

# レビュー類似度によるネットワーク構造分析

村松 優作<sup>†</sup> 風間 一洋<sup>††</sup> 斉藤 和巳<sup>†</sup>

<sup>†</sup> 静岡県立大学 経営情報学部 〒422-8526 静岡県静岡市駿河区谷田 52-1

<sup>††</sup> 日本電信電話株式会社 未来ねっと研究所 〒180-8585 東京都武蔵野市緑町 3-9-11

E-mail: †{b08103,k-saito}@u-shizuoka-ken.ac.jp, ††kazama@ingrid.org

あらまし レビューサイトにおいては参加者間の活発な意見交換が、ユーザの購買意欲を促進させる。しかし、ユーザは大量のレビューの中から、良いレビューを見つけなければならない。レビューサイトでは、自分のお気に入りのユーザに着目する機能が提供されており、それらの機能を用いることで、ユーザ同士が繋がり、ネットワークを形成する。その際に、ユーザ同士がどのような理由で繋がりを持つのかを知る為に、本稿においては、ユーザネットワークにおける、アイテムレビューの類似度を測定する。ユーザ間のリンクには双方向リンク、片方向リンク、リンク、全リンクの4つのパターンが存在する事に着目し、また、レビュー数によってユーザを絞る事でユーザのアクティビティ度を考慮した。化粧品の口コミサイトである@cosme、映画のレビューサイトである Yahoo!movie のデータを用いて評価を行い、その結果、ユーザ間にリンクがある方がよりレビュー類似度が高い事を確認した。

キーワード レビューサイト、ユーザネットワーク、レビュー類似度

## Analysis of network structure by review similarity

Yusaku MURAMATSU<sup>†</sup>, Kazuhiro KAZAMA<sup>††</sup>, and Kazumi SAITO<sup>†</sup>

<sup>†</sup> School of Administration and Informatics, University of Shizuoka

52-1 Yada, Suruga-ku, Shizuoka, 422-8526 Japan

<sup>††</sup> Nippon Telegraph and Telephone Corporation Network Innovation Laboratories

3-9-11 Midori-cho, Musashino-shi, Tokyo 180-8585, Japan

E-mail: †{b08103,k-saito}@u-shizuoka-ken.ac.jp, ††kazama@ingrid.org

**Abstract** In on-line review sites, customers' interests are promoted by exchanging their opinions. However, since some opinions might be inappropriate or prejudice, finding reliable users plays an important role for effectively utilizing such review sites. In this paper, by measuring the similarities of users' reviewed item sets, we explore its relationship with respect to the user network formed in review sites. More specifically, we classify user relations into four categories: all pairs, linked pairs, singly linked pairs, and mutually linked pairs. In our experiments using two review sites, @cosme and Yahoo!movie, we show that linked user pairs are likely to have high review similarity.

**Key words** review sites, user network, review similarity

### 1. はじめに

テレビや新聞などの様に、不特定多数の人間に一斉送信されるマスコミに対し、人と人との情報の共有により広まる情報伝達手段は口コミと呼ばれる。口コミはマスコミとは違い、双方向の情報の伝達が行われる。インターネットコミュニティにおける情報の共有や発信は、集まった利用者の自発的な活動を通じて行われ、共通の関心分野、価値観や目的を持ったユーザが集まることで、あるアイテムに対する情報が利用者間に広がり、その結果プロモーション効果やマーケティング効果が生まれるのである。

しかし近年のブログや SNS の普及により、情報量が爆発的に増加するとともに、信頼度が低い情報、類似情報もかなりの割合で含まれるので、自分にとって有益な情報を見つけ出す事が困難になってきた。そこで、我々が以前から注目していたのが、コミュニティを重視した新たな口コミサイトである。最近の口コミサイトの特徴として、ユーザが自分の情報を登録し、既に購入してる商品や気になっている商品の一覧が作成でき、更に他ユーザとそれを媒介にしたコミュニケーションが図れるようになっているのである。こういったコミュニティの形成には多種多様な要因が関係しているが、本稿ではコミュニティの形成要因に、ユーザが行ったレビューの類似度が関連している

事を検証する。コミュニティの形成要因を特定することは、円滑なコミュニティの形成に大きな利益をもたらさだろうと考えられる。

オンラインのレビュー口コミサイトに関する研究が幅広く行われている (e.g., [1])。しかしながら、ユーザ間のネットワークとレビュー類似度の関連については、あまり多くの研究は行われていない。一方で、ユーザ間ネットワークにおける欠損リンクの推定には、ある種の類似度が有効との報告もある [2]。そこで本研究では、ユーザ間ネットワークのリンク関係とレビュー類似度について、実データを用いて系統的に調査検討する。

本稿の構成は次の通りである。2 節では、レビュー類似度測定手法について述べる。3 節では類似度測定の詳細について述べるとともに、分析結果を考察する。最後に、4 節で本稿のまとめと今後の課題について述べる。

## 2. ユーザレビュー類似度測定手法

本節では、ユーザ間ネットワークとレビュー類似度について説明するとともに、ユーザレビュー類似度測定手法について述べる。

### 2.1 ユーザ間ネットワーク

レビューサイトなどにおけるユーザ集合を  $U = \{u_1, \dots, u_M\}$  とする。ここで  $M$  は全ユーザ数を表す。このようなサイトにおいて、発言が自分に嗜好と一致したり、有用な場合に各ユーザ  $u_m \in U$  は他のユーザ  $u_n \in U$  を「ファン」または「お気に入り」などとして登録できる。以下では、このような登録関係にあるユーザのペアを  $(u_m, u_n)$  として順序付き関係を定義する。いま、全てのユーザについて登録関係にあるペアの集合を  $E \subset U \times U$  とすれば、グラフ  $G = (U, E)$  を定義することができる。以下では、このグラフのことをユーザ間ネットワークと呼ぶ。

### 2.2 レビュー類似度

レビューサイトなどにおいて、映画や化粧品などレビューの対象集合を  $T = \{t_1, \dots, t_N\}$  とする。ここで  $N$  は全レビュー対象数を表す。いま、あるユーザ  $u_m \in U$  が対象  $t_n \in T$  をレビューしたとき、このペアを  $(u_m, t_n)$  として順序付き関係を定義する。すると、上記のユーザ間ネットワークと同様にして、全てのユーザについてレビュー対象とのペアの集合を  $D \subset U \times T$  とすれば、2部グラフ  $F = (U \cup T, D)$  を定義することができる。

この2部グラフを用いれば、ユーザ  $u_m \in U$  がレビューした対象集合  $R(u_m) \subset T$  を以下のように定義できる。

$$R(u_m) = \{t_n : (u_m, t_n) \in D\} \quad (1)$$

以下では、ユーザ  $u_m$  のレビュー対象集合  $R(u_m)$  に基づきレビュー類似度を定義する。本研究は、どのように類似度を定義するかに依存しない枠組みとなっているものの、このような目的で幅広く利用されている Jaccard 係数に着目して検証評価を行う。すなわち、ユーザ  $u_m$  と  $u_n$  のレビュー対象集合  $R(u_m)$  と  $R(u_n)$  に対して、Jaccard 係数を次式で求めるとする。

$$J(u_m, u_n) = \frac{|R(u_m) \cap R(u_n)|}{|R(u_m) \cup R(u_n)|} \quad (2)$$

ここで、 $|S|$  は集合  $S$  の要素数を表す。明らかに、 $0 \leq J(u_m, u_n) \leq 1$  となる。

### 2.3 ユーザレビュー類似度測定法

ユーザレビュー類似度測定法について述べる。いま、測定対象となるユーザペアの集合を  $H \subset U \times U$  とする。ここで、測定対象集合  $H$  は、全ユーザペアとしたり、ユーザ間ネットワークのリンク集合  $E$  に限定するなど多様に選定できるが、その選定法などについては、評価実験の節で詳述する。

本検討では、与えられたユーザペア集合  $H$  に対して、以下で定義する累積分布  $P(x : H)$  に基づきユーザレビュー類似度を測定する。

$$P(x : H) = \frac{|\{(u_m, u_n) \in H : J(u_m, u_n) > x\}|}{|H|} \quad (3)$$

ここで、パラメータ  $x \in [0, 1]$  は類似度の閾値を規定する。また、 $P(x : H)$  は単調現象で、 $P(0 : H)$  が最大で  $P(1 : H) = 0$  となる。すなわち、パラメータ  $x$  に対して、累積分布  $P(x : H)$  をプロットしたとき、 $x$  が比較的小さい段階で  $P(x : H)$  の減少が大きければ、測定対象  $H$  は相対的に類似度の低いユーザペアの集合となり、これに対して、 $x$  が比較的大きくなって  $P(x : H)$  の減少が小さければ、測定対象  $H$  は相対的に類似度の高いユーザペアの集合となる。

## 3. @cosme・yahoo!movie によるレビュー測定

本節では類似度測定の詳細について述べるとともに、分析結果を考察する。

### 3.1 評価に使用したデータセット

本稿では、@cosme と Yahoo!movie のデータを用いてユーザ間ネットワークを抽出し、「双方向リンク」、「片方向リンク」、「リンク」、「全リンク」の場合についてレビュー類似度を評価する。@cosme のデータは、2008 年 12 月から 2009 年 12 月の期間のものであり、Yahoo!movie のデータは、2008 年 6 月から 2009 年 6 月の期間のものである。分析対象データの基礎統計量を表 1 に示す。@cosme においては、ユーザ数が約 45000 人存在し、ユーザ同士は「お気に入りメンバー」「fan」という形で繋がっている。「お気に入りメンバー」というのは、ユーザが自ら気に入ったユーザを登録できるものであり、「fan」というのは逆に他のユーザから自分がお気に入り登録されるものである。yahoo!movie においても同様に「お気に入りレビュアー」「ファン」という形で繋がっている。ユーザは自分のお気に入りのユーザを任意で登録出来るが、逆にお気に入り登録されるのは、相手のユーザ次第という事になる。ユーザ間において双方向のリンクがある場合を「双方向リンク」、片方向にしかリンクが無い場合「片方向リンク」、双方向が片方向のリンクがある場合を「リンク」、全ユーザのペアを対象とした場合を「全リンク」とした。この4つの種類別で分析を試みる。また、ユーザのアクティビティ度を考慮して分析を行う為、レビュー数が1以上、10以上、100以上という場合についてユーザを絞り、上記の4つのリンク種の場合について評価する。

表 1 @cosme,yahoo!movie 基本統計量

データ名	ノード数	リンク数	平均次数	リンク密度	SCC 数	コリンク率
@cosme	45,024	351,299	7.802	0.00035	219	0.44312
yahoo!movie	5,524	76,314	13.815	0.00500	7	0.56076

以下では、上記の分類に従い測定対象となるユーザペアの集合  $H$  を定義する。まず、双方向リンク (mutually linked pairs) での定義  $H_m$  は以下となる。

$$H_m = \{(u_m, u_n) : (u_m, u_n) \in E \wedge (u_n, u_m) \in E\}$$

アクティビティ度の最小値を  $k$  とすれば、双方向リンクでの測定対象のユーザのペア集合  $H_m(k)$  は以下となる。

$$H_m(k) = \{(u_m, u_n) \in H : |R(u_m)| \geq k \wedge |R(u_n)| \geq k\}$$

なお、アクティビティ度を考慮した測定集合の定義は、上式と同様に行えるので以下では割愛する。次に、片方向リンク (singly linked pairs) での定義  $H_s$  は以下となる。

$$H_s = E \setminus H_m$$

ここで、 $\setminus$  により集合差を表す。また、リンク (linked pairs) での定義  $H_l$  は以下となる。

$$H_l = \{(u_m, u_n) : (u_m, u_n) \in E \vee (u_n, u_m) \in E\}$$

最後に、全リンク (all pairs) の定義  $H_a$  は  $H_a = U \times U$  となる。

3.2 リンクの種類、ユーザのアクティビティ度による変化  
以下に、「双方向リンク」、「片方向リンク」、「リンク」、「全リンク」の比較とともに、ユーザのアクティビティ度による変化を比較した結果について述べる。また、@cosme のデータと Yahoo!movie のデータでの違いを検証する。

@cosme のレビュー類似度の結果を図 1, 図 2, 図 3 に示す。図 1 は、レビューを 1 回以上行ったユーザが対象であり、図 2 については、10 回以上、図 3 は 100 回以上である。

Yahoo!movie のレビュー類似度の結果は図 4, 図 5, 図 6 であり、それぞれ図 4 が 1 回以上のレビューをしたユーザが対象、図 5 は 10 回以上、図 6 は 100 回以上である。

分析結果において、@cosme のグラフ図 1~図 3 では、ユーザのアクティビティ度が高くなるにつれて、リンクの有無の差が広がってきている。特にレビュー回数が 1 以上と 10 以上での差が大きく出ており、「双方向リンク」の場合が類似度が高い結果が出た。Yahoo!movie のグラフ図 4~図 6 の場合では、レビュー数が 100 以上のユーザに絞ると、リンクの有無による類似度の差がなくなっていることが分かる。また、2 つの違いを比べてみると、@cosme の場合はよりアクティビティ度の高いユーザが、自分が行ったレビューの類似度がより高いユーザとリンクしており、Yahoo!movie の場合は、アクティビティ度の高いユーザは、自分が行ったレビューの類似度と関係なくユーザとリンクしていた。ユーザのアクティビティ度が高くなるほどリンクの種類による類似度の差がなくなる為、自分とは

違うレビューを行ったユーザ、つまり多種多様な意見を取り入れる事を考え、自分とは嗜好の違うユーザをリンクしている可能性もある。@cosme と Yahoo!movie のレビューサイトとしての違いから結果を見てみると、@cosme のユーザは、自分の年齢や肌質に合った化粧品を手に入れる事を目的の一つとし、レビューサイトを参考にしていると考えられる。その為、基本的には自分と同じようなレビューを行ったり、自分と同じような属性や悩みを持ったユーザに興味を持つ事が推測できる。しかし映画のレビューサイトである Yahoo!movie においては、@cosme の様な事が全てに言える訳ではない。勿論、映画においても自分と同じような好みの映画やお気に入りの俳優を考慮してユーザをリンクしている場合もあるだろうが、ユーザのアクティビティ度がより高いユーザはレビュー類似度とは関係なくユーザをリンクしている為、人の属性や興味によって一様に決める事はできないだろう。2 つのジャンルの違うレビューサイトで検証を行って違う結果が出た様に、レビューサイトの扱う内容によって人間関係に違いが出てくるのかもしれない。

#### 4. ま と め

我々は、ユーザ間のレビュー類似度がコミュニティの形成要因の一因となっている事を検証した。ユーザ間のリンクがある場合とない場合では、リンクがある場合がレビュー類似度が高く、コミュニティの形成要因となると考えられる。本稿では、レビュー類似度の測定手法を述べ、実際に@cosme と yahoo!movie のデータを用いて分析を行った。分析内容としては、4 パターンのリンクで検証を行い、そしてユーザアクティビティ度を考慮する為、レビュー回数でユーザを絞って比較を行った。今回の結果において、リンクがない場合よりもリンクがある場合の方がレビュー類似度が高かったが、アクティビティ度を考慮すると、扱う情報によって違いが出た。よって、レビューサイトで扱っている内容も考慮して分析する事が必要である。今後の課題としては、ユーザアクティビティ度を考慮した場合の変化を、ネットワークの平均次数や密度など、調査対象ネットワークの性質から分析し、他のユーザネットワークにおいても引き続きレビュー類似度を検証する。

#### 謝 辞

本研究は、科学研究費補助金基盤研究 (C) (No. 22500133) の補助を受けた。

#### 文 献

- [1] D. Crandall, D. Cosley, D. Huttenlocher, J. Kleinberg, S. Suri. Feedback Effects between Similarity and Social Influence in Online Communities. Proc. 14th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining, 2008.
- [2] D. Liben-Nowell, J. Kleinberg. The Link Prediction Problem for Social Networks. Proc. 12th International Confer-

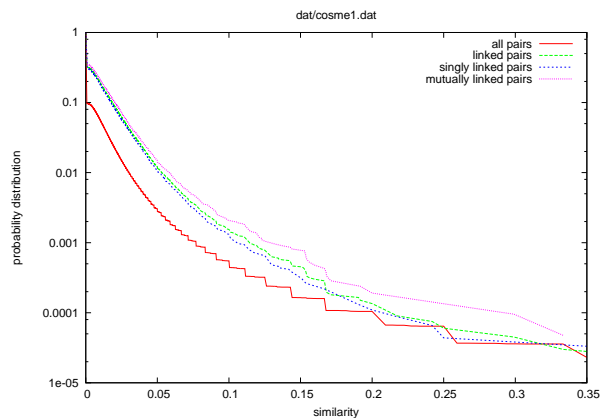


图 1 @cosme(1)

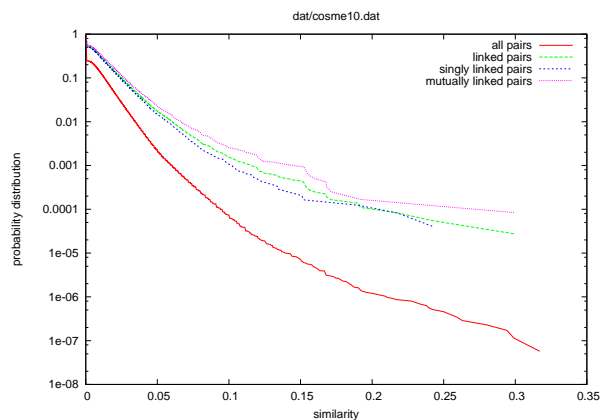


图 2 @cosme(10)

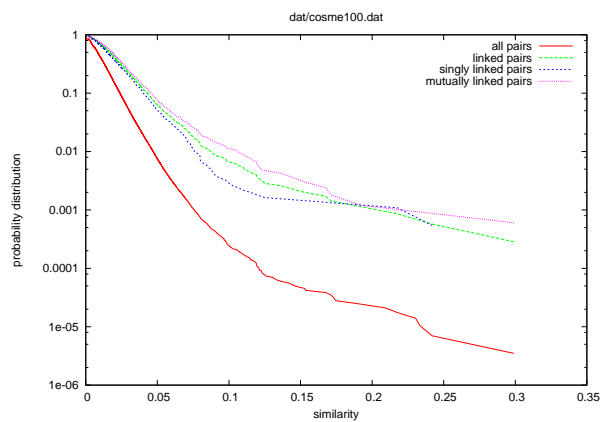


图 3 @cosme(100)

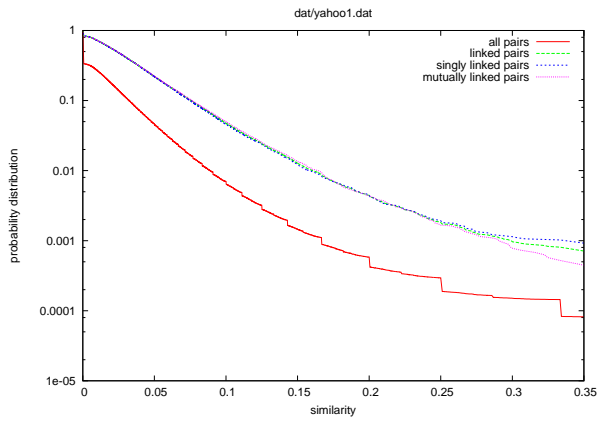


Figure 4: Yahoo!movie(1)

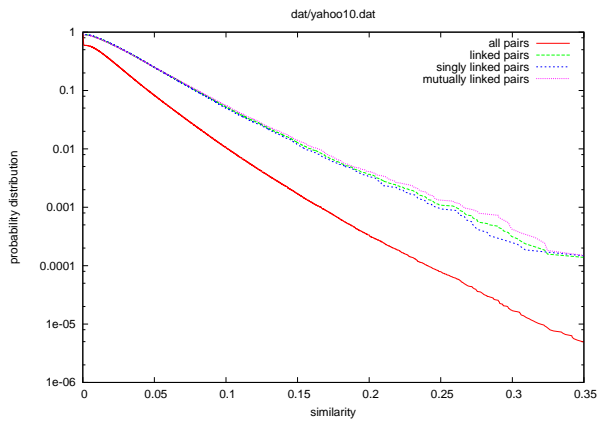


Figure 5: Yahoo!movie(10)

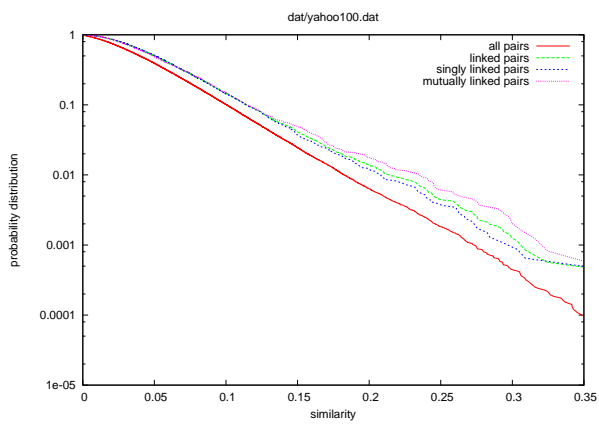


Figure 6: Yahoo!movie(100)