コミュニティQAにおける質問・回答間の依存関係抽出の検討

香川　雄一† 　佐藤　哲司†

† 筑波大学大学院図書館情報メディア研究科　〒 305–8550 茨城県つくば市春日 1–2
E-mail: †{kagawa,satoh}@ce.slis.tsukuba.ac.jp

あらまし　コミュニティQA では，1 つの質問に対して直接的な回答だけでなく，既に投稿された回答への補足や反論など追記型の投稿がなされることが多い．そのため，質問・回答間には何らかの依存関係があると考えられる．我々はこれまでに，質問・回答間の依存関係を手動で抽出し，依存関係を基にベストアンサーの特徴分析を行った．本論文では，グッドアンサー情報を用いて特徴分析を行い，ベストアンサーとの比較を行う．また，依存関係を自動的に抽出するために，依存関係を構成する観点を表現上の特徴と情報量を用いて抽出する．その結果，ベストアンサーとグッドアンサーに選ばれた回答の依存関係に基づく特徴の違いが明らかになり，良質な回答の特徴を新たに発見できる可能性を示した．また，質問・回答の観点を表現上の特徴と情報量を用いて自動抽出を行った結果について報告する．

キーワード　コミュニティQA，Web マイニング，談話解析，ベストアンサー，グッドアンサー，データベース

Extraction of dependence relationships among Question and Answers in Community QA

Yuichi KAGAWA† and Tetsuji SATOH†

† Graduate School of Library and Information Science and Media Studies,University of Tsukuba
1–2,Kasuga,Tsukuba,Ibaraki,305–8550 Japan
E-mail: †{kagawa,satoh}@ce.slis.tsukuba.ac.jp

Abstract　In Community QA, answerers post not only direct answers to the question but also to the answers posted before in order to support or object to them. That is why it is supposed that there are dependences among question and answers. We express dependence relationships among question and answers with network structure manually and analyze network charts expressing dependences using best answer information. In this paper, we analyze them using good answer information and compare the result with that of best answer information. And in order to extract dependences automatically, we extract perspectives in each article by text feature words and entropy. In the result, we find the difference about features between best answer and good answer, and prove the possibility about finding new features of good quality answers by dependences among question and answers. And we report on the result of extracting perspectives in each article by text feature words and entropy.

Key words　Community QA，Web Mining，Conversation analysis，Best answer，Good answer，Database

1. はじめに

近年，電子掲示板，コミュニティQA などウェブ上の情報交換・知識共有サイトの利用が盛んである．コミュニティQA とは，電子掲示板上で利用者同士が知識や知恵を共有するナレッジコミュニティである [1]．例として，日本語の質問回答サイトの中で最大級の Yahoo! 知恵袋(注1)や教えて!goo(注2)などが挙げられる．コミュニティQA では，利用者は仮名 (ユーザ ID) で質問や回答を投稿でき, 年齢，性別，職業の異なる利用者から多様な質問・回答を得ることができる．

コミュニティQA では，1 つの質問に対する直接的な回答だけでなく，自らの体験や意見の提示, 既に投稿された回答への補足や反論など追記型の投稿がなされることが多い [2]．その際に,2 つ目以降の回答を投稿する回答者は，質問だけでなく自分より先に投稿された回答を見て，回答するかどうかの判断も含めて自分の回答の内容を決めると考えられる．例えば,「遠距離恋愛を続けるべきか」という質問がある時に，続けるべきと言

(注1)：http://chiebukuro.yahoo.co.jp/
(注2)：http://oshiete.goo.ne.jp/

う先行した回答に対して，自らの体験を補足する，あるいは，続けるべきとした回答の根拠が曖昧として，やめた方がいいとする回答が投稿されるなど，先の回答を参照しながら賛同や反論を示す回答がなされている．このように，ある質問とその質問に対する回答，更には回答間同士には，依存関係があると考えられる．

一方で，コミュニティQA に投稿された質問・回答は膨大な件数に達しており，利用者は直接質問を投稿せずとも，過去に投稿された質問から，自分の状況にマッチした解決策を探すことが可能となってきている．こうした状況では，検索エンジンに入力する検索クエリに単に適合する質問を探すだけではなく，良質な回答が含まれる質問記事を探す必要がある [3]．従来の手法では，各回答を独立したものとして評価し，良質な回答の特徴を導き出している．しかし，ある質問とその質問に対する回答，更には，回答間同士の依存関係を明らかにすることで，新たに良質な回答の特徴を導くことができると考えられる．

そこで本研究では，質問に連なる回答間の依存関係をネットワーク構造で表現し，質問・回答間の依存関係を定量的に評価するという新しいアプローチを試みる．我々はこれまでに，依存関係を表すネットワーク図作成の実験を行った [4]．また，依存関係を表すネットワーク図に基づいて新たな良質な回答の特徴を明らかにすることを目指し，ベストアンサーの特徴分析を行った．また，大規模データに対して依存関係を明らかにするため，質問記事および回答記事から依存関係を構成する観点を自動抽出するための手法を提案した [5]．本論文では，新たにグッドアンサー情報を用いて特徴分析を行う．また，観点を自動抽出するための手法として，新たに情報量を用いた手法について提案する．

本論文の構成は以下の通りである．2 章で関連研究について説明する．3 章では，質問・回答間の依存関係の概要と，依存関係を表すネットワーク図作成の実験の概要について説明する．4 章では，依存関係に基づいたグッドアンサーの特徴分析について説明する．5 章では，依存関係を構成する観点の表現上の特徴と情報量を用いた自動抽出について説明する．6 章で，4 章と 5 章でそれぞれ得られた分析結果について考察を行う．最後に 7 章で，本論文のまとめと今後の課題について述べる．

2. 関 連 研 究

これまでに，コミュニティQA 利用者が投稿する質問や回答の意図を明らかにする研究がなされている．Long Chen ら [6] は，質問者の質問の意図を明らかにすることで，コミュニティQA において類似した質問を同定し，関連する回答や潜在的な回答者を発見できると述べている．そのために，テキスト上の特徴やメタデータを特徴量とする機械学習を用いて，質問の意図ごとに 3 つのカテゴリに分けて質問を分類する予測モデルを提案している．Yoon ら [7] は，query を入力するユーザの検索意図を明らかにするため，コミュニティQA をコーパスとして用い，コミュニティQA 内の intent(意図) をチャンク間で引き継ぐ研究を行なっている．具体的には，意図の素性を明らかにし，素性スコアを算出する方法を用いている [8]．山本ら [9] は，ウェブ検索においてユーザがより興味を引く単語を提示するために「修飾語付き観点」に着目し，Q&A コンテンツの質問からウェブ検索に有用な修飾語付き観点を抽出手法を提案している．柳瀬ら [10] は，質問応答事例検索における質問事例から質問内容の中心表現を取り出すために，文末表現に着目した表層表現のパターンマッチングを利用している．佐々木ら [11] は，手掛かりとなる表現を用いて how 型 QA と why 型 QA からそれぞれ行動表現と理由の回答部分を抽出し，回答の根拠を提示する意思決定支援型の提案している．

また，良質回答の特徴分析に関する研究については，石川ら [3] は，2 人の判定者とベストアンサーの 3 者が一致して選択した回答を良質回答と見なし，良質回答の特徴分析を行なっている．特徴として【詳しさ】【丁寧さ】【根拠】を挙げ，これらに基づいて，良質な回答の推定モデルを提案・構築している．栗山ら [12] [13] は，質問を情報検索型と社会調査型の 2 つのタイプに分類し，ベストアンサーの特徴を記述形式や表現パターンの観点から分析している．また，ベストアンサーを推定するために頻繁にサイトを使用している利用者およびベストアンサーに数多く選ばれている回答者の質問・回答行動の特徴について分析している．Blooma ら [14] は，良質回答の特徴として，non-textual feature と textual feature について述べている．non-textual feature としては，質問者と回答者のリンク構造を解析し，質問者の権威 (Asker Authority) や回答者の権威 (Answer Authority) などを特徴として利用している．textual feature としては，回答の長さ，回答文中のユニークワード数，質問と回答の共通ワード数などを特徴として利用している．

従来研究では，ユーザの検索またはチャンク間の関係を基に質問や回答の意図を明らかにしている．本研究では，大規模データに対して依存関係を明らかにするため，各チャンク内において質問記事および回答記事から依存関係を構成する観点を抽出することを目指す．さらに良質な回答の特徴分析について従来研究では，各回答を独立したものとして評価し，主にテキスト情報を用いて良質な回答の特徴分析を行なっている．また，質問者や回答者などコミュニティQA 利用者の情報に基づいて，分析を行なっている．本研究では，各回答を独立したものとしてではなく，チャンク内における質問や回答の依存関係について着目して，良質な回答の特徴分析を行う．

3. 質問・回答間の依存関係

本章では，質問・回答間の依存関係の概要と，依存関係を表すネットワーク図作成の実験の概要について説明する．

我々は，質問記事および回答記事を構成する観点を単位として，依存関係を表すネットワーク図を作成する手法を提案している [4]．以下に，ネットワーク図を作成するために策定した生成規則を概説する。

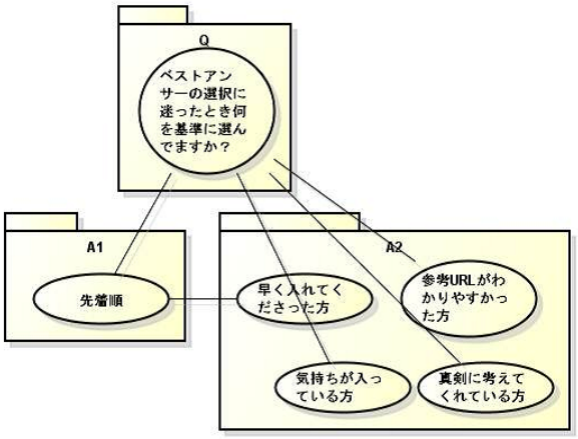


図 1　質問・回答間のネットワーク図の例

ネットワーク図の作成規則（初訂版）

(1)　質問文を読んで,「質問の聞きたいこと, 知りたいこと」を質問の観点として抜き出し，質問内のノードとする.
(2)　回答文を読んで,「各回答ごとに応えていること，応えようとしていること」を一つ一つ回答の観点として分解し，各回答内のノードする.
(3)　作成したノードにおいて，質問内のノードと対応して関連のある回答の観点のノードの間でリンクを張る.
(4)　また，読み取った回答の観点のノード間で対応しているもの同士で無向のリンクを張る.
(5)　賛同，反論，具体化（例示），抽象化（まとめ）から選択し，リンクにラベル付を行う.

以上の作業を行うことで，質問・回答間の関係を再構成し，ネットワーク図を作成する．例えば,「ベストアンサーの選択に迷ったとき何を基準に選んでますか？」という質問に対して，時系列順に「迷ったら先着順です。」,「早く入れてくださった方や 参考 URL がわかりやすかった方、 気持ちが入っていて、真剣に考えてくれている方、 など、ですかねぇ。」という回答があったとする．この場合，質問の聞きたいこと, 知りたいこととして,「ベストアンサーの選択に迷ったとき何を基準に選んでますか？」というノードが抽出される．質問に対する回答として前者の回答は「先着順」，後者の回答は「早く入れてくださった方」,「参考 URL がわかりやすかった方」,「気持ちが入っている方」,「真剣に考えてくれている方」というノードがそれぞれ抽出できる．これらの抽出したノードに基づいて同じ観点をリンクでつなぐとき，回答の「先着順」,「参考 URL がわかりやすかった方」,「気持ちが入っていて、真剣に考えてくれている方」というノードはそれぞれ質問「ベストアンサーの選択に迷ったとき何を基準に選んでますか？」に対して最初に出た観点なので，直接質問と回答のノード間でリンクを張る．ここで，回答の「早く入れてくださった方」というノードは,「回答の早さ」という観点で「先着順」というノードと同じであると考えられるので，両ノード間にリンクを張る．同様に，後から投稿された回答に, 既に投稿された回答と同じ観点がある場合は，直近の回答のノードとリンクを張る．以上の結果を，図 1 に示す.

上記の規則を用いて我々はこれまでに，依存関係を表すネットワーク図作成の実験を行った [4]．データは，大学共同利用機関法人国立情報研究所が提供をしている，Yahoo!知恵袋を用いた．実験では，NTCIR-8 CQA（コミュニティQA テストコレクション）(注3) に含まれる質問と回答を用いる．これは 2004 年 4 月から 2005 年 10 月に投稿された質問から無作為に 1500 件を抽出したもので，全 14 のカテゴリからなる．このうち特に回答数が多く，回答間の依存関係が多いと思われる「恋愛相談，人間関係の悩み」,「Yahoo!オークション」の 2 カテゴリから無作為に 81 件抽出して実験を行った.

4.　依存関係に基づいたグッドアンサーの特徴分析

本章では，質問・回答間の依存関係に基づくグッドアンサーの特徴分析について説明する．分析は,3 章で説明した実験によって得られたネットワーク図を用いて行った.

「グッドアンサー」情報とは，酒井ら [15] [16] によって行われた 4 名の判定者による回答評価実験の結果によって付与された情報である．質問者とは異なる 4 名の判定者は，各回答の質を A,B,C の 3 段階で絶対評価を行っている．判定者に与えた回答の判定基準は以下の通りである.

A　質問内容を十分満たす答えが含まれている

B　質問内容に部分的に適合している，もしくは部分的に不適合

C　質問内容と全く関係がない

本研究では，特に評価の高い，3 名以上が A と判定した回答を「グッドアンサー」と定義し，特徴分析を行う．ベストアンサーは各チャンクに 1 つしかないが，グッドアンサーはチャンクによって数が異なる．その内訳を図 2 に示す.

ネットワーク図から明らかとなった回答の特徴としては，以下の 4 つについて着目する [5].

a. 他の回答とのつながりが最も多い：リンク数が最も多い

b. 他の回答に比べ，最も多くの観点を提示している：ノード数が最も多い

c. 独自の観点を提示している：他の回答の観点とつながりを持たない観点がある

d. 上記 3 つの特徴をいずれも持たない

実験によって得られたネットワーク図を用いて，グッドアンサーが 4 つのどの特徴を持つかを検証をした．その結果をベン図にして図 3 に示す.「Yahoo!オークション」カテゴリでは，40 件中 9 件が最多で a と b の両方の特徴を持つことがわかった．また,「恋愛相談，人間関係の悩み」カテゴリでは，165 件中 41 件が最多で c の特徴を持つことがわかった.

さらに，グッドアンサーのうち，4 名の判定者が A と判定した最も高い評価が付与された回答が 4 つのどの特徴を持つか検

(注3)：http://research.nii.ac.jp/ntcir/permission/ntcir-8/perm-ja-CQ.html#cqa_task

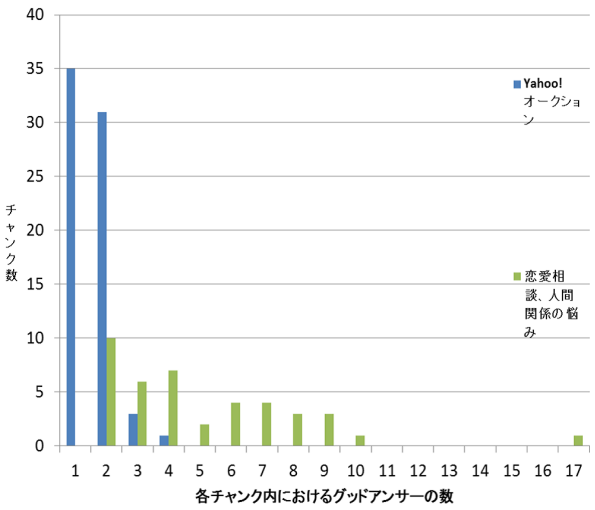


図 2　カテゴリごとのグッドアンサーの件数別チャンクの数

証を行った．その結果をベン図にして図 4 に示す.「Yahoo!オークション」カテゴリでは，18 件中 3 件が最多で a と b の両方の特徴を持つことがわかった．また,「恋愛相談，人間関係の悩み」カテゴリでは，41 件中 12 件が最多で a と b の両方の特徴を持つことがわかった．

最後に，ベストアンサーかつグッドアンサーに当てはまる回答が 4 つのどの特徴を持つか検証を行った．その結果をベン図にして図 5 に示す.「Yahoo!オークション」カテゴリでは，17 件中 5 件が最多で a と b の両方の特徴を持つことがわかった．また,「恋愛相談，人間関係の悩み」カテゴリでは，29 件中 11 件が最多で c の特徴を持つことがわかった．

5. 表現上の特徴と情報量を用いた観点の自動抽出

我々はこれまでに，大規模データに対して依存関係を明らかにするため，質問記事および回答記事から依存関係を構成する観点を自動抽出するための手法を提案した [5]．具体的には，3 章のネットワーク図作成実験で得られた観点を基に観点を出現させると考えられる特徴的な表現を拾い上げ，その周辺の語にスコアを付与する．その後，各語のスコアにおいて，閾値以上のものを抽出する．形態素単位で区切られた語 w のスコアを式（1）を用いて算出する．特徴的な表現の直前あるいは直後に隣接する語のスコアは θ（0 ≦ θ ≦ 1）とし，n 形態素単位で離れた語のスコアは θ の n 乗とする．本論文では，広い範囲にわたって特徴語によるスコアを付与するために，θ の値を 1 に近い 0.9 と固定して算出する．

$$Score(w) = \sum \theta^{n} \tag{1}$$

また，一般性の高い語を除外するために，idf 値を用いた観点候補の算出を試みた．具体的には，特徴的な表現によるスコアの総和に idf（逆文書頻度）値をかけて，各語のスコアとする．スコアは式（2）によって算出する．

$$\begin{aligned} Score(w) &= (\sum \theta^{n}) \times idf(w) \\ &= (\sum \theta^{n}) \times (log\frac{N}{df(w)} + 1) \end{aligned} \tag{2}$$

本論文では，情報量を用いた観点候補の算出を試みる．各チャンク内における各単語の情報量を求める．情報量は，各単語集合の中である単語が出現する確率，つまり，ある単語の出現しにくさを表す．これまでと同様，特徴的な表現によるスコアの総和に情報量の値をかけて，各語のスコアとする．情報量を用いたスコアは式（3）によって算出する．

$$\begin{aligned} Score(w) &= (\sum \theta^{n}) \times H(w) \\ &= (\sum \theta^{n}) \times P(X = w) logP(X = w) \end{aligned} \tag{3}$$

最終的に，チャンク内で最もスコアの高い語のスコアが 1 になるようにスコアをチャンクごとに正規化し，全ての語のスコアを 0 以上 1 以下とする．

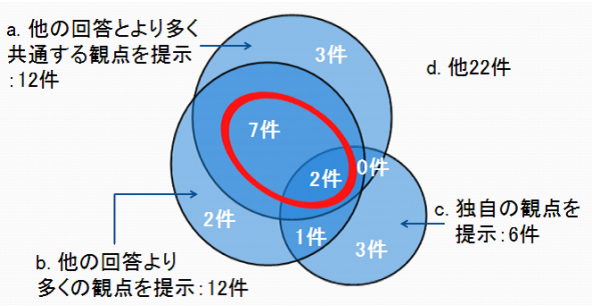
3 章の実験で用いたデータセットに対して，以上の方法でスコアを付与し，実験で得られたデータを観点の正解データとし，スコアに対して閾値を変えて観点抽出を行った結果について示す．閾値は 0 から 1 まで 0.1 ごとに変え，閾値よりスコアが大きい語を観点として抽出する．閾値が 0 のときは，スコアが付与された全ての語を抽出し，閾値が 1 のときは，どの語も抽出しないことを意味する．

idf 値を用いた場合と情報量を用いた場合との比較を図 6 に示す．図 6 よりいずれの場合も，Q（質問）に対して観点抽出を行うと，F 値は閾値が 0.9 のとき，最大で約 0.7 となった．また，BA（ベストアンサー），NA（ノーマルアンサー）に対して観点抽出を行うと，いずれの場合も，F 値は閾値が 0 のとき最大でそれぞれ約 0.4，約 0.3 となった．

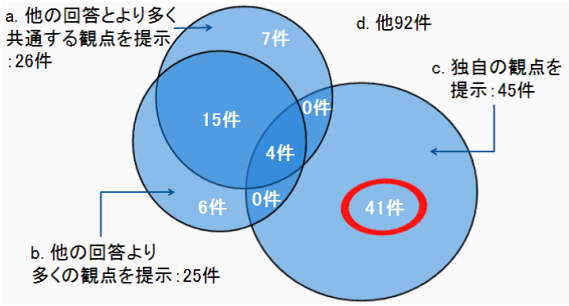
6. 考　　察

6.1 良質な回答の特徴分析に関する考察

「グッドアンサー」情報を用いた分析の結果，図 3 より,「Yahoo!オークション」カテゴリでは，a,b の特徴を持つグッドアンサーが相対的に多く,「恋愛相談，人間関係の悩み」カテゴリでは，c の特徴を持つグッドアンサーが多い．図 2 より,「Yahoo!オークション」カテゴリでは，ほぼ各チャンクに 1 つのグッドアンサーがあるが,「恋愛相談，人間関係の悩み」カテゴリでは，各チャンクに 1 つとは限らず，チャンクによってグッドアン

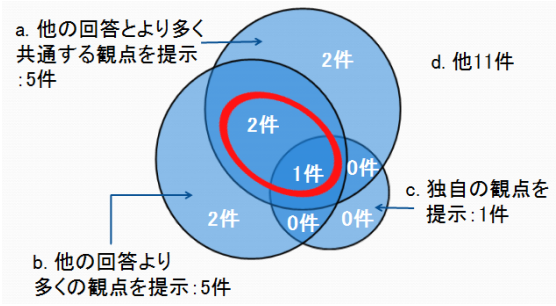


(a) Yahoo!オークション」カテゴリ

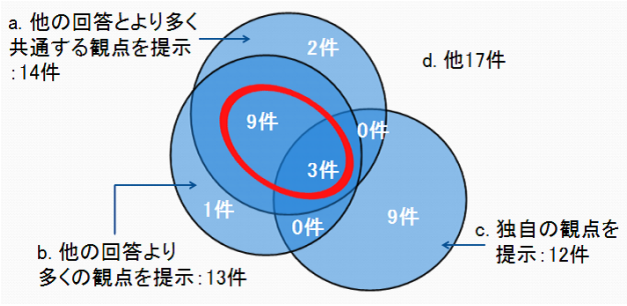


(b) 「恋愛相談，人間関係の悩み」カテゴリ

図 3　ネットワーク図上の観点間のつながりに基づくグッドアンサーの分布

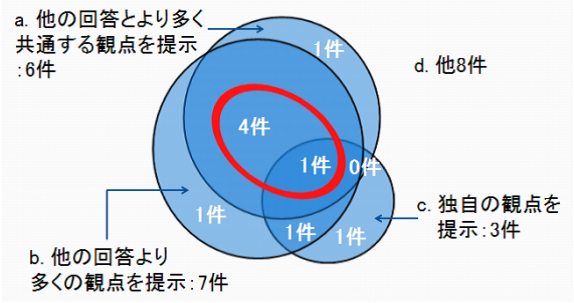


(a) Yahoo!オークション」カテゴリ

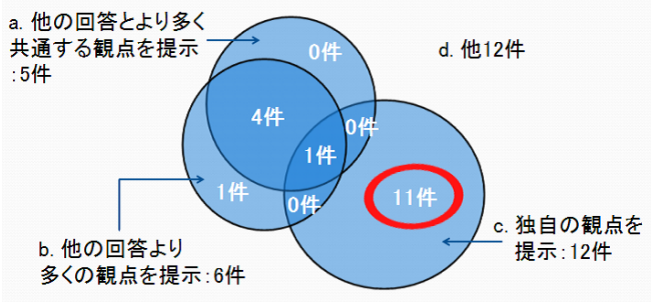


(b) 「恋愛相談，人間関係の悩み」カテゴリ

図 4　ネットワーク図上の観点間のつながりに基づく「AAAA」である回答の分布

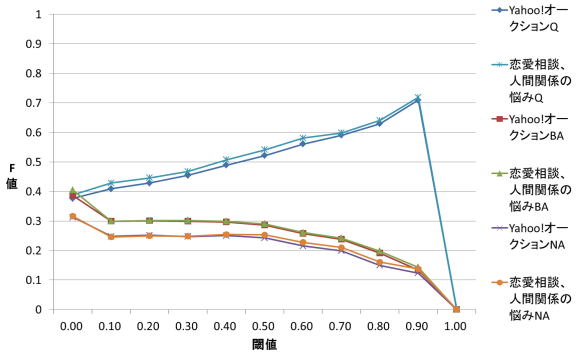


(a) Yahoo!オークション」カテゴリ

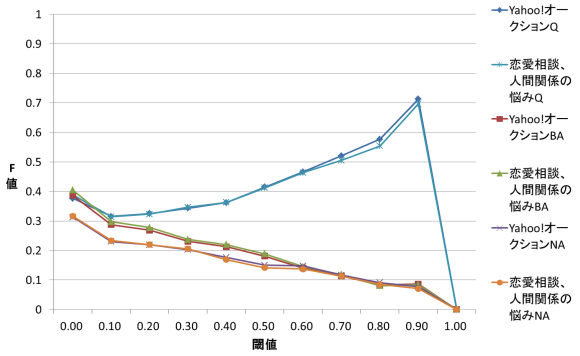


(b) 「恋愛相談，人間関係の悩み」カテゴリ

図 5　ネットワーク図上の観点間のつながりに基づくベストアンサーかつグッドアンサーの分布



(a) idf 値用いた場合



(b) 情報量を用いた場合

図 6　Q（質問）、BA（ベストアンサー）、NA（ノーマルアンサー）に対する観点抽出を行ったときのカテゴリ別 F 値

サーの数が大きく異なることがわかる．チャンク内に複数グッドアンサーがある場合でも，a，b の特徴を持つ回答は各チャンクに 1 つの回答にしか当てはまらない．そのため，相対的に c の特徴を持つグッドアンサーの割合が大きくなるにも関わらず，

表 1 「独自の観点を提示している」特徴を持つ回答の割合

「Yahoo!オークション」カテゴリ	
特徴分析に用いた回答	割合
ベストアンサー	32.5%
グッドアンサー	15.0%
評価「AAAA」の回答	5.56%
ベストアンサーかつグッドアンサー	17.6%

「「恋愛相談，人間関係の悩み」カテゴリ	
特徴分析に用いた回答	割合
ベストアンサー	39.0%
グッドアンサー	27.2%
評価「AAAA」の回答	29.3%
ベストアンサーかつグッドアンサー	41.4%

表 2 「3 つの特徴をいずれも持たない」回答の割合

「Yahoo!オークション」カテゴリ	
特徴分析に用いた回答	割合
ベストアンサー	45.0%
グッドアンサー	55.0%
評価「AAAA」の回答	61.1%
ベストアンサーかつグッドアンサー	47.1%

「「恋愛相談，人間関係の悩み」カテゴリ	
特徴分析に用いた回答	割合
ベストアンサー	43.9%
グッドアンサー	55.8%
評価「AAAA」の回答	41.5%
ベストアンサーかつグッドアンサー	41.4%

ベストアンサーの結果と比較するとどちらのカテゴリでも，表 1 より，c の特徴を持つ回答の割合は少ない．これより独自の観点を持った回答は，ベストアンサーに比べて比較的グッドアンサーになりにくい傾向にあることがわかる．この結果は，第 3 者視点という客観的評価をされたグッドアンサーと質問者視点という主観的評価をされたベストアンサーのちがいが反映されていると考えられる．また，評価が「AAAA」である回答の結果に着目すると，図 4 より，どちらのカテゴリも a,b の特徴を持つ回答が最も多い．特に「恋愛相談，人間関係の悩み」カテゴリでは，図 4. のグッドアンサー情報を用いた結果では，c の特徴を持つ回答が最も多かったが，評価が「AAAA」である回答を用いた結果では，a,b の特徴を持つ回答が多くなった．これより，客観的な評価で最も高い評価を得た回答は，カテゴリを問わず，独自の観点よりも，他の回答の観点を踏まえたり多くの観点を提示している回答が多いことがわかる．また，図 5 の「ベストアンサーかつグッドアンサー」の結果に着目する．表 2 より，どちらのカテゴリでも比較的 d の特徴を持つ回答が少ない．これより，ベストアンサーかつグッドアンサーである主観的評価でも客観的評価でも高い評価を得られた回答は，依存関係において何らかの特徴を持っている可能性が高いことがわかる．

6.2 観点の自動抽出に関する考察

図 6 より Q（質問）に対して観点抽出を行うとき,「Yahoo! オークション」カテゴリでは，情報量を用いて閾値を 0.9 にすると F 値は約 0.7 で最も高い値を示し，観点抽出により有効な手法であることがわかった.「恋愛相談，人間関係の悩み」カテゴリでは，idf 値を用いて閾値を 0.9 にすると F 値は約 0.7 で最も高い値を示し，観点抽出により有効な手法であることがわかった．いずれの場合も閾値を大きくすればするほど F 値は大きくなるため特徴的な表現に近い語ほど観点が含まれている可能性が高く特徴的な表現による手法の有効性を示せた．また，一般性の高い語の除外や単語の出現確率を考慮することは，Q（質問）に対する観点抽出において，より有効な手法であることがわかった．BA（ベストアンサー）では，いずれの場合も閾値を 0 にすると F 値は最も高い値を示した．閾値が 0 のときに最も高くなった理由としては，回答記事に特徴語が含まれていないためスコアが付与されない観点が多い，あるいは特徴語から遠くの位置に観点があるために観点のスコアが小さいため抽出されない，などが挙げられる．特徴語が含まれていない例としては,「ガイドラインに違反した物品」という回答などが挙げられる．また，特徴語から遠くの位置に観点があるために観点のスコアが小さい例としては,「単に違反で取り消されただけでしょうね．落札された商品の内容次第なのでなんとも申せませんが，出品者がヤフーに取り消され，ココに不可解なことを書いて憂さ晴らししている方が最近多いです.」など,「違反」が観点であるが，特徴語「です」から遠くの位置にあるためにスコアが小さくなっている回答などが挙げられる．NA（ノーマルアンサー）では，F 値は閾値が 0 のときに最大で約 0.3 となった．全体の傾向は BA と同じであるが，どの手法の場合も BA より値が下回っていることがわかった．以上より，本手法はカテゴリに依存せず，NA より BA に対して有効であったということがいえる．

7. おわりに

本論文ではまず，依存関係を表すネットワーク図に対して新たにグッドアンサー情報を用いて特徴分析を行った．その結果，2 カテゴリ間でベストアンサー，グッドアンサーに選ばれやすい回答の特徴に違いがあることがわかり，質問・回答間の依存関係を用いて，新たに良質な回答の特徴を発見できる可能性を示せた．また，大規模データに対して依存関係を明らかにするため，表現上の特徴と情報量を用いて観点の自動抽出を行った．その結果，質問の観点抽出に対する本手法の有効性を示せた．回答に対する観点抽出は，質問に対しての結果と比べ低い値となったが，ノーマルアンサーよりもベストアンサーでより有効であるということがわかった．

今後の課題としては，良質な回答の特徴分析においては，さらにカテゴリを増やして分析を行い，カテゴリごとや質問タイプごとなど様々な側面から評価を行うことなどが挙げられる．また，観点の自動抽出においては，回答の観点抽出精度を改良することが挙げられる．そのために表現上の特徴語やスコア付与の関数を改良することで特徴語によるスコア付与の範囲を広

げることや，質問の観点を用いた回答の観点抽出などが挙げられる．また，評価を行うカテゴリ数を増やすことでカテゴリ毎の違いを明らかにしていきたいと考えている．

謝　　辞

本研究の一部は筑波大学図書館情報メディア系プロジェクト研究による助成を受けたものである．また，実装・評価に際し，大学共同利用機関法人国立情報学研究所から提供を受けた，Yahoo!知恵袋のデータを利用している．ここに記して謝意を示す．

文　　献

[1] 三浦麻子, 川浦康至, 地福節子, 大瀧直子, 岡本真. 知識共有コミュニティを創り出す人たち. 人工知能学会全国大会大会論文集, pp. 1–4, 2006.

[2] 島田諭, 福原知宏, 佐藤哲司. 質問回答サイトにおける利用者行動に基づく記事の関連付け手法の検討. 電子情報通信学会 第 19 回データ工学ワークショップ（DEWS2008）, No. B6-5, 2008.

[3] 石川大介, 酒井哲也, 関洋平, 栗山和子, 神門典子. コミュニティ qa における良質回答の自動予測. 情報知識学会誌 21(3), pp. 362–382, 2011.

[4] 香川雄一, 島田諭, 神門典子, 佐藤哲司. コミュニティQA における質問・回答が互いに及ぼす影響の可視化に関する検討. 第 4 回データ工学と情報マネジメントに関するフォーラム (DEIM2012), No. C3-1, 2012.

[5] 香川雄一, 佐藤哲司. コミュニティQA における質問・回答間の依存関係の評価と観点の自動抽出の検討. 情報社会学会, 第 5 回知識共有コミュニティワークショップ論文集, pp. 61–64, 2012.

[6] Long Chen Birkbeck, Dell Zhang, and Levene Mark. Understanding user intent in community question answering. WWW 2012 − CQA'12 Workshop, pp. 823–828, 2012.

[7] Soungwoong Yoon, Adam Jatowt, and Katsumi Tanaka. Connecting qas to show intent underlying web query. 第 4 回データ工学と情報マネジメントに関するフォーラム（DEIM2012）, No. D2-4, 2012.

[8] Soungwoong Yoon, Adam Jatowt, and Katsumi Tanaka. Intent feature discovery using q&a corpus and web data. *Proceedings of the 4th International Conference on Ubiquitous Information Management and Communication (ICUIMC 2010)*, pp. 40:1–40:7, 2010.

[9] 山本岳洋, 中村聡史, 田中克己. Q&a コンテンツからの観点抽出に基づくウェブ検索支援. 情報処理学会論文誌（トランザクション）データベース, Vol.4No.2（TOD50）, pp. 74–87, 2011.

[10] 柳瀬隆史, 木田敦子, 難波功, 落谷亮. 質問応答事例検索のための談話構造の認定. 情報処理学会研究報告. 情報学基礎研究会報告 2001(86), pp. 97–104, 2001.

[11] 佐々木智, 藤井敦. 回答の根拠を提示する意思決定支援型の質問応答システム. 言語処理学会第 17 回年次大会発表論文集, pp. 252–255, 2011.

[12] 栗山和子, 神門典子. Q&a サイトにおける質問と回答の分析. 情報処理学会研究報告. 情報学基礎研究会報告,Vol.2009-FI-95, pp. 1–8, 2009.

[13] 栗山和子, 神門典子. Q&a サイトにおける質問と回答の分析 (3)-質問・回答履歴を用いたベストアンサー推定-. 情報処理学会研究報告. 自然言語処理研究会報告 2010-NL-195(7), pp. 1–8, 2010.

[14] M.J. Blooma, A.Y.K. Chua, and D.H.L. Goh. Selection of the best answer in cqa services. *2010 Seventh International Conference on Information Technology*, pp. 534–539, 2010.

[15] 酒井哲也, 石川大介, 栗山和子, 関洋平, 神門典子. コミュニティ qa における良質な回答の選定タスク: 評価方法に関する考察. 第 9 回情報科学技術フォーラム (FIT2010), pp. 13–20, 2010.

[16] Tetsuya Sakai, Daisuke Ishikawa, and Noriko Kando. Overview of the ntcir-8 community qa pilot task (part 2): System evaluation. *Proceedings of NTCIR-8 Workshop Meeting*, pp. 433–457, 2010.