** なぜ使用感が想像できるか、想像できないか (自由記述)

この事前実験結果である項目 1 の評価と項目 2 の自由記述において、項目 1 の評価が高い評価文に頻繁に記述された項目 2 の自由記述の内容を得点付け項目として設定した。

### 3.4 手続き III: 因子分析による有用性指標群抽出

手続き III では、手続き II にて作成したサンプルを用いて有用性指標群を抽出する。抽出手法には因子分析 [2] を用いる。手続き III では、因子分析における観測変数として各サンプルの「有用度」を除いた得点付け項目の得点を用い、主因子法 [3] にて因子の抽出を試みる。ここで抽出できた因子を有用性指標群とする。手続き III では、因子分析において固有値 1 以上の因子を 1 つの有用性指標の基として抽出する。よって抽出される因子は、

表 2: 得点付け項目の点数とその意味

得点付け項目	点数
有用度	とても影響した: 5, 影響した: 4, わからない: 3 影響しなかった: 2, とても影響しなかった: 1
影響	購入する: 1, 購入しない: 0
文の構成	とてもしっかりしている: 2, しっかりしている: 1, しっかりしていない: 0
形容詞	文の最後: 1, 形容詞が存在しない: 0, 文の最後以外: -1
主観客観	主観的: 1, どちらでもない: 0, 客観的: -1
数値	記述されている: 1, 記述されていない: 0
否定肯定	肯定: 1, どちらでもない: 0, 否定: -1
初知り	初めて知ったこと: 1, 初めて知ったことではない: 0
比較表現	含まれている: 1, 含まれていない: 0
具体抽象	具体的: 1, どちらでもない: 0, 抽象的: -1

観測変数に影響を一定以上与えているものとなる。また主因子法により、抽出された第 1 因子が他の因子より仮定された観測変数に影響を与えるようになる。さらに斜交回転により、抽出された因子がどの観測変数に影響を与えているかが回転前より単純に解釈できる。したがって、本節にて説明した設定での因子分析を実行することで、有用性指標群が、他の因子分析手法より少数かつ多くの情報量を持つ形で抽出できる。

## 4. 被験者実験による有用性指標群抽出

### 4.1 被験者実験によるサンプル作成

手続き II においてサンプルを作成するために、20 代の男性 18 名、女性 2 名の被験者に対して実験を実施した。詳細を以下に記す。被験者に「東芝 dynabook satellite B451」(以下、「dynabook」という)というノート PC をオンラインショッピングサイトで購入する場面を想定してもらった。そのノート PC に対する情報はサイト内のスペック表からのみとし、他のサイトや実世界からは入手できないものとした。以上の想定で、「dynabook」に対するロコミ情報から抽出できた 137 文の評価文を 1 件ずつ被験者に提示し、被験者に表 1 の得点付け項目すべてで点数付けしてもらった。本研究では、楽天市場 [4] が提供する楽天データセット [5] の中で、「dynabook」に対して記述されているロコミ情報全てを使用した。表 1 と対応する項目の具体的な点数を表 2 に示す。

### 4.2 因子分析を用いた有用性指標群抽出

手続き III では、手続き II にて作成したサンプルを使用して因子分析を実行し、有用性指標群を抽出する。この有用性指標群を使用して、ロコミ情報から商品を選択する上で有用な情報を抽出するが、ロコミ情報には商品とは関係のない情報も多く含まれる。したがって、有用性指標群は商品を選択する上で有用な情報を抽出する精度が高いものが好ましい。そこで本研究では、手続き III における因子分析において使用するサンプルを 3 種類の方法で分割し、それぞれの分割方法において因子分析を実行する。そして、それぞれの分析で抽出できた有用性指標群の精度についても検証する。3 種類の分割方法にて実行する分析を表 3 に示す。分析 A では、すべてのサンプルに対して手続き III を実行した。分析 B では、すべてのサンプルの内、有用度を 4, 5 と得点付けされたものに対して手続き III を実行した。分析 B は、有用だと判断された評価文のみから有用性指標群を抽出する。分析 C では、被験者に対して「重視するノート PC の特徴」についてアンケートを実施し、アンケートの回答ごとに被験者をクラスタリングし、クラスタごとに属する被験者が作成したサンプルについて手続き III を実行し

表 3: 分割方法別の分析手法

被験者		サンプル	
		すべて	有用度別 点数 4, 5
すべて		分析 A	分析 B
重視 項目別	処理速度	分析 C	
	持ち運びやすさ		

表 4: 分析 A : 因子を意味づける評価文の性質

因子 1	因子 2	因子 3
初めて知ったこと 具体的	構成がしっかりしている 具体的	形容詞が文の途中にある 客観的 否定的 比較表現である

た。つまり、分析 C では、被験者の志向ごとにサンプルを分割し、それぞれに対して手続き III を実行した。分析 C では、いくつか被験者に対してアンケートを取ったが、「重視するノート PC の特徴」別のクラスタリングが最もサンプルの得点付けにばらつきが出たため、このクラスタリング法に対してのみ手続き III を実行した。以下、各分析について詳しく記述する。

#### 4.2.1 分析 A : サンプル全体での因子分析

分析 A では、全てのサンプルを用いて手続き III の因子分析を実行した。このとき抽出された因子と 8 つの観測変数との関連度合いである因子負荷量を確認する。この因子負荷量の絶対値が、すべての因子において 0.2 未満であった観測変数である「数値」観測変数を除いて再度因子分析を実行した。その結果、因子が 3 つ抽出できた。ここで、各因子において因子負荷量の絶対値が 0.2 以上となった観測変数群を抽出された 3 因子の本質を表す観測変数群として採用した。それぞれの因子で採用された観測変数群は、因子 1 は「初知り」、「具体抽象」観測変数群、因子 2 は「文の構成」、「具体抽象」観測変数群、因子 3 は「形容詞」、「主観客観」、「否定肯定」、「比較表現」観測変数群となった。ここで、サンプルがそれぞれの因子に対してどのくらいの重みを持っているかの得点である因子得点を確認する。サンプルを有用度の 5 段階で分類し、それぞれ因子得点の平均を算出する。算出した因子得点の平均から、因子 1、因子 2 は有用度が上昇するごとに因子得点が増加し、因子 3 は有用度が上昇するごとに因子得点が増加することが確認できた。つまり、因子 1、因子 2 は有用度に正の相関関係、因子 3 は有用度に負の相関関係がある。よって、因子ごとの観測変数を因子 1、因子 2 はそのまま、因子 3 は正負を逆転させて因子を意味づける評価文の性質として採用した。これを表 4 にまとめる。

#### 4.2.2 分析 B : 高有用度のサンプルに対する因子分析

分析 B では、被験者が有用度を 4, 5 と得点付けたサンプルのみを用いて手続き III の因子分析を実行した。分析 B でも分析 A と同様に、まず 8 つすべての観測変数を用いて因子分析を実行した。その後、因子負荷量の絶対値がすべての因子において 0.2 未満であった観測変数である「主観客観」観測変数を除いて再度因子分析を実行した。その結果、因子が 3 つ抽出できた。ここで、各

表 5: 分析 B : 因子を意味づける評価文の性質

因子 1	因子 2	因子 3
形容詞が文の途中にある 否定的 比較表現である	構成がしっかりしている 形容詞が文の最後にある 初めて知ったこと 具体的	構成がしっかりしている 具体的な数値が無い 初めて知ったこと

表 6: 分析 C 「処理速度」:

因子を意味づける評価文の性質

因子 1	因子 2	因子 3	因子 4
形容詞が文の途中にある 否定的 比較表現である	構成がしっかりしている 形容詞が文の最後にある 初めて知ったこと 具体的	構成がしっかりしている 具体的な数値が無い 初めて知ったこと	客観的 抽象的

因子において因子負荷量の絶対値が 0.2 以上となった観測変数群をそれぞれの因子の本質を表す観測変数群として採用した。ただし、因子 3 においては因子負荷量の絶対値が 0.2 以上となった観測変数が 1 つのみとなったため、1 度目の因子分析において抽出できた因子 3 を採用した。よってそれぞれの因子で採用された観測変数群は、因子 1 は「形容詞」、「否定肯定」、「比較表現」観測変数群、因子 2 は「文の構成」、「形容詞」、「初知り」、「具体抽象」観測変数群、因子 3 は「文の構成」、「数値」、「初知り」観測変数群となる。分析 B において抽出できた因子は、分析 A にて抽出できた因子と類似する部分が多く見受けられた。したがって、分析 A においてそのまま採用した因子と類似する因子はそのまま、正負を逆転させて採用した因子と類似する因子は正負を逆転させて分析 B においても採用した。これを表 5 にまとめる。

#### 4.2.3 分析 C : 被験者の志向ごとの因子分析

分析 C では、被験者に対して「重視するノート PC の特徴」についてアンケートを取り、アンケートの回答ごとに被験者をクラスタリングし、クラスタごとに属する被験者が作成したサンプルについて手続き III の因子分析を実行した。アンケートでは、被験者をそれぞれ「処理速度」、「持ち運びやすさ」、「値段」について重視するクラスタに分割できた。本論文では、「処理速度」、「持ち運びやすさ」、「値段」をそれぞれについて重視するクラスタと定義する。クラスタごとに分析 A と同様の因子分析を実行し、サンプルを有用度の 5 段階で分類し、それぞれ因子得点の平均を算出した。因子得点の平均と有用度の関係をクラスタごとに確認すると、「処理速度」、「持ち運びやすさ」には何らかの相関関係が見られたが、「値段」には相関関係が見られなかった。そこで分析 C では、「処理速度」、「持ち運びやすさ」について着目する。それぞれクラスタごとのサンプルについて分析 A と同様の因子分析を実行し、因子負荷量を確認し、この絶対値がすべての因子において 0.2 未満であった観測変数を除いて再度因子分析を実行した。その結果、「処理速度」では因子が 3 つ、「持ち運びやすさ」では因子が 3 つ抽出できた。ここで、各因子において因子負荷量の絶対値が 0.2 以上となった観測変数群をそれぞれの因子の本質を表す観測変数群として採用した。採用した観測変数群を分析 A と同様に因子得点の有用度ごとの平均と有用度との相関関係から解釈し、因子を意味づける事柄として採用した。これを表 6、表 7 にまとめる。

表 7: 分析 C 「持ち運びやすさ」:

因子を意味づける評価文の性質

因子 1	因子 2	因子 3
形容詞が文の途中にある 否定的 初めて知ったこと 比較表現である	構成がしっかりしている 形容詞が文の最後にある 初めて知ったこと 具体的	構成がしっかりしている 否定的 具体的な数値が無い 初めて知ったこと

表 8: 分析ごとの抽出された有用性指標群の有効性評価

	分析 A	分析 B	分析 C 「処理速度」	分析 C 「持ち運びやすさ」
適合率 (%)	0.359	0.486	0.511	0.512

## 5. 有用性指標群の有効性評価と議論

### 5.1 有用性指標群の有効性評価

分析 A から分析 C にて抽出された因子群は、口コミ情報の有用性指標群である。本章では、分析 A から分析 C にて抽出された因子群の、有用性指標群としての有効性を評価する。有効性は、各有用性指標群を用いることで、商品を選択する上で有用な評価文を精度良く抽出できるか否かで評価する。本節では、有用性指標を持つサンプルを、分析 A から分析 C それぞれにて抽出できた因子を意味づける性質をすべて持つものと定義する。被験者が有用度に 4 または 5 と点数付けしたサンプルを正解データ、分析 A から分析 C それぞれの有用性指標群を持つサンプルを、当該有用性指標群に関する抽出データとした。分析 A の有用性指標群に関する抽出データの、正解データとの一致度合いが高いほど、分析 A にて抽出された有用性指標群の有効性が高いといえる。分析 B, C に対しても同様である。以上の仮定で、分析 A, B と分析 C の各クラスターのそれぞれで、適合率を (1) の式を用いて算出した。

$$\text{適合率} = \frac{\text{抽出データ中の正解データ数}}{\text{抽出データ数}} \quad (1)$$

本研究では口コミ情報という非常にノイズが多い情報を扱っているため、できるだけ商品選択に影響を与える上で非有用な情報を排除すべきだと考え、適合率を重視した。この結果を表 8 に示す。

この結果に着目すると、サンプルを詳細に分類して抽出した有用性指標群を用いるほど適合率が上昇していることがわかる。よって、本研究にて抽出された有用性指標群の抽出はサンプルを詳細に分類して因子分析を細かく実行することでより精度の高いものとなったといえる。今回、被験者のクラスタリングを実施した際、価格を重視する被験者からはうまく因子が抽出されなかった。これは、価格に関してはスペック表を見れば自身にとって適切かどうかの判断が容易であるからだと考えられる。もし価格が自身にとって適切でなかった場合、価格の次に重視する機能や特長を評価文にて判断したと考えられる。よって、価格を重視する被験者の中で評価する観点にばらつきが出たため、被験者ごとの評価文への評価もばらつきが出たと考えられる。

### 5.2 有用性指標群を用いた口コミ情報抽出の実現可能性

分析 C で抽出できた有用性指標群を用いたオンラインショッピングにおいて有用な情報抽出の実現可能性について考察する。本論文では、分析 C において抽出された因子において、因子を意味づける事柄として多く出て

きたものに着目する。まず、最も多くの因子に出現した事柄として「初めて知ったこと」が挙げられる。この事柄は、消費者の知識量に関することなので、提示する情報が初めて知ることかを判定するのは困難である。しかし、イノベーター理論に基づく消費者の革新性から消費者の消費行動を推定することで、ある程度の商品に対する知識量は推定できると考える [6]。また、多くの因子に出現した事柄として「文中の形容詞の位置」、「構成がしっかりしている」が挙げられる。これらの事柄は、口コミ情報を形態素解析することで判定できると考えられる。また、文章の構造を係り受け解析により解析し、商品に対する評価表現を抽出する研究 [7] がされていることから、これらの事柄を持つ情報が抽出できると考えられる。最後に、「否定的」の事柄について考える。文章が否定的な表現か、肯定的な表現かを判定する研究 [8] が存在することから、この事柄を持つ情報は口コミ情報からできると考えられる。

## 6. おわりに

本論文では、オンラインショッピングにおける消費者の後悔を防止するため、口コミ情報を解析し、商品選択に影響を与える口コミ情報に共通する指標群を発見した。これらの指標群を発見するために、実際のオンラインショッピングサイトにおける口コミ情報内から評価文を抽出し、それらの各評価文の有用性と評価文の特徴を実験において被験者に得点付けしてもらった。この点数を観測変数として、因子分析を 3 回実行した (分析 A, 分析 B, 分析 C)。その結果、それぞれの分析で抽出できた指標群を用いて有用な情報を抽出した場合の適合率は、分析 C がもっとも高くなった。このことから、商品選択に影響を与える口コミ情報に共通する指標群はサンプルを詳細に分類して因子分析を細かく実行する方が、より精度が高く抽出できるといえる。

今後は、抽出できた商品選択に影響を与える口コミ情報に共通する指標群を用いた、口コミ情報から商品選択において有用な情報の抽出アルゴリズムを考察する。

## 参考文献

- [1] 総務省：平成 23 年通信利用動向調査, 2011
- [2] 松尾 太加志, 中村 知靖：誰も教えてくれなかった因子分析, 北大路書房, 2002
- [3] 村瀬 洋一, 高田 洋, 廣瀬 毅士：SPSS による多変量解析, オーム社, 2007
- [4] 楽天市場：<http://www.rakuten.co.jp/>
- [5] 楽天データセット：  
<http://rit.rakuten.co.jp/rdr/index.html>
- [6] 中村 美穂, 市川 裕介, 後藤 真一郎, 黒川 裕彦, 中川 哲也：ユーザ革新性の推定手法に関する評価, 人工知能学会第 22 回全国大会 (JSAI 2008), June 2008, 3E1-03 (2008)
- [7] 二本木 智洋, 住田 一男：文の構造化による口コミ評価の分析・検索, インタラクシオン 2002 論文集, pp.175-176, 2002
- [8] 藤村 滋, 豊田 正史, 喜連川 優：文の構造を考慮した評判抽出手法, 電子情報通信学会第 16 回データ工学ワークショップ (DEWS2005), March 2005, 6C-i8 (2005)