

ネットワーク構造を対象とした 特徴量抽出とその応用

羽室 行信, 中原 孝信

近年、テキストデータや顧客データを始め、アクセスログ、ソーシャルメディア、そしてセンサーデータなど、膨大な量、非定型な形式、そして多様性をもったビッグデータの収集が行われ、その利活用への期待が寄せられている。本稿は、グラフ構造を利用した二つの応用研究を示す。一つ目は、Twitter データに対して単語の類似度グラフを日々構築し、グラフのクラスタリングを行うことで構造変化の抽出を目的にした研究である。二つ目は、ID 付き POS データを利用した商品購買グラフを構築し、グラフ構造を用いた特徴抽出に関する研究である。

キーワード：マイクロクラスタリング, NetSimile 法, グラフ特徴量, データ研磨

1. はじめに

近年、グラフ構造を対象にした研究が盛んになってきた。データマイニング分野では、ソーシャルネットワーク、ハイパーリンク、そして遺伝子ネットワークなどを対象にしたリンクマイニングと呼ばれる研究が行われている [1]。そこでは、リンク関係を用いたノードのクラスタリング、ノードのランキング、リンクの予測などが研究されている。一方で、社会ネットワーク分析や複雑ネットワークに関する研究は古くから行われており、要素間の関わりについて、複雑なシステムをモデル化し解析することで、その相互作用や因果関係などの構造的な特徴を明らかにしてきた [2, 3]。いずれの分野にも共通する点は、個々のリンクだけに着目するのではなく、グラフ構造全体を扱うことで初めて明らかになる関係性に着目していることである。

また、グラフ構造は、本来備わっている構造をネットワークとして表現する場合と、関係性を距離関数で定義してグラフ構造として表現する場合がある。前者は、ソーシャルネットワークとしての人と人のつながりや、ハイパーリンクによるページとページのとつながり、企業間取引ネットワークによる企業と企業のとつながりなどであり、本来備わっている構造をネットワークと

して表現している。一方で後者は、商品や単語など事物の関係性を距離関数で定義することにより、グラフ構造として表現したものであり、コンビニエンスストアなどの POS データ、ニュース記事や Twitter の投稿内容などのテキストデータの解析に利用される。たとえば、POS データの場合は、複数の同時購買の関係性をつないでいくことで、商品の類似性に関する巨大なネットワークが生成できる。また、ニュース記事や Twitter などのテキストデータも同一文書内に出現する単語の共起関係を定義し、共起頻度がある閾値よりも高い場合に単語間に枝を張ることで、文章に出現する単語の関連性をネットワークとして表現できる。

本稿は、ID 付き POS データ、Twitter データを対象に、共起関係としての距離関数を定義することで商品間の関係、単語間の関係をグラフ構造で表現する。そして、それぞれのグラフ構造から特徴的な構造を抽出する方法を示し、その方法を用いた応用研究を紹介する [4, 5]。

本稿では、2 節でグラフ構造を対象にしたデータ研磨と呼ばれるグラフクリーニングの方法と、Netsimile 法と呼ばれるグラフ構造からの特徴量の抽出方法を紹介し、それらを利用した応用研究を 3 節、4 節でそれぞれ紹介する。

2. 類似度グラフと解析技術

グラフ構造でデータを表現するために、共起関係を利用した距離関数を定義し、その値がある閾値以上の場合に節点間に枝を張ることでグラフ構造を構築する。これを類似度グラフと呼ぶ。次にこの類似度グラフを対象にしたデータ研磨と、グラフの特徴量を計算する

はむろ ゆきのぶ
関西学院大学経営戦略研究科
〒 662-8501 兵庫県西宮市上ヶ原一番町 1-155
hamuro@kwansei.ac.jp
なかはら たかのぶ
専修大学商学部
〒 214-8580 神奈川県川崎市多摩区東三田 2-1-1
nakapara@isc.senshu-u.ac.jp

NetSimile 法を紹介する。

2.1 類似度グラフ

以下では、Twitter データのツイートを対象に、単語の類似度グラフを作成する方法を説明する。単語の類似度グラフとは、単語を節点で表し、類似した単語間に枝を張った無向グラフである。ここで類似度は、単語の共起情報に基づいて定義される PMI (Pointwise Mutual Information) を用いる。単語 u の生起確率を $p(u)$ 、単語 v との共起確率を $p(u, v)$ で表すと、 u と v の距離関数 PMI は式 (1) で定義される。

$$\text{PMI}(u, v) = \log_2 \frac{p(u, v)}{p(u)p(v)} \quad (1)$$

PMI の値が 0 より大きければ、二つの単語は共起しやすく、0 より小さければ共起しにくいと解釈できる。そしてユーザが指定した最小 PMI の γ について、 $\text{PMI}(u, v) \geq \gamma$ を満たすような二つの単語 u, v に枝を張る。 γ を小さな値にすると密なネットワークとなり、逆に大きな値にすると疎なネットワークが構成されることになる。

この方法は ID 付き POS データにも利用可能である。たとえば、ユーザの購入した商品を u 、その生起確率を $p(u)$ 、商品 v との共起確率を $p(u, v)$ で表すと同じように式 (1) が適用できる。距離関数については、式 (1) の PMI 以外にも Jaccard 係数 [6] が利用されることも多い。

2.2 グラフ研磨

類似度グラフは、互いに類似した節点群には密に枝が張られ、逆に類似度の低い節点群には枝が張られにくくなり疎な構造となる。そこで、類似度グラフの密な部分グラフをクラスタとして抽出することで、互いに関係性の類似したコミュニティ構造を得ることができる。一般グラフのクラスタリングについては、ニューマンクラスタリング、グラフ分割、極大クリーク列挙など、これまでさまざまな手法が提案されてきたが、どの手法も問題点を抱えており、決定打になっていないというのが現状である。

たとえば極大クリーク列挙では、現実データにおいては多くの場合、非常に多数の類似した極大クリークが列挙されてしまうという問題がある。列挙された極大クリークの類似関係を用いて、極大クリークをさらにクラスタリングするという方法も提案されているが、列挙される極大クリークの数によっては計算量が問題となる。このような問題の多くは、そもそも対象とするグラフにノイズが含まれるために起こる問題とも考えられる。

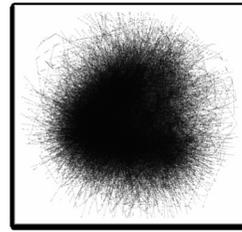


図 1 オリジナルネットワークの概略図

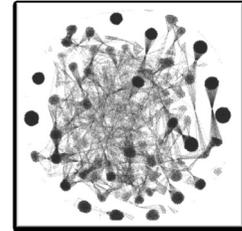


図 2 データ研磨後の概略図

そこで、宇野らは、対象とするグラフをクリーニングする「グラフ研磨」手法を提案している [7]。これは、グラフをクラスタリングする前に、枝を張り直すことでグラフを再構成し、できる限り構造を明確化しておくというものである。図 1 はオリジナルネットワークで、それに研磨を適用することで図 2 に例示されるように、直感的には枝密度の濃い部分グラフはより濃く、薄い部分グラフはより薄くするというものである。

グラフを研磨することの利点の一つとして、グラフ構造が明確化されるために、列挙されるクリークの数劇的に少なくなることが挙げられる。本研究の育児休暇に対する意見抽出で利用したデータにおいても、研磨前のグラフに比べて平均約 89.5% の削減効果が確認されている。

研磨のアルゴリズムを Algorithm 1 に示す。ここに示すアルゴリズムは、効率よりも理解のしやすさを優先させた記述となっている。効率的なアルゴリズムについては文献 [7] が詳しい。研磨の方法は至ってシンプルで、すべての節点ペアについて、その類似度 ($\text{sim}(u, v)$) がユーザの指定した閾値以上であれば接続し、そうでなければ接続しないというルールに従って、新たなグラフを再構成する。

類似度としてはさまざまな定義を用いることができるが、ここでは類似度グラフの構築で用いた PMI とした。グラフ上での二つの節点 u, v の類似度 PMI は、式 (1) の定義において $p(u) = |N(u)|/|V|$ 、 $p(u, v) = |N(u) \cap N(v)|/|V|$ として出現確率を定義したものに

Algorithm 1 グラフ研磨アルゴリズム

```

1: function POLISHING( $G = (V, E), \sigma$ )
2:    $V$ : 節点集合,  $E$ : 枝集合,  $\sigma$ : 類似度下限値
3:    $E' = \phi; V' = \phi$  ▷ 研磨後の枝集合と節点集合の初期化
4:   for all  $u \in V$  do
5:     for all  $v \in V$  do ▷ # 全節点ペア  $u, v$  について調べる
6:       if  $sim(u, v) \geq \sigma$  then ▷ 節点ペア  $u, v$  が似ていれば新たに枝として加え、似ていなければ加えない
7:          $E' = E' \cup \{(u, v)\}$ 
8:          $V' = V' \cup \{u\}$ 
9:          $V' = V' \cup \{v\}$ 
10:      end if
11:    end for
12:  end for
13:  return( $V', E'$ )
14: end function

```

相当する。ここで $N(u)$ は節点 u の近傍節点（直接接続のある節点）の集合を、 V は研磨対象のグラフを構成する節点集合を表している。すなわち大雑把に言えば、共通する近傍節点が多い節点間に枝が張られ、少ない節点間の枝は切断される。これは、SNS における友達紹介のアルゴリズム（すなわち共通する友達が多い二人は友達である可能性が高い）と同様なもので、グラフ構造のプリミティブな変化予測（リンク予測）を行っているとも解釈できる。そして新たに構成されたグラフを入力として同様の研磨手法を繰り返し適用し、グラフの構成に変化がなくなるか、もしくはユーザの指定した最大繰り返し回数に達すれば終了する。最終的に得られたグラフが研磨グラフである。この研磨グラフから列挙された極大クリークをマイクロクラスと呼ぶ。

2.3 NetSimile 法

NetSimile 法は、複数のネットワーク間の類似度を測定するために提案された手法である [8, 9]。この手法は、1) 異なるサイズのネットワークに適用でき、2) 枝数に線形な時間で計算でき、3) 節点や枝の対応関係がなくてもよい、といった特徴をもつ。

NetSimile 法は、1) 節点特徴量の抽出、2) グラフ特徴量の集約、3) グラフの類似度計算、の三つのステップから構成される。本稿で紹介する応用研究では、複数ネットワークの類似度を測定する目的に NetSimile 法を利用するのではなく、最初の二つのステップで得られるグラフ特徴量を利用する。以下では、本稿に関連するステップ 1), 2) について説明する（図 3 に概略図を示す）。

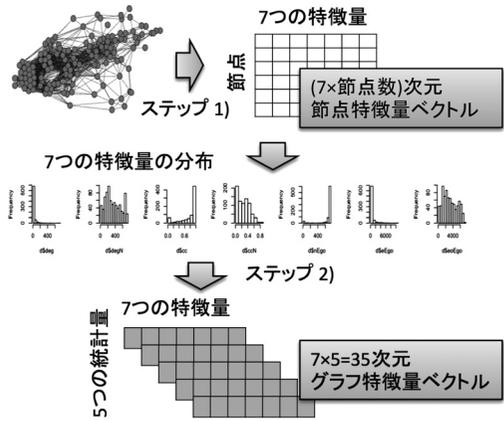


図 3 NetSimile の概略図

2.3.1 NetSimile 法による節点特徴量の抽出

類似度グラフ $G = (V, E)$ のすべての節点 $v \in V$ について、以下に定義される七つの特徴量を計算する。これらの特徴量は節点に定義されるため、特に「節点特徴量」と呼ぶことにする。以下の定義において、 $ego(v)$ は節点 v のエゴ・ネットワーク（「エゴネット」と略称する）を表しており、ここでは節点 v および節点 v から 1-hop で到達できる節点集合から誘導される G の部分グラフのことである。

1. 次数 $d_v = |N(v)|$: 節点 v と接続のある節点数。
2. クラスタ係数 $c_v = \binom{d_v}{2}^{-1} |\{(u, w) | (u, w) \in E, u, w \in N(v)\}|$: 近傍節点間の枝の数を近傍節点の二つの組合せで割ったもの。
3. 近傍平均次数 $\bar{d}_{N(v)} = \frac{1}{d_v} \sum_{u \in N(v)} d_u$: 近傍節点の平均次数。
4. 近傍平均クラスタ係数 $\bar{c}_{N(v)} = \frac{1}{d_v} \sum_{u \in N(v)} c_u$: 近傍節点の平均クラスタ係数。
5. エゴネット枝数 $eego_v$: エゴネット $ego(v)$ 内の枝の数。
6. エゴネット接続枝数 $eego_v^o$: エゴネット $ego(v)$ に接続される枝の数。
7. エゴネット近傍節点数 $nego_v$: エゴネット $ego(v)$ の近傍節点数。

2.3.2 グラフ特徴量の集約

前節の節点特徴量を抽出した段階で、節点 × 特徴量行列が得られるが、次のステップでは、これらの特徴量を集約することでグラフ全体の特徴量（「グラフ特徴量」と呼ぶことにする）を求める。グラフ特徴量は、七つの節点特徴量それぞれについて、節点をサンプルと考えた場合の分布により定義される。ここには節点特徴量の分布の形状によりそのグラフを識別するとの

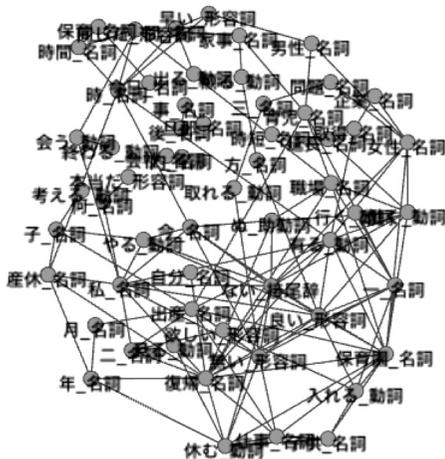


図 4 単語の類似度グラフ

仮定がある。そして NetSimile 法では、分布の形状は、中央値、平均値、標準偏差、歪度、尖度の五つの統計量によって要約される。以上の集約により、七つの節点特徴量について五つの統計量の 35 次元特徴量ベクトルが得られ、これをグラフ特徴量として用いる。

3. 育児休暇に対する意見抽出

本節では、育休についての Twitter 投稿に注目し、類似度グラフ (2.1 節) とグラフ研磨手法 (2.2 節) を利用して一般の人々の声を要約する方法を示す。具体的には、安倍政権の育休 3 年の要請という発言 (2013 年 4 月 18 日) によってユーザの話題がどのように変化したかを解析する。利用したデータは「育休」「育児休暇」のいずれかを含む 2013 年 4 月 10 日から 10 日間につぶやかれた約 26,000 ツイート (約 6,400 ユーザ) を用いた。

ツイートは、1 週間を単位として 1 日ごとにずらした移動窓を設定し、それぞれの単位で単語の類似度グラフを構築する (図 4)。そして、グラフ研磨を適用することで、グラフのクリーニングを実施する。研磨した類似度グラフを図 5 に示す。また、研磨グラフから抽出したマイクロクラスタを表 1 に示す。

研磨しない類似度グラフからクラスタを抽出した場合は 77 個のクラスタが抽出されたが、研磨後のグラフのクラスタ数は 14 に減り、1 クラスタに含まれる単語数は増えた (平均 4.64 語、26 語からなるクラスタを除くと平均 3.0 語)。また似たようなクラスタが複数列挙されるという問題も回避できている。

次にツイート内容の変化検出として、単語クラスタ (研磨された類似度グラフの極大クリーク) を構成する

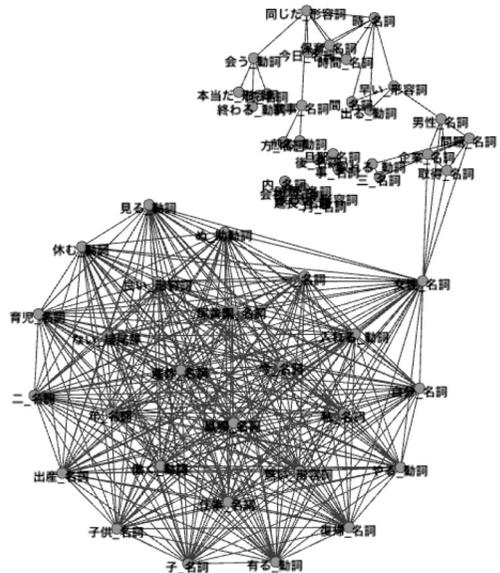


図 5 単語の類似度グラフを研磨したグラフ

表 1 研磨したグラフから抽出したマイクロクラスタ

{会社, 内}	{今日, 保育, 同じだ, 時, 時間}
{事, 後, 旦那}	{企業, 取得, 問題, 女性, 男性}
{延長, 時短}	{三, 取れる, 問題}
{早い, 男性}	{会う, 何, 本当だ, 終わる}
{会う, 同じだ}	{出る, 早い, 時, 間}
{月, 欲しい}	{家事, 方, いる}
{同じだ, 家事}	:

単語がどのように時系列で変化したかをとらえる。本稿では図 6 に示す Sankey ダイアグラムによって視覚化することで変化を主観的にとらえる。Sankey ダイアグラムとは、閉路のない有向グラフ (DAG) を視覚化する手法の一つで、枝の重みとして定義される流量が節点間でどのような割合で流れていくかを直感的に理解することができ、送電ネットワークの視覚化などに利用される。

ある期における研磨グラフの各クラスタを節点と考え、次の期の各クラスタと共通する単語数を流量として Sankey ダイアグラムを描画する。図 6 では、節点 (クラスタ) は棒で示されているが、その高さはクラスタに含まれる単語数に対応する。そして同じ期のクラスタはすべて同じ水平位置に描画されている。このチャートから、ツイート内容について以下の三つの性質を読み取ることができる。

- 内容変化 (枝の錯綜) : 枝の分岐が多い場合、単語の結びつきに変化が生じたということを意味し、全体としての投稿内容に何らかの変化が生じた

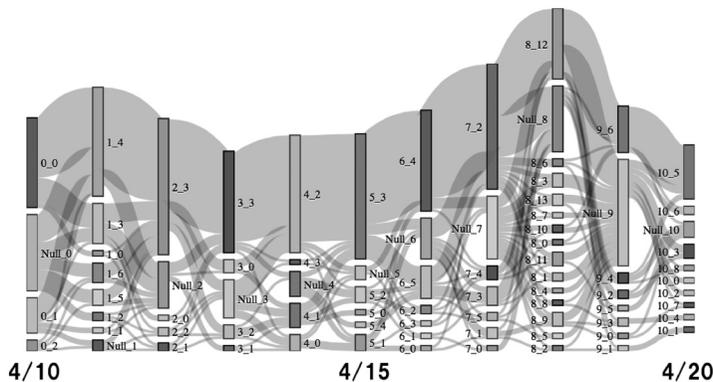


図6 Sankey ダイアグラム

考えられる。

- 多様性（節点の高さ）：ある期におけるすべての棒の合計が相対的に長くなるということは、それだけ多様な単語が利用されていることを意味し、意見に多様性が出てきたと考えられる。
- 独立性（節点の多さ）：ある期における節点数が多い場合、単語の結びつきが細分化されたことを意味し、ユーザによって投稿される内容が分化してきたことが伺える。

4月10日から20日のSankeyダイアグラムを図6に示す。ここで、図中、“Null”へ流れる単語はどのクラスタからも消えた単語で、また“Null”から流れる単語は新たにクラスタに現れた単語を意味する。10日から11日にかけてもクラスタ構造の変化が見取れるが、クラスタの多様性、独立性の観点から見て4月18日が突出しているのがわかる。安倍首相の発言を受けてツイート数自体が増加したこともあるが、Twitterで展開される話題に大きな変化があったことが読み取れる。また、マイクロクラスタに含まれる単語の内容から、話題の時系列差異を分析したところ、安倍首相発言前後で、4月17日には{出産、為、取れる、ある、本当だ}、{ママ、頑張る、出す、...}など、育児を取得している、取得しようとしているユーザによると思われる話題が抽出されていた。ところが安倍首相発言のあった4月18日には{企業、取得、問題、男性、女性}や{3、取れる、問題}など、安倍首相の発言を受けたと思われる話題が現れ、4月19日になると{男性、育児}、{自分、子、期間、考える}など、ツイートの増加に合わせて意見の表明や議論が進んでいることを思わせる話題が抽出されるようになった。

このように、Twitterに投稿されたツイートの話題に関する変化検出、および話題要約の手法として、研

磨した類似度グラフからのクラスタリング手法（極大クリーク列挙）にはいくつかの興味深い知見を得ることができた。

4. 購買履歴データからのグラフ特徴量の抽出

本節では、あるスーパーマーケットの2013年7月から2014年6月までの1年間で品川と横浜にそれぞれある1店舗の購買履歴データを利用する。これらは共に同程度の売上と床面積をもっており、ネットワーク特徴量を利用し比較することで二つの店舗の違いを明らかにする。

分析対象とする顧客は、各店舗で1年間の来店回数が50回以上90回未満を一般顧客、90回以上来店のある顧客を優良顧客と定義した。NetSimile法により求めた節点特徴量、グラフ特徴量を説明変数とし、顧客の来店回数の多寡を目的変数とした罰則付きロジスティック回帰モデルを構築する。2.1節の方法で、顧客ごとに日別の購入商品群から類似度グラフを作成した。その際の最小PMIの γ は、横浜店、品川店ともに0.005と定めた。次に、各顧客から生成した類似度グラフにNetSimile法を適用し、節点ごとに7次元の節点特徴量とグラフ全体の特徴量として35次元のグラフ特徴量を生成した。それ以外にも類似度グラフから、節点数、枝数、枝密度、を計算し説明変数として利用した。類似度グラフ以外の説明変数は、1商品の購入頻度(1-item)、2商品の共起頻度(2-item)を用いた。これらの説明変数を利用して、優良顧客の判別を目的に罰則付きロジスティック回帰モデルを構築した。

表2は特徴量別の正解率を示している。正解率は予測したクラスと実際のクラスが一致した割合である。 $\#p$ は選ばれた説明変数の数を示す。ランダムに予測

表 2 特徴量別正解率

特徴量	横浜店		品川店	
	正解率	#p	正解率	#p
1-item	0.856	115	0.900	211
2-item	0.727	80	0.948	284
節点数	0.626	1	0.599	1
枝数	0.609	1	0.582	1
枝密度	0.636	1	0.595	1
グラフ特徴量	0.681	3	0.774	17
節点特徴量	0.670	50	0.799	112

表 3 横浜、品川店に共通する変数

商品名	オッズ比	中分類	売上ランク
リンゴデニッシュ	1.091	パン	215
粒チョコ	1.361	パン	380
ロースハム 3P	1.011	加工肉	2
しらす干 並	1.003	塩干	10
のり弁当	1.004	惣菜	2
コロッケ	1.001	惣菜	3
しらたき	1.045	日配	46
こんにゃく 黒	1.023	日配	61
野菜かき揚げ	1.002	惣菜	12
いなり 3ヶ入	1.038	惣菜	96
にんじん	1.017	野菜	9
長ねぎ	1.011	野菜	13
アスパラ	1.018	野菜	16
キャベツ 1/2 切	1.008	野菜	66
さんま	1.033	鮮魚	8

した場合の正解率は品川店 0.52, 横浜店 0.59 である。横浜店、品川店の両方で 1-item の正解率が最も高いが、選択されている説明変数の数は 115 個、211 個とそれぞれ多く、すべての説明変数を解釈することは困難である。

そこで、品川店と横浜店の回帰モデルで共通して得られた 1-item の結果から、優良顧客に寄与している変数を抜き出したものが表 3 である。売上ランクは中分類別の売上金額のランキングを示している。両店舗に共通する商品で売上金額ランキングの高い商品は、惣菜と加工肉で「のり弁当」、「コロッケ」、「ロースハム」は他の顧客と同様に優良顧客にも好まれる商品である。一方でパンカテゴリの中でもランキングの低い商品である「リンゴデニッシュ」、「粒チョコ」は優良顧客が好む商品であり、売上ランキングが下位でもストックの必要な商品と判断できる。次に品川店にのみ出現し優良顧客に寄与する商品は、「プチチョコビスケット」や「チョコボールキャラメル」などの菓子と、タバコである。これは品川店というビジネス街にある店舗が

表 4 各店舗で優良顧客に寄与するグラフ特徴量

品川店			横浜店		
特徴量	統計量	回帰係数	特徴量	統計量	回帰係数
ccN	mean	-13.58	ccN	mean	-5.56
degN	kurt	-0.0088	ccN	median	-2.84
deg	median	-0.0036	ccN	skew	-0.034
deg	kurt	-0.0033			
deg	sd	0.0082			
deg	skew	-0.013			
eEgo	mean	0.00019			
eEgo	kurt	-0.00067			
eEgo	skew	6.31			
eoEgo	median	-6.37			
eoEgo	kurt	0.0068			
ccN	kurt	0.044			
nEgo	median	0.00054			
ccN	sd	0.094			
cc	median	-0.10			
cc	kurt	0.035			
cc	sd	-0.60			

特徴量は、deg：次数、cc：クラスタ係数、degN：近傍節点の平均次数、ccN：近傍節点の平均クラスタ係数、eEgo：ego ネットワークの枝数、eoEgo：ego ネットワークに接続された枝数、nEgo：ego ネットワークに接続された節点数を表す。

その要因であると考えられる。また、横浜店の優良顧客に特徴的な変数は、「木綿豆腐」や「極小粒納豆」などの日配と、「そうめん専科」、「カルボナーラ」などの調味料が多く出現しており、賞味期限の短い日配や調味料を購入する主婦層が優良顧客として考えられる。

次に、グラフ特徴量を見ると、驚くべきことに、節点特徴量との比較においてより少数の変数によって同等の精度を達成していることがわかる(表 2)。特に横浜店では、三つのグラフ特徴量で節点特徴量のモデル精度をしのいでいる。この結果は、個々の商品についての購買行動を見なくても、商品全体の関係性に顧客の購買行動の特徴が現れ、それが優良顧客の内在的な購買行動として現れていると考えられる。また表 4 に各店における選ばれたグラフ特徴量一覧を示す。横浜店と品川店で共通して現れるグラフ特徴量は、近傍クラスタ係数平均だけであり、優良顧客化の要因に関する一般理論の導出には至らないが、近傍密度が優良顧客の購買行動に影響を与えていることが伺える。

来店回数から優良顧客を定義し、優良顧客の購買行動を 1 商品の購入頻度や 2 商品の共起頻度を利用して表層的な特徴としてとらえた場合と、商品間グラフ構造を利用して内在的な特徴としてとらえた場合の両方で結果を紹介した。表層的な関係性は、購買の直接的な影響をとらえることが可能であり、売上ランクは高くないが、優良顧客に好まれる商品を明らかにした。一方で、商品間のグラフ構造については、NetSimile を

適用することでグラフ特徴量を算出し、ある商品から接続されている商品間の関係性が密になることが優良顧客の購買行動として明らかになった。

5. おわりに

本稿では、Twitter から生成した単語の類似度グラフの構造変化をとらえた。またスーパーマーケットの購買に対しては、表層的な関係だけではなく、商品購買グラフからグラフ特徴量を利用しモデルを構築した。Twitter を対象にした分析では、安部首相の育児休暇に関する発言の前後で、ツイート内容に関する構造が大きく変化していることをとらえた。また、商品間の購買グラフを利用したグラフ特徴量では、優良顧客や店舗別に見られる購買傾向の違いを構造から把握することができた。

グラフ構造によって要素間の関わりが複雑なシステムをモデル化することができ、構造を把握、解明するうえでグラフ構造の可能性は大きく、マーケティングやビジネス応用としても今後ますますグラフ構造を扱うことの重要性は高まってくるであろう。

参考文献

- [1] L. Getoor and C. P. Diehl, “Link mining: A survey,” *SIGKDD Explorations*, **7**(2), pp. 84–89, 2005.
- [2] D. J. Watts and S. H. Strogatz, “Collective dynamics of small-world networks,” *Nature*, **393**, pp. 440–442, 1998.
- [3] R. Albert and A.-L. Barabasi, “Statistical mechanics of complex networks,” *Reviews of Modern Physics*, **74**, pp. 47–97, 2002.
- [4] 前川浩基, 内田将史, 大内章子, 宇野毅明, 羽室行信, “データ研磨手法を用いた Twitter ユーザの関係構造変化の検出,” 人工知能学会全国大会論文集, **28**, pp. 1–4, 2014.
- [5] 中原孝信, 羽室行信, “グラフ特徴量を用いた識別モデルによる内在的購買行動の抽出,” 人工知能学会全国大会論文集, **29**, pp. 1–4, 2015.
- [6] H. Kauts, B. Selman and M. Shah, “The hidden web,” *AI Magazine*, **18**(2), pp. 27–36, 1997.
- [7] 宇野毅明, 中原孝信, 前川浩基, 羽室行信, “データ研磨によるクリーク列挙クラスタリング,” 情報処理学会アルゴリズム研究会報告書, 2014-AL-146(2), pp. 1–8, 2014.
- [8] M. Berlingerio, D. Koutra, T. Eliassi-Rad and C. Faloutsos, “NetSimile: A scalable approach to size-independent network similarity,” *CoRR*, abs/1209.2684, 2012.
- [9] M. Berlingerio, D. Koutra, T. Eliassi-Rad and C. Faloutsos, “Network similarity via multiple social theories,” In *Advances in Social Networks Analysis and Mining*, pp. 1439–1440, 2013.