DBSJ Journal, Vol.11, No.1
June 2012

コミュニティ QA を用いたクエリ拡張のためのコンテキスト抽出に関する一考察

A Study on the Seasonally Context Extraction Method for Query Expansion using Community Question Answering Resources

大塚淳史♥　関洋平♦
神門典子♠　佐藤哲司♦

Atsushi OTSUKA　Yohei SEKI
Noriko KANDO　Tetsuji SATOH

コミュニティQAにおける質問記事には，質問者が知りたいとする情報要求だけでなく，投稿時に選択したカテゴリや投稿日時など，質問のコンテキストが含まれている．本論文では，質問記事に含まれるコンテキストを用いて，ウェブ検索におけるクエリ拡張を実現する手法を提案する．特に，コミュニティQAからの投稿日時と関連の高い季節性コンテキストを抽出する．質問記事を投稿時期に応じた4季節に分割し，変動係数と投稿のバースト周期を用いて，季節性を持つ単語を抽出できることを明らかにした．更に，それらの単語を用いることで，多様なWeb検索を実現するクエリ拡張が実現できることを報告する．

Question articles in CQA describe user's information needs, and they contain contexts like category or submitting-date. Thus, CQA is one of the useful resources for query expansion. Contexts are helpful for diversity of question. In this paper, we create web search queries by using question articles. Submitting-date of the question, i.e. season, and its category are applied for enhance the diversity of search queries. We evaluated CQA resources in the view of seasonality using Coefficient of Variation (C.V) and burst term of submitting. We demonstrate availability of season for diversified query expansion.

1. はじめに

Web情報の大規模化に伴い，所望の情報を入手するWeb検索の重要性はますます高まってきている． ユーザは，自らが調べたいことである情報要求を“疑問”として想起し，その疑問から検索クエリを作成しWeb検索エンジンに入力する．一つのキーワードやトピックに対して多様な情報を入手する広域なWeb探索を行う場合には，ユーザは，調べたいキーワードに対する様々な関連語を自身で想起し，絞り込み検索を行う必要がある．しかし，ユーザ個人が考えうる“疑問”の範囲には限りがあり，絞り込み検索も，ユーザが想定できる範囲でしか実行することができない．また，単純なキーワードでのWeb検索では，キーワードに関する一般的な内容のWebページが上位に検索されてしまうため，多様な情報を入手することは難しい．

著者らは，Web検索ユーザが想定できないようなWeb検索での“疑問”を，他ユーザの疑問によって補完することで，多様なWeb探索を実現する手法について検討を行なっている．そのひとつとして，Webサービスの一つである，コミュニティQA（CQA）の質問記事を用いたWebクエリ拡張手法について提案している[1]．CQAに投稿される質問記事はWebユーザが持つ“疑問”そのものであり，質問記事からWeb検索用のクエリを作成することで，疑問ベースの多様なWeb探索を実現することができる．また，クエリ拡張で使用した質問記事を自然言語のままユーザに提示することは，クエリの背後に隠された拡張の根拠を示す点で有用である．CQAに投稿される質問記事は膨大な数に及び，同じキーワードが使用されている質問記事でも，質問内容は異なっている場合が多い．本論文では，このような，多種多様な質問記事の中からWeb検索ユーザが興味をもつ“疑問”となる質問記事を提示するために，“カテゴリ”と“季節”をコンテキストとしたクエリ拡張のナビゲーション手法について提案する．特に，CQAにおける投稿時期により質問記事の“季節性”を明らかにするため，質問記事中の単語出現確率の分散とバースト抽出による手法を組み合わせた，CQAでの季節性を持つ単語の抽出手法について説明する．

本論文の構成は以下のとおりである．まず，2章で関連する研究について述べる．3章では，実装したクエリ拡張インターフェースについて説明し，4章でCQAからの季節性抽出手法について詳述し，5章で提案手法の評価を行う．6章で考察し，7章でまとめと今後の課題について述べる．

2. 関連研究

2.1 コンテキストアウェアなクエリ拡張

著者らが提案しているクエリ拡張は，ユーザは自らの状況（コンテキスト）に合わせて推薦されるクエリを変更できるという点に特徴がある．このような状況に合わせたクエリ推薦は盛んに研究されている．Cao ら[2]は，Web 検索でのセッションを元にしたクエリ拡張を提案している．Cao らは Web 検索エンジンのセッションデータと検索結果のクリックログを用いて，検索回数に応じて最適なクエリを推薦する手法を提案している．Semgstock ら[3]は，クエリログデータから，検索時間とドメインを自由に変更することで，その状況に応じた最適なクエリを推薦するシステムを実装している．

従来のコンテキストアウェアなクエリ拡張ではクエリログを始めとした Web 検索エンジンから入手したデータを用いることが一般的であった．本論文では，クエリ拡張のデータとして，外部リソースである CQA の質問記事を用いている点に大きな特徴がある．また，コンテキストはシステムが自動で推定するのではなく，複数のコンテキスト候補を提示し，ユーザ自身に自らの状況に合ったコンテキストを選択するという点でも，従来研究と異なる．

♥ 学生会員　筑波大学大学院図書館情報メディア研究科 博士前期過程 aotsuka@slis.tsukuba.ac.jp
♦ 正会員　筑波大学図書館情報メディア系 yohei@slis.tsukuba.ac.jp satoh@ce.slis.tsukuba.ac.jp
♠ 正会員　国立情報学研究所 kando@nii.ac.jp

2.2 クエリ拡張の意味提示

本研究では，コミュニティ QA の質問記事から拡張クエリを作成すると同時に，質問記事本文もユーザに提示する．質問記事は自然言語で記述されており，ユーザは質問記事を読むことで，拡張クエリの背後に隠された“疑問”を明確に知ることができる．クエリの背後に隠された意味や根拠を提示する手法に関しては，Guo ら[4]は，クエリ拡張にソーシャルアノテーションを付与することで，クエリを意味ごとに分類し，提示する手法を提案している．クエリの意味提示に自然言語処理のアプローチを適用した研究には Reisinger ら[5]の研究がある．Reisinger らは Web ページから is-a 関係を抽出することでラベルを作成し，確率文脈自由文法によりクエリとラベルを紐付けすることで，クエリに意味を説明するラベルを付与している．Zha ら[6]は，拡張クエリを画像と共に提示することで視覚的にクエリの意味を提示する手法を提案している．

本研究の手法は，クエリの拡張に用いた質問記事をそのまま提示するという特徴がある．従来の先行研究では，意味を推定するためにクエリ拡張との意味提示をそれぞれ別の外部リソースを使用していたが，本研究では CQA を質問記事のみでクエリ拡張と意味提示を同時に行える利点がある．

3. コンテキスト切り替えによるクエリ拡張インターフェース

コンテキスト切り替えによるクエリ拡張を実現するため，2 次元タブとタグクラウドを組み合わせたインターフェースを提案する．インターフェースを図 1 に示す．インターフェースは初期クエリの入力窓，カテゴリ・季節の 2 次元タブ，そして関連語のタグクラウドから構成される．検索したいキーワードを検索窓に入力すると，システムはキーワードに関連するカテゴリをカテゴリタブに．春から冬までの季節を季節タブに表示する．カテゴリタブは初期クエリと関連順に上からランキングされている．初期状態では，最も関連度が高いカテゴリ（最上位に表示されているカテゴリ）が選択されている．季節タブの初期状態はアクセスタイムをもとに，現在の時期に最適な季節が自動で選択される．

ユーザはカテゴリと季節を，コンテキストとしてタブを用い自由に切り替えることができる．タグクラウドに表示される関連語はタブを切り替えるごとに変化する．タグクラウドに表示される関連語のフォントサイズは，関連語が含まれる CQA の質問記事数に依存しており，より多くの質問記事で出現する関連語は大きなフォントで表示されている．関連語の表示順は LDA によって算出した，入力クエリとの潜在的な関連度順である．

本章では，コンテキストとなるカテゴリと季節のデータセットの作成法について説明する．

3.1 カテゴリによるコンテキスト切り替え

本研究では，CQA のコーパスとして国立情報学研究所が提供する Yahoo!知恵袋[1]コーパス[2]を使用している．Yahoo!知恵袋では質問記事を投稿する際に，質問記事の内容にふさわしいカテゴリを選択する必要があり，質問記事はカテゴリごとに分類されている状態になっている．そのため，カテゴリ別にクエリ拡張を行うことで，話題の異なる拡張クエリを作成

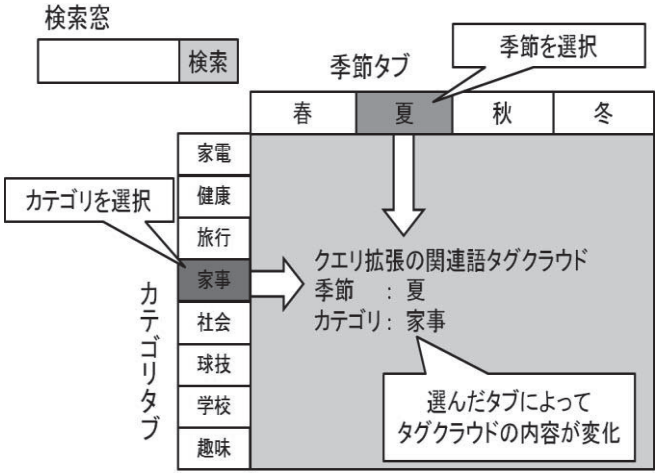

図 1 コンテキスト切り替えによる
クエリ拡張インターフェース

することができる．しかしながら，CQA は非常に多くのカテゴリを有しており，Yahoo!知恵袋にも 100 を超えるカテゴリが存在している．また，カテゴリは非常に細かく分類されているため，内容が非常に近いカテゴリも存在する．本研究では，内容の近いカテゴリ同士を統合することで，カテゴリ数を圧縮している．カテゴリ統合の指標として，Yahoo!知恵袋の階層構造を用いる．Yahoo!知恵袋のカテゴリは 3 階層の階層構造になっている．最上位の大カテゴリは実際には質問記事は投稿できず，抽象的なカテゴリとなっている．ここでは，この大カテゴリを基準としたカテゴリ統合を行う．大カテゴリの下位カテゴリである中カテゴリに投稿されている質問記事を集約し．最終的に 20 のカテゴリに統合する．統合した 20 カテゴリをナビゲーションカテゴリと呼ぶことにする．

3.2 季節によるコンテキスト切り替え

CQA は誰でも，自由な時間に投稿できるという利点がある．そのため，投稿される質問記事の内容は，投稿時期に大きく影響を受けると考えられる．また，内容が類似する質問記事は，1 年などの周期性を持って投稿されている．そこで，投稿時期によって質問記事を分割することで，より多様な拡張クエリを提示する手法を提案する．データの周期を 1 年とし，季節ごとにデータを分割する．Yahoo!知恵袋コーパスは，2004 年 4 月から 2009 年 4 月までのデータが収録されている．Yahoo!知恵袋が正式サービスとしてスタートした 2005 年 11 月以降のデータである，2006 年 1 月から 2008 年 12 月までの 3 年分のデータを使用することとした．3 年分のデータを春から冬までの四季に分割する際に，以下のように 3 月毎に質問記事を分割した．

- **春**：3 月～5 月に投稿された質問記事
- **夏**：6 月～8 月に投稿された質問記事
- **秋**：9 月～11 月に投稿された質問記事
- **冬**：12 月～2 月に投稿された質問記事

以上の 4 季節と 20 カテゴリを組み合わせ，80 通りのデータを作成する．表 1 に作成したナビゲーションカテゴリと，ナビゲーションカテゴリに使用した中カテゴリ，そして季節

[1] http://chiebukuro.yahoo.co.jp/
[2] http://www.nii.ac.jp/cscenter/idr/yahoo/chiebkr2/Y_chiebukuro.html

表 1 ナビゲーションカテゴリと質問記事数

ナビゲーションカテゴリ	中カテゴリ	春	夏	秋	冬
趣味	アニメ, コミック, 本, おもちゃ, 占い, くじ	46,281	55,793	67,941	89,893
エンタメ	映画, 音楽, 芸能人, ミュージカル, テレビ/ラジオ, 伝統芸能	77,018	85,588	94,995	106,841
デジタル・家電	パソコン, デジタルカメラ, インターネット, ソフトウェア, AV 機器, 携帯電話	92,981	91,920	96,057	110,821
旅行	国内旅行, 海外旅行, 交通案内, 路線案内, テーマパーク	85,900	96,284	99,627	107,634
恋愛	恋愛相談, 人生相談	149,830	169,239	196,009	220,074

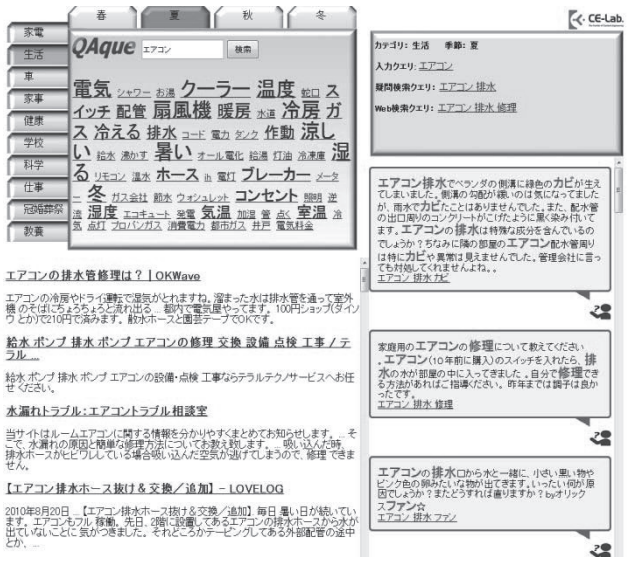

図 2 コンテキスト切り替えによる Web 検索システム

ごとの質問記事数を示す．趣味，エンタメカテゴリは元々一つの大カテゴリであるが，中カテゴリ数，質問記事数共に他のカテゴリよりも多いため，趣味カテゴリとエンタメカテゴリに分割した．また，中カテゴリである恋愛相談カテゴリは，質問記事投稿数が全カテゴリの中で最も多いが，人生相談カテゴリは非常に少ない投稿件数[3]であるため，恋愛カテゴリは他のナビゲーションカテゴリよりも質問記事数が多い．

3.3 コンテキスト切り替えによる Web 検索システム

提案したクエリ拡張の応用として，著者らは，Web 検索システムを実装している[7]．作成したシステムを図 2 に示す．タグクラウドから関連語を選択すると，初期クエリと関連語を含む質問記事が検索される．同時に質問記事からキーワードを抽出し，Web 検索用のクエリを作成する．ユーザには質問記事本文とクエリをセットで提示する．質問記事と拡張クエリのセットを CQA クエリと呼ぶ．初期クエリからタグクラウド，CQA クエリを順に選択していくことで，徐々に具体的な検索クエリが作成される．ユーザがタグクラウド中の関連語や CQA クエリを選択した時点で，Web 検索エンジン API により，Web 検索結果を得る．システムインターフェースは HTML5，Ajax により実装しており，タブやクエリの切り替えを非同期通信で実現しているため，インタラクティブに Web 検索を行うことができる．

[3] 2008 年 1 年間での投稿数 145 件

4. コミュニティ QA の季節性コンテキスト抽出

本研究では，カテゴリと季節性をコンテキストとして利用している．著者らは，実験により，カテゴリ数を増やすことで，より多様な Web 検索結果を得られることを実証している[1]．本論文では，CQA の季節性に着目する．CQA に投稿される質問記事は季節によって内容が変化すると仮定している．そこで，質問記事の季節的傾向を検証する．特に，単語ベースでの季節性を検証する．拡張クエリを作成する際，拡張のために追加する関連語として，質問記事中に出現する単語を使用している．そのため，投稿時期の単語の出現頻度の分布から，ある特定の季節に多く使用されている単語が存在するのかどうかを検証する．

本章では，季節性を持つ単語の抽出手法について詳述する．季節性を持つ単語の抽出手法として，変動係数を用いる手法と単語の出現確率のバーストの周期性による判定手法を提案する．

4.1 分散による季節性抽出

CQA における季節性を検証するため，月毎の単語の出現確率を計算する．出現確率は各カテゴリにおいて，ひと月に投稿された質問記事のうち，ある単語が含まれる質問記事の割合となる．カテゴリCにおいて，ある単語wの出現確率$P_{C,w}$は以下のように計算できる．

$$P_{C,w} = \frac{N_{C,w}}{N_C}$$

N_CはカテゴリCに投稿された全質問記事数，$N_{C,w}$はカテゴリCに投稿された質問記事のうち，単語wを含む質問記事数である．出現確率は，2006 年 1 月から 2008 年 12 月までの 36 ヶ月分計算する．

季節性を持つ単語は，ある月では高い出現確率を示すが，他の月では，出現確率が低い単語であると考えられる．そこで，各単語の月毎の出現確率の分散を計算し，分散の大きい単語を，季節性を持つ単語として抽出する．しかし，出現確率は非常に小さい値をとるため，単純な分散では，単語間の比較を行うことができない．そこで，分散を出現確率の平均値で正規化する変動係数を用いる．元々の出現確率が高い頻出語と稀にしか出現しない単語の分散を変動係数によって比較することで，ある特定の月に頻出する（バーストする）単語を抽出する．変動係数$C.V$は以下の式で計算する．

$$C.V = \frac{\sqrt{\sigma^2}}{\bar{x}}$$

$\sqrt{\sigma^2}$は標準偏差，$\bar{x}$は出現確率の平均である．変動係数を各単語で計算し降順に並べたとき，上位の単語を，季節性を持つ単語として抽出する．

4.2 バーストによる季節性抽出

単語の出現確率が，他の月と比較し急激に上昇する月が存在する．本論文では，これをバーストと呼ぶこととする．バーストは CQA に限らず，多くの情報源で発生する．Vlachos ら[8]は，クエリログから，Web 検索クエリのバースト性を検証している．また，山家ら[9]は，Vlachos らの手法を用いてソーシャルブックマークのバーストと周期性を発見している．CQA においても単語のバーストが発生すると仮定し，バーストが周期的に発見できる単語は季節性がある単語として抽出する． Vlachos らの手法を用いて，バーストを特定し，バーストが毎年同じ月に発生しているものだけを抽出する．Vlachos らのバースト発見法の手順を以下に示す．

(1) データ系列$t = (t_1, \ldots, t_n)$に対して，長さwの移動平均MA_wを計算

(2) MA_wの平均$mean(MA_w)$と標準偏差$std(MA_w)$より
$$cut\_off = mean(MA_w) + x * std(MA_w)$$
を計算

(3) i番目のデータに対して，
$$MA_w(i) > cut\_off$$
の場合，i番目のデータはバーストしたデータである

データ系列tは 2006 年 1 月から 2008 年 12 月までの 36 ヶ月分のデータである．移動平均の長さwは季節の区切りと対応させ，$w = 3$とした．またxは，閾値のための重み付けパラメータであり，予備実験の結果，$x = 2.5$とした．各単語において 2006 年から 2008 年までで，毎年同じ月にバーストしている単語のみを抽出し，それ以外の単語は除外する．

5. 評価実験

本章では，4 章で提案した季節性抽出手法に関する評価実験について述べる．まず，各手法によって抽出できた単語と，抽出単語の月毎の出現確率の推移を示し，手法の妥当性を検証する．次に，季節性が拡張クエリのコンテキストとして有用であるかを，抽出した季節性を持つ単語によって実証する．

5.1 コミュニティ QA からの季節性を持つ単語の抽出

4 章で提案した手法を Yahoo!知恵袋コーパス上で実験する．3.1 節で提案したナビゲーションカテゴリのなかから，デジタル・家電カテゴリでの実験結果を示す．デジタル・家電カテゴリはどの時期においても安定して多数の質問記事が投稿されている点，一見して季節性を認識しにくい点において，CQA に隠されている季節性を抽出する実験に適しているといえる．比較対象として，3 つの手法によって単語抽出を行う．

(a) **Freq** ： 出現確率の 36 ヶ月の平均値
(b) **C.V** ： 36 ヶ月分の出現確率の変動係数
(c) **C.V + Burst**： 変動係数結果にバースト周期性を考慮

実験結果を表 2～表 4 に示す．表 2 が出現確率の平均が上位の単語，表 3 が変動係数上位の単語，そして表 4 が変動係数上位かつ毎年同じ月にバーストしている単語である．また，表 2 には出現確率の平均値，表 3 には変動係数，そして表 4 には変動係数に加えて，バースト判定手法によってバーストと判定された月を示す．(a)では“教える”，“使う”といった

表 2 出現確率による抽出単語(Freq)

順位	単語	出現確率(平均)
1	教える	0.233
2	できる	0.211
3	使う	0.146
4	パソコン	0.138

表 3 変動係数による抽出単語(C.V)

順位	単語	変動係数
1	年賀状	1.93
2	ボーダフォン	1.26
3	湿る	1.20
4	流出	1.17

表 4 変動係数とバーストによる抽出単語(C.V+Burst)

順位	単語	変動係数	バースト
1	年賀状	1.93	11 月,12 月
2	湿る	1.20	6 月,7 月
3	除	1.13	6 月,7 月,8 月
4	4 月	0.98	3 月,4 月

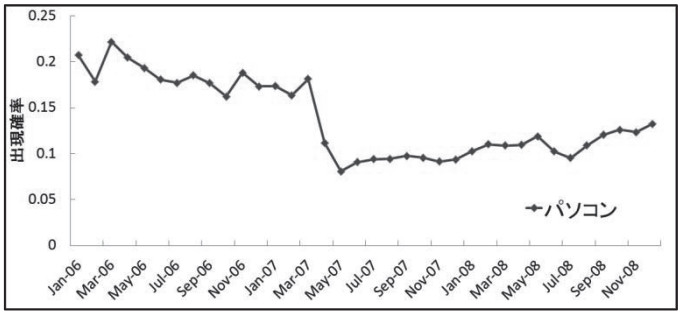

図 3 “パソコン“の出現確率の推移

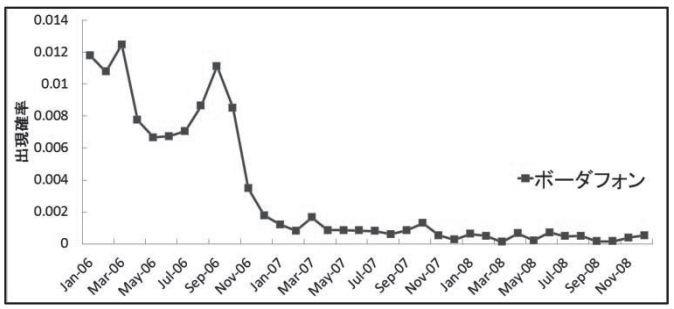

図 4 “ボーダフォン “の出現確率の推移

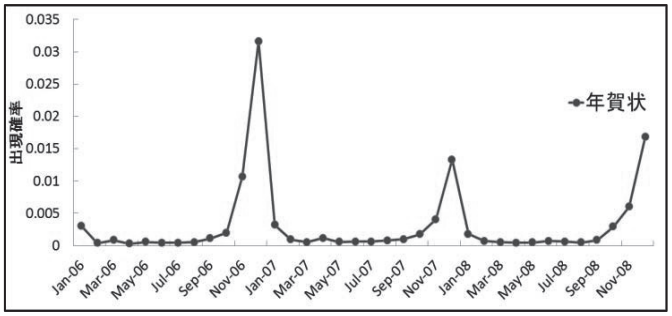

図 5 “年賀状“の出現確率の推移

CQA でよく使用される語が上位に来ているが，デジタル・家電カテゴリの特徴的な語として“パソコン”が抽出されている．また(b)では，“ボーダフォン”，“流出”が上位で抽出されているが，(c)では，それらの語が出現せず，“冷房”，“除”といった単語が抽出されていることがわかる．

次に，それぞれの抽出手法で抽出できた単語の月毎の出現確率の推移を図 3～図 5 に示す．図 3 は“パソコン”，図 4 は“ボーダフォン”，図 5 は“年賀状”の推移を示している．“パソコン”は月毎であまり大きな変化は見られない．“ボーダフォン”は 2007 年以降一気に出現確率が低くなり，以後はほぼ出現しない．“年賀状”は毎年，11 月，12 月に高い出現確率を示しているが，その他の月では，出現確率がほぼゼロに近いことがわかる．

5.2 季節性を持つ単語のクエリ拡張

5.1 節で，CQA から季節性を持つ単語を抽出した．ここでは，実際に季節性を持つ単語を入力した際の，クエリ拡張結果について示す．本研究での拡張クエリはタグクラウドの形式で提示される．タグクラウドは，関連する質問記事が多いほど，多くの関連語が推薦されるという特徴がある．また，本手法においての季節性とは，季節タブを変更した際，他の季節には存在しなかったユニークな関連語が提示されることである．そこで，季節ごとのユニークな関連語の数（ユニーク語数）を比較することでシステムの季節性を検証する．

初期クエリは，5.1 節の(c)手法で抽出した季節性を持つ単語として，“年賀状”，“チョコ”，“桜”の 3 つを用いた．“チョコ”は恋愛カテゴリ，“桜”は旅行カテゴリから抽出した単語である．図 6 にシステムの単語出力数を示す．横軸は季節，縦軸はユニーク語数である．また，棒グラフは，タグクラウドが表示した総関連語数である．“年賀状”は冬，“桜”は“春”，“チョコ”は冬から春にかけてユニーク語数が他の季節よりも増加している．

6. 考察

6.1 コミュニティ QA からの季節性抽出に関する考察

5 章では，CQA からの季節性を持つ単語を抽出するために 3 つの手法を比較した．出現確率の平均である(a) **Freq** では，CQA で用いられる一般的な単語が抽出されている．またデジタル・家電カテゴリで多く使用されている代表語として“パソコン”が抽出されている．図 3 では“パソコン”の出現確率は月ごとの変動がほとんど無いことがわかる．このことから，頻出語は投稿時期にかかわらず，ほぼ一定の高い割合で使用されているといえる．変動係数による順位付け(b) **C.V** では，“年賀状”などの単語に加えて“ボーダフォン”，“流出”といった単語が上位になっている．図 4 のグラフでは，“ボーダフォン”は 2006 年 9 月から 11 月にかけて出現確率が増加しているが，2007 年以降はほぼ出現していない．これは，携帯電話のキャリアであったボーダフォンがソフトバンクモバイルに切り替わる時期に一致している．また“流出”は，2006 年から 2007 年にかけてのファイル共有ソフトウェアによる情報流出に関連するものであると考えられる．これらは，突発的な事象であるため，バーストは一回発生するだけで，周期的に発生するものではない．バーストの周期性を考慮した(c) **C.V＋Burst** では，“ボーダフォン”，“流出”という単語がなくなり，“除”，“4 月”という単語が抽出されている．“4 月”は月を示す単語そのものである．“除”はバースト時期が 6 月から 8 月であるため，除湿に関する単語であると考えられる．“年賀状”の出現確率の推移を示した図 5 においても，毎年同じ時期にバーストが発生していることがわかり，同時に他の時期はほとんど使用されない単語である．これは，4.1 節で仮定した季節性の特徴に一致していることから“年賀状”は季節性を持つ単語であるといえる．また“湿る”，“除”，“4 月”の出現確率も同様の傾向を示し，主観的にも季節に関連すると思われることから，変動係数とバースト周期を考慮することで季節性を持つ単語を抽出できたと考えられる．

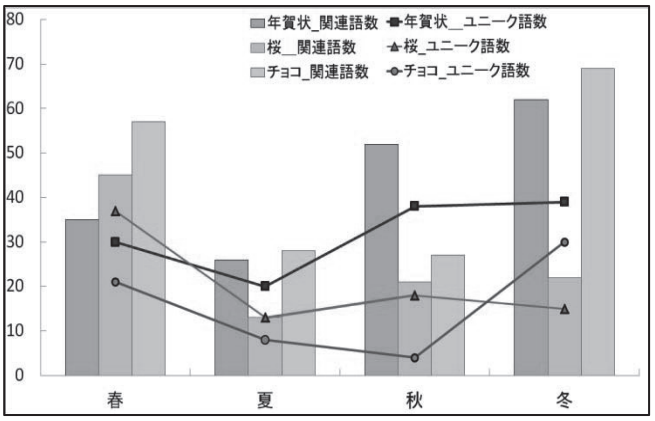

図 6 タグクラウドの関連語数とユニーク語数

季節性抽出手法によって抽出された“年賀状”は，出現確率の推移に特徴を持っている．また同様の手法を恋愛カテゴリ，旅行カテゴリに適用した結果，“チョコ”と“桜”というキーワードを抽出することができた．“チョコ”は 1 月と 2 月にバーストしており，バレンタインデーに関連するものであると考えられる．また“桜”は 3 月と 4 月にバーストしており，桜の開花と一致している．以上のことから，CQA では，どのカテゴリにおいても，投稿時期や季節が投稿される質問記事に影響を与えていると考えられる．

6.2 クエリ拡張における季節の有用性

図 6 の結果より，提案手法のクエリ拡張は単語の季節に，ユニーク語数が最大になることがわかる．これは，該当する季節には，投稿される質問記事数が多く，内容も多岐にわたるため，ユニークな関連語が多く提示されたのではないかと考えられる．また，他の季節においても関連語が提示されるが，質問記事数が少ないため，ユニーク語が少なくなったのではないかと考えられる．以上のことから，単語と一致する季節に対しては，クエリ拡張に季節性を反映されているといえる．

その他の季節では，単語によってグラフの形状が異なっている．“年賀状”では，春から秋にかけてユニーク語数が増加していっていることがわかる．図 5 の出現確率のグラフを見てみると，“年賀状”は 10 月から出現確率が増加し，11 月，12 月にピークを迎え，1 月には急激に低下している．これは，年賀状の作成時期に対応していると考えられる．デジタル・家電カテゴリにおいては，PC などを用いた年賀状の作成が主なトピックとして多くの質問記事が投稿される．そのため，実際に年賀状が届く 1 月には出現確率は急激に低くなっていると考えられる．“チョコ”に関しては 3 月にホワイトデーがあるため，春にもユニーク語が多く存在していると考えられる．それに対して“桜”は，春は多くのユニーク語が提示されるが，その他の季節は急激に少なくなっている．

これは桜の時期が 3 月から 4 月下旬までで，提案手法で定義した春の時期に完全に一致しているためであると考えられる．また，桜は開花している時期以外は観光ではあまり話題にあがることがないため，他の季節にはほとんどユニーク語が現れなかったのではないかと考えられる．以上より，単語に該当する季節以外にも単語によって様々な影響を投稿時期から受けているため，単に該当する季節のクエリを提示するのではなく，季節をコンテキストとして自由に切り替えることでより多様なクエリを提示できると考えられる．

7. まとめ

本論文では，コミュニティ QA（CQA）からのクエリ拡張のためのコンテキスト抽出手法を提案した．CQA には季節性が存在し，ある時期には多くの質問記事で使用されている単語が他の時期では全く使用されないといった季節性を持つ単語を抽出することで CQA の季節性を示した．また，CQA のカテゴリと季節をコンテキストとして自由に切り替えることによる，多様で特徴的なクエリ拡張手法を示した．

今後の課題として，より季節性の高い関連語をクエリ拡張で提示する手法を検討する予定である．現在は質問記事を投稿月ごとに分割し，それぞれのデータでクエリ拡張を行ったものを提示しているが，本論文で示した季節性を持つ単語の抽出手法と組み合わせることで，季節に合ったユニーク語を数多くタグクラウドで提示する手法を検討している．

[謝辞]

本研究の一部は科研費：基盤研究 C(2150009)の助成を受けたものである．本研究の実装・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] 大塚淳史，関洋平，神門典子，佐藤哲司．情報要求の言語化を支援するクエリ拡張型 Web 検索システムに関する一検討．情報処理学会 論文誌 データベース（TOD），Vol.4, No.3, pp.1－11, 2011.

[2] H.Cao, D. Jiang, J.Pei, Q. He, Z.Liao, E. Chen, and H. Li. Context-aware Query Suggestion by Mining Click-through and Session Data. *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'08)*, pp.875－883, 2010.

[3]C.Sengstock and M.Gertz. CONQUER: A System for Efficient Context-aware Query Suggestions. *Proceedings of the 20th International Conference on World Wide Web (WWW'11),* pp. 265－268, 2011.

[4]J. Guo, X. Cheng, G. Xu, and H. Shen. A Structured Approach to Query Recommendation with Social Annotation Data. *Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM'10)*, pp.619 － 628, 2010.

[5] J. Reisinger and M. Pasca. Fine-Grained Class Label Markup of Search Queries. Proceedings of the 49th Annual Meeting of the *Association for Computational Linguistics (ACL'11)：Human Language Technologies*, pp.1200－1209, 2011.

[6] Z.-J. Zha, L. Yang, T. Mei, M. Wang, Z. Wang, T.-S.Chua, and X.-S. Hua. Visual Query Suggestion: Towards Capturing User Intent in Internet Image Search. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, pp. 13:1－13:19, 2010.

[7] Atsushi Otsuka, Yohei Seki, Noriko Kando, Testuji Satoh. QAque: Faceted Query Expansion techniques for Exploratory Search using Community QA Resources. *Proceedings of the 21st International Conference Companion on World Wide Web (WWW '12 Companion)*, pp.799－806, 2012.

[8] M. Vlachos, C. Meek, Z. Vagena, and D. Gunopulos. Identifying Similarities. Periodicities and Bursts for Online Search Queries. *Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data (SIGMOD'04)*, pp. 131－142, 2004.

[9]山家雄介，中村聡史， アダム ヤトフト，田中克己．ソーシャルブックマーキングの周期性発見と時期連動型検索ランキングへの適用．情報処理学会 論文誌 データベース（TOD），Vol.2, No.3, pp.130－140, 2009.

大塚 淳史　Atsushi OTSUKA

筑波大学大学院図書館情報メディア研究科博士前期課程在籍．2011 年筑波大学情報学群知識情報・図書館学類卒業．Web 情報検索，大規模データからの知識処理技術に興味を持つ．情報処理学会，日本データベース学会各学生会員．

関 洋平　Yohei SEKI

1996 年慶應義塾大学大学院理工学研究科計算機科学専攻修士課程修了．2005 年総合研究大学院大学情報学専攻博士後期課程修了．博士（情報学）．同年豊橋技術科学大学工学部情報工学系助手．2008 年コロンビア大学コンピュータサイエンス学科客員研究員．2010 年筑波大学大学院図書館情報メディア研究科助教，現在に至る．自然言語処理，意見分析等の研究に従事．ACM, ACL，情報処理学会，電子情報通信学会，言語処理学会，日本データベース学会各会員．

神門 典子　Noriko KANDO

1994 年慶應義塾大学大学院文学研究科博士課程修了．博士（図書館・情報学）．同年学術情報センター助手．1995 年米国シラキウス大学情報学部客員研究員，1996～1997 年デンマーク王立図書館情報大学客員研究員．1998 年学術情報センター助教授．2000 年国立情報学研究所助教授，2002 年より総合研究大学院大学助教授を併任，2004 年より国立情報学研究所ならびに総合研究大学院大学教授，現在に至る．テキスト構造を用いた検索と情報活用支援，探索や学習のための情報アクセス技術，情報検索システムの評価等の研究に従事．ACM-SIGIR,ASIS&T, 言語処理学会，日本図書館情報学会各会員．

佐藤 哲司　Tetsuji SATOH

1980 年山梨大学工学部電子工学科卒業．同年日本電信電話公社武蔵野電気通信研究所に入所．以来，論理回路の大規模一括集積技術，データベースマシン，マルチメディアデータベースの研究・開発に従事．1994 年工学博士(大阪大学)取得．2007 年 4 月より現職．分散並列処理，情報検索，社会インタラクションに興味を持つ．電子情報通信学会，情報処理学会各会員．