

問い合わせ拡張による情報要求の言語化支援技術
に関する研究

筑波大学
図書館情報メディア研究科
2020年4月
大塚淳史

目次

第 1 章	序論	1
1.1	背景	1
1.2	目的とアプローチ	2
1.3	オープンナレッジとクローズドナレッジ	3
1.4	本研究の構成	4
第 2 章	関連研究	7
2.1	クエリ拡張に関する研究	7
2.1.1	コンテキストアウェアなクエリ拡張に関する研究	8
2.1.2	拡張の根拠を提示するクエリ拡張手法に関する研究	9
2.2	コミュニティ QA と Web 検索に関する研究	9
2.3	機械読解に関する研究	10
2.4	質問の書き換え・パラフレーズに関する研究	11
2.5	質問生成に関する研究	12
第 3 章	オープンナレッジにおける情報要求の言語化	13
3.1	はじめに	13
3.1.1	本章の構成	14
3.2	単語の出現確率推移による話題変動抽出	14
3.2.1	分散とバーストによる周期性を持つ単語抽出	16
3.2.2	単語の出現確率推移による話題遷移抽出結果	17
3.3	トピックモデルによる CQA の話題変動抽出	23
3.3.1	時系列トピックモデルによる話題変動抽出手法	24
3.3.2	CQA の話題変動抽出結果	25
3.3.3	周波数解析による周期的なトピックの抽出	33
3.3.4	CQA からの周期的なトピックの抽出実験	35
3.4	情報要求を提示するクエリ拡張型 Web 検索システム	43

3.4.1	質問記事付き拡張クエリ	44
3.4.2	タブとタグクラウドによるファセット検索インターフェース	44
3.4.3	データセット	46
3.4.4	タブによるコンテキスト提示の実装	46
3.4.5	タグクラウドの実装	47
3.4.6	質問記事検索と拡張クエリ作成の実装	49
3.5	評価実験	50
3.5.1	キーフレーズ抽出による Web 検索の多様性評価	50
3.5.2	季節の違いによるクエリ拡張の評価	56
3.6	考察	57
3.6.1	クエリ拡張におけるカテゴリと多様性	57
3.6.2	CQA の季節性とクエリ拡張への影響	59
3.7	まとめ	60
第 4 章	クローズドナレッジにおける情報要求の言語化	61
4.1	はじめに	61
4.1.1	本章の構成	62
4.1.2	問題定義	62
4.2	改訂質問生成による機械読解モデル	64
4.2.1	Encoding layer	66
4.2.2	Matching layer	67
4.2.3	Decoding layer	68
4.2.4	改訂質問生成モデル学習	70
4.2.5	文圧縮による学習データ作成	70
4.3	評価実験	71
4.3.1	質問セット	71
4.3.2	改訂質問生成モデルによる質問の復元実験	72
4.3.3	改訂質問による機械読解実験	76
4.4	考察	82
4.4.1	短文質問に対する改訂質問生成に関する考察	82
4.4.2	長文質問に対する改訂質問生成に関する考察	82
4.5	まとめ	83
第 5 章	考察	84
5.1	オープンナレッジにおける情報要求の言語化	84

目次	iii
5.2 クローズドナレッジにおける情報要求の言語化	85
5.3 オープン/クローズドにおける情報要求の言語化の違いと今後の展望	85
第 6 章 結論	88
6.1 本研究のまとめ	88
参考文献	92
業績一覧	101

目次

1.1	情報要求の言語化	2
1.2	オープンナレッジとクローズドナレッジ	3
3.1	“Christmas” に関する GoogleTrend でのトレンド変化	15
3.2	“Christmas Movies” に関する GoogleTrend でのトレンド変化	15
3.3	“The Nightmare Before Christmas” に関する GoogleTrend でのトレンド変化	15
3.4	Freq により抽出された単語の出現確率推移	19
3.5	C.V により抽出された単語の出現確率推移	20
3.6	C.V+Burst により抽出された単語の出現確率推移	21
3.7	旅行カテゴリでの単語の出現確率推移	22
3.8	恋愛カテゴリでの単語の出現確率推移	22
3.9	DTM のグラフィカルモデル (時間分割数 3 のとき)	25
3.10	PC カテゴリ全トピックの JS ダイバージェンス (2006 年 1 月基準)	26
3.11	旅行カテゴリ全トピックの JS ダイバージェンス (2006 年 1 月基準)	27
3.12	経済カテゴリ全トピックの JS ダイバージェンス (2006 年 1 月基準)	27
3.13	PC カテゴリ, トピック 3 の JS ダイバージェンス推移 (2006 年 1 月基準)	28
3.14	PC カテゴリ, トピック 3 内の単語の確率変動	29
3.15	PC カテゴリ, トピック 2 の JS ダイバージェンス推移 (2006 年 1 月基準)	30
3.16	PC カテゴリ, トピック 2 内の単語の確率変動	30
3.17	旅行カテゴリ, トピック 1 の JS ダイバージェンス推移 (2006 年 1 月基準)	31
3.18	旅行カテゴリ, トピック 1 の単語の確率変動	32
3.19	旅行カテゴリ 200 トピックの周波数解析によるクラスタリング結果	35
3.20	旅行カテゴリ, トピック 0 の話題遷移	36
3.21	旅行カテゴリ, トピック 0 のパワースペクトル分布	37
3.22	旅行カテゴリ, トピック 0 の各月の単語分布によるクラスタリング結果	37
3.23	旅行カテゴリ, トピック 171 の話題遷移	38

3.24	旅行カテゴリ, トピック 171 のパワースペクトル分布	39
3.25	旅行カテゴリ, トピック 171 の各月の単語分布によるクラスタリング結果 . .	39
3.26	旅行カテゴリ, トピック 15 の話題遷移	40
3.27	旅行カテゴリ, トピック 15 のパワースペクトル分布	41
3.28	旅行カテゴリ, トピック 15 の各月の単語分布によるクラスタリング結果 . .	41
3.29	クエリ拡張型 Web 検索システムのプロトタイプ画面	43
3.30	クエリ拡張型 Web 検索システムのシステム構成図	44
3.31	質問記事付き拡張 (CQA クエリ)	45
3.32	タブ-タグクラウドインターフェース	45
3.33	提案法 (cqa) と既存手法 (yahoo) のキーフレーズ抽出数	54
3.34	“日本語”でのキーフレーズ抽出数の推移	54
3.35	“ソフト”でのキーフレーズ抽出数の推移	55
3.36	抽出カテゴリ数とキーフレーズ数の相関	56
3.37	ユニークキーワード数の季節ごとの推移	59
4.1	改訂質問生成の概要	63
4.2	SQuAD データセットの質問文中の単語について.	64
4.3	改訂質問生成と機械読解による質問応答.	65
4.4	改訂質問生成モデル (specific question generation model : SQGM). . . .	66
4.5	実験で使用する質問セットの概要.	72
4.6	SQuAD の圧縮質問における, 単語の削除数毎の質問の復元結果.	76
4.7	jpWiki における改訂質問による機械読解の例.	81
4.8	SQuAD における改訂質問による機械読解の例.	81

表目次

3.1	出現確率平均上位の単語 (Freq)	19
3.2	変動係数上位の単語 (C.V)	20
3.3	変動係数とバースト抽出による単語 (C.V+Burst)	21
3.4	旅行カテゴリでの単語抽出結果	22
3.5	恋愛カテゴリでの単語抽出結果	22
3.6	PC カテゴリトピック 3 の特徴語と変動語	28
3.7	PC カテゴリトピック 3 の特徴語と変動語	30
3.8	旅行カテゴリトピック 1 の内容語と変動語	32
3.9	旅行カテゴリでの階層的クラスタリング結果 (分割数 3)	35
3.10	旅行カテゴリ各トピックの代表的な単語	40
3.11	PC, 健康, 経済カテゴリでのトピック分類結果	42
3.12	PC, 健康, 経済糧での代表的な単語	42
3.13	Navigation Category とデータセット例	47
3.14	入力語 “ウィルス” の拡張クエリ	52
3.15	入力語 “mac” の拡張クエリ	52
3.16	各検索エンジンのキーフレーズ抽出数	52
3.17	入力語とそのカテゴリ	53
3.18	入力語 “ソフト” の拡張クエリ	53
3.19	全クエリに対する提案法と既存手法の比較	55
3.20	タグクラウドの出力例とユニークキーワード数	58
3.21	CQA クエリの例	59
4.1	質問の復元実験で使用する機械読解データセット.	73
4.2	質問の復元実験に関する実験結果.	75
4.3	改訂質問による機械読解で使用する日本語機械読解データセット.	76
4.4	SQuAD における機械読解の回答精度の比較結果.	79

4.5	jpWiki における機械読解の回答精度の比較結果.	80
-----	------------------------------------	----

第 1 章

序論

1.1 背景

情報社会と呼ばれる現在，画像や音声、テキストなど様々なメディアは電子化され，大量に蓄積・流通されるようになった．特に，世界中の情報を共有可能な World Wide Web (Web) と，大量の情報から所望の情報を取得できる，情報検索をはじめとした情報アクセス技術の発達によって，我々は，自身が必要と思った情報を瞬時に探し出し，取得できるようになった．情報アクセスの手段も多様化しており，コンピュータだけでなく，スマートフォンやスマートスピーカー，スマートウォッチ，車載システムなど様々なデバイスから Web 上の情報にアクセス可能となっている．また，情報検索は，大規模なデータの中から，利用者が所望する情報が含まれる資料（テキストなど）を見つけ出す技術であるが [43]，近年では，質問と回答の形式で，利用者が必要とする情報そのものを提示する質問応答も，深層学習を始めとする人工知能技術の発達に伴い，広く実用化されるようになった．

Web に蓄積される情報も日々増加の一途を辿り，情報アクセス手段も多様化する中，誰もが自身の所望の情報にアクセスできるとは限らない．情報検索システムや質問応答システムを使って，情報を得るためには，最初に，自身が取得したい情報について，明確に言語化されている必要がある．例えば，情報検索の場合には，取得したい情報である情報要求を特徴的に表現するキーワード組をクエリとして，システムに入力する必要がある．情報要求とは，所望する情報を入手することで達成できる意識的，または無意識的な欲求である．近年の研究でも，検索におけるクエリの性能は，背後に存在する情報要求と関連させて評価することが望ましいことが報告されており [74]，情報検索において，情報要求は非常に重要な観点といえる．情報アクセスにおける情報要求は Taylor[63] が，情報要求を曖昧なレベルから visceral need, conscious need, formalized need, compromised need の 4 つのレベルに分類できることを報告している．このとき，情報検索や質問応答を有効に使用できるのは，情報要求が具体的に言語化されている formalized need 以上の情報要求であり，これ以上曖昧な情報要求の場合，入

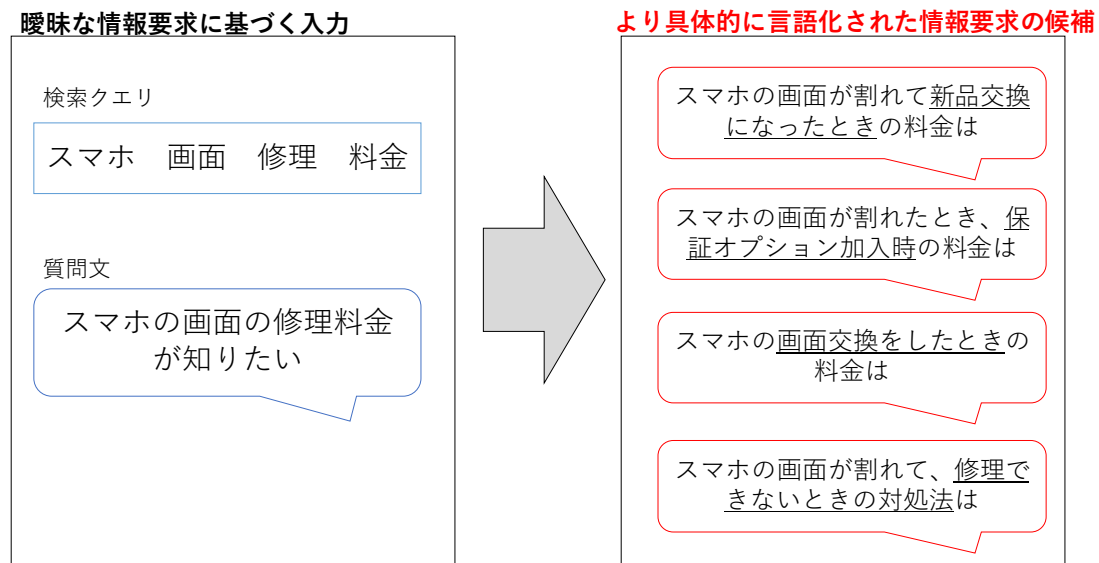


図 1.1 情報要求の言語化

力となるクエリや質問文を作成することは難しく，少数の単純なキーワードによるクエリや曖昧な質問文しか作成することができない．短いクエリや曖昧な質問文では情報検索や質問応答の性能を十分に発揮することができず，必要な情報を得られないという課題がある [68, 35]．

1.2 目的とアプローチ

本研究では，情報要求の言語化に着目した情報アクセスの支援技術に取り組む．システムが情報要求の言語化を支援することで，曖昧な情報要求を持った利用者であっても，最終的には自身の所望する情報を入手できるようにすることが目的である．

情報要求の言語化支援のアプローチとして，利用者が最初に入力したクエリや質問文から想定される情報要求の候補を利用者に複数提示し，選択させるという手法を採用する．ここで，提示する情報要求とは，自然言語で記述されたテキストである．本研究における情報要求の例を図 1.1 に示す．利用者は提示された情報要求を閲覧することで，自身の入力に不足していた情報が何であるかを知ることができるようになる．また，候補の中から自身の情報要求に最も合致していると考えられる候補を選択することで，選択した情報要求に関する情報検索や質問応答が実行可能となる．選択式にすることで，利用者は再入力の手間を必要とせず，提示された候補を比較することで，自身の情報要求を言語化しながら，様々な情報にアクセスできるようになる．

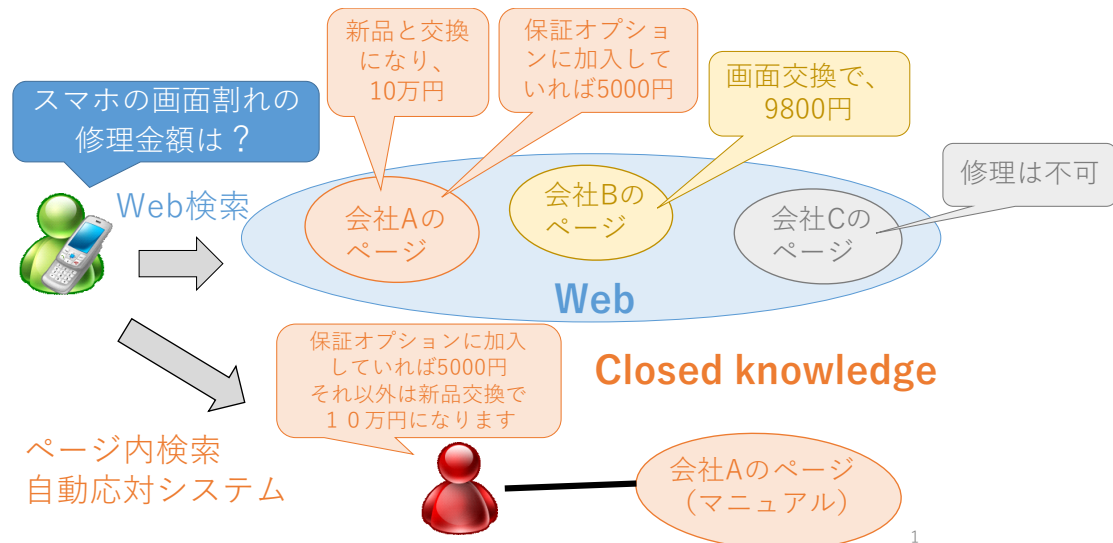


図 1.2 オープンナレッジとクローズドナレッジ

1.3 オープンナレッジとクローズドナレッジ

本研究では、情報アクセスを行う利用者に対して、情報要求の言語化を支援するために、自然言語で記述された情報要求を提示する。本研究の中心となるリサーチクエションは、利用者に提示する自然言語の情報要求をどのように生成するかという点である。近年では、深層学習を用いることで、比較的容易に自然言語を生成することが可能であることが知られている[62]。しかしながら、深層学習を利用するためには、モデルを学習するための学習データが必要となる。学習データを作成するためには、利用者の入力とその背後にある情報要求との関係を明らかにする必要がある。

本研究では、利用者が使用する情報アクセスの手段や、アクセス対象となるリソースの違いによって、必要とする情報要求は異なると仮定する。例を図 1.2 に示す。例えば、修理金額を調べたいという情報要求があったとき、Web 検索を使用した場合は、様々な Web ページから多様な金額に関する情報を得ることができる。このとき、これらのどの情報が利用者が真に知りたい情報であるかを判別することは難しいため、「新品交換となる場合」、「保証オプションに加入している場合」、「画面交換の場合」など様々な状況に対する情報要求の候補を提示し、利用者自身が近いと思うものを選択させることが必要となる。一方で利用者がある特定の Web ページにアクセスし、その中での検索やページ内に設置されている自動応対システムを利用することを考える場合、存在する情報はそのページ内に限定される。ここで、そのページ内に存在しない情報に関する情報要求を提示しても、その情報要求を満たす情報が存在しないだけでなく、間違った情報を利用者に伝えてしまう可能性がある。図 1.2 の例では、会社 A のページ

内の探索で、画面交換や修理不可の場合の情報要求を提示することは避ける必要がある。

本研究では、Web に代表される膨大な情報源をオープンナレッジ、Web ページやマニュアルなどの特定の範囲に限定された情報源をクローズドナレッジと定義し、それぞれで別のアプローチによって言語化された情報要求を生成し、提示する。オープンナレッジでは、情報の探索空間が膨大であることの利点を生かして、様々な観点から多様な情報にアクセスするための情報要求の候補を提示する。一方でクローズドナレッジにおける言語化支援では、情報の特定性に着目し、探索空間を絞り込むための情報要求を提示する。本研究では、オープンナレッジの代表として、Web 検索における情報要求の言語化支援技術について取り組む。クローズドナレッジについては、近年マニュアル検索や自動応対で利用される機械読解による質問応答技術を対象とした言語化支援を扱う。

1.4 本研究の構成

第 1 章では本研究の背景、目的と問題設定を行い、本研究の目的が、情報アクセスにおける情報要求の言語化支援技術であること、アプローチとして、自然言語で記述された情報要求の候補を提示すること、そして提示する情報要求の生成・提示手法に対して、膨大な情報源から探索するオープンナレッジと、特定の情報源を探索するクローズドナレッジに分けてそれぞれの観点から異なる手法を提案することを説明している。

第 2 章では、本研究の関連技術を示す。まず、オープンナレッジにおける言語化支援に関する関連研究として、情報検索における利用者支援技術であるクエリ拡張技術についての関連研究を紹介する。また本研究の対象である Web 検索やコミュニティ QA に関する関連研究について述べる。次に、クローズドナレッジの情報要求の言語化支援で対象とする質問応答および機械読解技術に関する先行研究を示す。また、機械読解の基盤技術となっている深層学習による言語処理技術、言語生成技術についての関連研究も示す。

第 3 章では、オープンナレッジにおける言語化支援として、Web 検索における情報要求の言語化支援技術について提案する。オープンナレッジにおいては、より多様な情報にアクセスできることを目的とする。そこで、提示する情報要求の知識源として、コミュニティ QA (CQA) の質問記事を利用する。コミュニティ QA は、身近な様々な疑問や悩みについて利用者は QA 形式で自由にコミュニケーションを行えるサービスである。ここで、投稿された質問記事を、言語化された情報要求とみなすことで、検索利用者に提示する情報要求の候補として採用する。さらに質問記事から、検索で使用する拡張クエリを作成しセットで提示することで、選択した質問記事に関する Web 検索を実行可能にする。コミュニティ QA は多くの利用者が様々な質問記事を投稿することから、多様な観点に基づいた情報要求の候補を多数提示で

きるという利点がある。また、質問記事を投稿したカテゴリや投稿時期といったコンテキストを考慮することで、同じキーワードであってもより多様な情報要求を提示できるようになることを実験的に示し、コミュニティ QA の質問記事を情報要求の候補として、Web 検索に活用することの有効性について述べている。

第4章では、クローズドナレッジにおける言語化支援として、機械読解による質問応答における、改訂質問生成技術を提案する。機械読解では、利用者は自然言語で記述した質問文を入力することで、質問に対する回答を対象テキストから発見する技術であるが、このとき、曖昧な質問文を入力してしまうと、回答箇所を特定することができず、回答精度が低下するという課題があった。改訂質問生成では、曖昧な質問文を、機械読解で使用するテキストの内容に基づいて具体化した質問文を生成し、ユーザに提示する。具体化された質問を再度機械読解に入力することで、機械読解の回答が変化し、より特定性の高い回答を得られるようになる。改訂質問生成では、自然言語の質問文が複数生成されるため、これらの候補を利用者に選択させることでクローズドナレッジにおいて、特定性を高めるための、言語化された情報要求を提示することができる。本研究では、改訂質問生成を実現するための、深層学習モデルについても提案する。提案モデルでは、言語生成のモデルをベースに、機械読解とコピー機構を組み合わせることで、対象テキストの内容に即した改訂質問が生成できる。また、深層学習モデルを学習するための学習データについて、機械読解の学習コーパスから自動生成する手法を提案している。提案手法では、機械読解コーパスの質問文を文圧縮によって、短くすることで擬似的に曖昧な質問文を作成している。実験では、日英の機械読解コーパスにおいて、短い質問文を入力すると機械読解の回答精度が低下すること、そして、短い質問文から改訂質問生成によって作成された改訂質問をしようすることにより、回答精度が改善されることを示す。

第5章では、オープンナレッジとクローズドナレッジそれぞれに対して、言語化された情報要求を提示することの効果と、本研究の提案手法の効果について考察を行う。Web 検索においては、自然言語で記述された情報要求を提示することは、特に同音異義語のような語に対して効果的であるといえる。従来のクエリ拡張では、追加されたキーワードのみしか利用者に提示されないため、同音異義語の、異なる話題に関する追加キーワードが混在して表示された場合、利用者がそれらを判別することは難しい。そこで、本研究の提案手法のように自然言語で記述された質問記事を閲覧することで、クエリを作る際に消えてしまったコンテキストを含めて理解できるため、同音異義語の区別を判別しやすくなる。コンテキストがわかるというのは、多様な情報を探索するために大量の拡張クエリに対しても、その背後に隠れている情報要求を素早く理解できるという点においても重要であったと考えられる。機械読解においては、改訂質問生成が、機械読解の回答に置いて重要となる情報をピンポイントで抽出、追加している点が重要であると考えられる。機械読解では、自然言語の質問文を入力するが、自然言語の

場合、主語などの情報を省略してしまう場合が多い。これは日本語において特に顕著に発生するが、改訂質問生成では、これら省略されやすい主語等に該当しそうな固有名詞などの語を多く抽出している。また、改訂質問生成モデルに機械読解とコピー機構を導入したことで、対象文書で使用されている範囲の単語のみで具体化するという点が機械読解の回答精度の改善に繋がったと考えられ、クローズドナレッジにおける情報要求の言語化の方向性を示すことができたのではないかと考えられる。

最後に 6 章では、本研究をまとめている。情報アクセスの手段の高度化および蓄積される情報量の増加によって、利用者がアクセス可能な情報は格段に増加したと言える。本研究では、その中で、利用者自身の“何を知りたいか”という情報要求に着目することによって、利用者 1 人ではアクセスできなかった情報にアクセス可能にする情報要求の言語化技術について取り組んだ。その結果、膨大な情報を有する Web 検索では、情報要求を明確にしながらより多様な Web 検索を実現可能とするクエリ拡張手法を実現した。また、特定文書を対象にした機械読解では、入力した質問文から不足している情報を推定し、その情報を文書中の情報を用いて補強する改訂質問生成技術について取り組み、具体化した質問文の候補を提示することで、回答の特定性が高まることを示した。

第 2 章

関連研究

2.1 クエリ拡張に関する研究

本研究では，コミュニティ QA の質問記事を用いたクエリ拡張により，Web 検索支援を行う．クエリ拡張は多くの研究がなされていると共に，商用の検索エンジンでも提供されており，Web 検索エンジンにおいて重要な機能となっている．Kato ら [36] は検索エンジンにおいてどのような場合にユーザが推薦された拡張クエリを使用するのかを分析し，ユーザが推薦されたクエリを使用するのは，クエリのエラー修正を行う場合のみならず，ユーザが入力したクエリが単一の単語のみである場合や，ユーザが検索結果の URL をいくつかクリックした後にもよく使用されることを明らかにしている．

商用の Web 検索エンジンでは，検索エンジンに入力されたクエリログを元にクエリ拡張を行っている．クエリ拡張に他の Web ページなどの外部リソースを使用することも有用であることが多くの研究で示されている．Yin ら [71] は，Web 検索エンジンで検索した Web ページのスニペットから拡張クエリを作成することが効果的であることを示している．村田ら [88] は，多くの検索ユーザからアクセスされる Web ページのスニペットから拡張クエリを作成することで，追加する語が少ない場合の検索精度を向上させている．堀ら [80] は，Web 百科事典 Wikipedia から作成した拡張クエリは，Web 検索結果の疑似適合フィードバックから作成した拡張クエリよりも，ユーザ満足度が高くなることをユーザ実験によって示している．Web 上ではユーザ自身が積極的に情報を発信していることから，水野ら [83] は，ユーザが記述した blog やブックマークから作成したユーザプロフィールを情報源とすることで，ユーザの趣向にあった拡張クエリを作成できるとしている．Jaech らは，RNN 言語モデルに基づいて，クエリに追加するキーワードを生成する際に，ユーザに関する情報を埋め込んだベクトルを追加することで，個人の趣向に合わせたクエリが推薦できる手法を提案している [31]．Jiang らの研究

では、一連の情報検索行動を追跡し、閲覧した Web サイトやそこで使用されている単語情報をクエリ生成のニューラルネットワークに組み込むことによる、コンテキスト依存なクエリ推薦を実現している [33]。Li ら [39] は、ユーザのフィードバックから検索意図を反映した拡張クエリを正しく推薦するため、敵対的学習を用いたクエリ推薦手法を提案している。Web の文書の観点からクエリ拡張を検討する研究には、Han らの研究がある [25]。Han らは Web 文書を与えたとき、文書にヒットするクエリを推論・生成するため、グラフとニューラルネットワークを組み合わせた生成モデルを提案している。クエリ推薦は主に情報検索をターゲットにし、キーワードの追加修正を行う技術である。本論文の改訂質問生成は、質問応答をターゲットに自然文の質問文を書き換えるのでこれらの研究とは異なる。

情報検索型の質問応答では、入力質問を解析し、質問に対する適合文書を発見することで、回答となる情報を抽出する。このとき、より文書に適合しやすくするために、クエリ拡張やクエリの変形などの処理を行う。しかし、これらの処理は知識源の冗長性に依存して行われるので [8, 40, 77]、ある特定の知識について質問を書き換える必要のある機械読解では有効に機能しない。

以下では、本研究に関連の深いコンテキストアウェアな Web アクセスを実現するためのクエリ拡張、拡張クエリの根拠を提示する研究に関して述べる。

2.1.1 コンテキストアウェアなクエリ拡張に関する研究

本研究では、CQA のカテゴリと投稿の季節を検索の切り口（ファセット）として、ファセットを切り替えることにより、様々なコンテキストに対応した拡張クエリを提示している。ユーザのコンテキストを推定し、コンテキストに対応したクエリを提示する手法では、Cao ら [10] や Semgstock ら [56] の研究がある。Cao らはクリックログとセッションデータから、現在のセッションに近いクエリをログデータから推薦する手法を提案している。また、Semgstock らはクエリログから、時間とドメインに依存したクエリを抽出し、時間と場所を変化させることで、拡張クエリが変化するシステムを提案している。Ahmad ら [2] は、系列を保持できるニューラルネットワークモデルである LSTM を用いて、ユーザの検索過程から、検索結果のランキングとクエリ推薦を同時に行うマルチタスク学習モデルを提案している。Völske ら [67] は、クエリ推薦の強化を目的として、クエリログのクエリをタスク別にマッピングする技術を提案している。

ファセット検索は単純な検索クエリから多様な Web 検索を実現できる手段として、探索的情報検索分野において、多くの研究がなされている。Jonathan ら [37] は、ファセット検索とユーザプロファイルと組み合わせることにより、個人の趣向に合わせたインタラクティブな情報検索を実現する手法を実現している。Hearst [28] は、ファセットを階層構造化させる Web 検索インターフェースを提案している。廣嶋ら [84] はユーザが入力したクエリを“グルメ”、“

スポーツ”，“企業名”などのタイプに分類し，タイプに応じた Web 検索結果を提示している．クエリの意味を提示する研究では，クエリログやアクセスした URL のログから自動でタグやタイプを推定している．多義的なクエリを推薦する手法として今井ら [81] は，クエリと URL からなる 2 部グラフを用いたクラスタリングを行い，話題が偏らないクエリ推薦を行うことが可能であることを明らかにしている．

2.1.2 拡張の根拠を提示するクエリ拡張手法に関する研究

本研究では，CQA の質問記事を拡張クエリと共に提示することにより，拡張クエリの根拠となる情報要求を提示する．クエリの意味や根拠を提示する手法については，Guo ら [23] や，Lin ら [41] が，ソーシャルアノテーションに基づくクエリ拡張を提案している．拡張の際に用いたソーシャルアノテーションをそのまま用いることで，クエリの分類と意味付与を同時に行っている．自然言語処理の技術を用いてクエリにラベルを付与する手法では，Reisinger ら [55] 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，クエリに付与するラベルを作成し，確率文脈自由文法を用いてクエリとラベルの関係を紐付けている．また，クエリの意味として文章ではなく画像ファイルをクエリと共に提示する手法を Zha ら [75] が提案している．Zha らは画像共有コミュニティに投稿されている画像と画像に付与されているタグを用いて，画像付きの拡張クエリを提示するシステムを作成している．

2.2 コミュニティ QA と Web 検索に関する研究

コミュニティ QA (Community Question Answering: CQA) は，ソーシャルベースの新たな情報アクセス手段として注目を集めている．CQA 内で議論される話題に関する研究は，Adamic ら [1] の研究がある．Adamic らは，CQA での話題の最も基本的な単位であるカテゴリに着目し，カテゴリごとにコミュニケーションのタイプが異なることを明らかにしている．また，Miao ら [44] は，CQA の中から新たなカテゴリとなる話題を，トピックモデルによって発見する手法を提案している．

CQA と既存の情報アクセス手段である Web 検索と結びつけることで，より高度な情報検索を実現する研究がなされている．Liu ら [42] は Web 検索ユーザが CQA ユーザとなるまでの経過を分析している．CQA ユーザは質問記事を投稿する際に，Web 検索結果中の CQA ページを閲覧してから利用するケースが多く，また，Web 検索クエリに関しても，より具体的なクエリを投稿する特徴があることを明らかにしている．Yoon ら [72] の研究では，ユーザの要求とコミュニティ QA のカテゴリを関連付け，Web 検索結果をコミュニティ QA のカテゴリにより分類，再ランキングを行う手法を提案している．山本ら [79] は，コミュニティ QA から形容詞と名詞の組み合わせによる修飾語付き観点を抽出し，タグクラウドとしてユーザに

提示する．修飾語付き観点はユーザがより直感的に分かりやすい表現となっている．実験により，修飾語付き観点は，これまでの検索ではなかなか思い浮かばない意外な組み合わせの語が推薦されることを明らかにしている．高田ら [78] は，コミュニティ QA の質問記事と回答記事に着目し，回答記事の別解情報を含む Web ページを収集することで，質問記事に関連のある Web ページを効率的に閲覧できる手法を実現している．

2.3 機械読解に関する研究

機械読解とは，システムが自然言語で記述された文書（パッセージ）を読み解き，そのパッセージに関する質問に対して，回答となる情報をパッセージ中から発見，抽出することで回答を可能とする質問応答技術である．従来の情報検索と異なり，質問に対する回答部分となる情報のみを提示するという特徴がある．機械読解は，Seo らの BiDAF [57] や，Yu ら QANet [73] などのニューラルネットワークを用いたモデルにより，高い回答精度が実現できることが知られている．近年では，文字や単語の意味をベクトルとして保持するというニューラルネットワークモデルの特徴を生かし，異なる言語間での質問応答を実現する機械読解モデルも提案されている [15, 34]．

機械読解は，質問に対する回答を文書中の表現から抜き出す抽出型のアプローチが一般的であったが，近年では，生成モデルにより，回答を生成する手法も提案されている．Tan ら [13] は，機械読解したテキストをもとに回答を生成するモデルである S-Net を提案している．Nishida ら [49] は，機械読解における回答を生成モデルによって生成すると同時に，回答スタイルを定義し，短い回答や長い回答などスタイルの異なる回答を同一モデルで任意に選択し，生成する手法を提案している．

機械読解では質問と，質問に対する回答を抽出するための知識源となる約 1 段落分の文章を入力とする．従来のデータセットは入力となる文書の中に必ず回答が含まれていることを保証していたが，機械読解モデルの性能向上に伴い，複数の文章を参照する必要があるデータセット [70] や，“回答なし”が回答となるようなデータセット [53] が提案されるようになった．複数の段落を参照する機械読解では，Nishida ら [48] が複数の段落の文書から回答で使用する根拠となる情報のみを系列モデルで抽出していくモデル（Query Focused Extractor：QFE）を提案している．Tu ら [46] は，複数の段落の文章の中からまず不要と思われる文章を除外してから，単語，文章で異なる粒度での回答のマッチングを見ることで，計算量を抑えながら高精度に回答できるモデルを提案している．

ニューラルネットワークによる機械読解モデルでは，単語や文章の意味をすべてベクトルとして表現するが，入力の初期値として，Word2vec[45] や Glove[51] といった単語埋め込みモデルが使用されていた．しかしながら近年では，初期値として大量のコーパスで事前学習された言語モデルを使用することで，機械読解をはじめ，多くの自然言語処理タスクで従来研究より

も高い精度が達成できることが報告されている。Peter ら [52] が提案した Elmo では、ニューラルネットワークの初期値として、双方向の LSTM[30] で学習した言語モデルを使用する。Devlin ら [16] が提案した BERT では、ニューラルネットワークの構造に Transformer[65] を用い、学習に Masked Language Model を採用することで、文の前後双方向のコンテキストを解釈可能な言語モデルを学習することができる。BERT は、学習済みの BERT モデルの上位層にタスク固有の層を追加してファインチューニングを行うことで、多くのタスクに対しても高い精度を達成できる。機械読解においても BERT の出力層に抽出層を 1 層追加するのみで QANet 等の従来手法よりも高い精度が達成できることが知られている。近年では、より高精度を達成するモデルとして XLNet[69] や、軽量モデルである ALBERT[38] といった BERT の派生モデルが提案されている。

本論文では、クローズドナレッジにおける情報要求の言語化として機械読解タスクを対象とする。従来の機械読解研究では、人手で作成されたオープンデータセットの質問を対象とするが、本論文では、これらのデータセットに含まれないような不完全な質問を対象とする点が異なる。また、機械読解において生成モデルは回答を生成することを目的に使用されていたが、本論文では、質問文に対して生成モデルを適用するという違いがある。

2.4 質問の書き換え・パラフレーズに関する研究

質問応答の回答精度の向上を目的に、質問文を書き換える研究では、Buck らの研究がある [9]。Buck らは、質問者と質問応答システムとの間にブラックボックスで質問書き換えモデルを設置し、入力質問を質問応答システムが回答可能な質問に書き換えてから質問応答システムに入力することで、回答精度を高めている。質問書き換えモデルには、生成モデル (Seq2Seq) を用いて、回答精度を最大化するように強化学習を行うことでモデルパラメータの学習を行う。Buck らの手法では、事前学習済みの質問の書き換えモデルが必要となる。しかしながら、モデル学習のための言い換えコーパスを用意することが難しいという課題がある。また、書き換えモデルを学習したコーパスに含まれていない話題に関する質問が入力されたとき、後段の質問応答システムに無関係な質問が生成されてしまうという可能性がある。本研究が提案する改訂質問生成モデルは、言い換えコーパス等の別のデータセットが不要であるという点異なる。また、後段の機械読解で使用するパッセージ中の単語をコピーすることにより、改訂質問を生成するため、パッセージと無関係な質問が生成されにくいという特徴がある。

言い換え文を作成するパラフレーズも、質問の書き換えに関するタスクの一つといえる。Dong らは、言い換え文生成と質問応答を同時に行うモデルを構築し、学習を行うことで入力質問と同じ回答が得られるような言い換え文を生成する手法を提案している [17]。Gupta らの手法では、Variational Autoencoder (VAE) を用いて、言い換え文の変換スタイルを学習

させることで，文生成時には，スタイルを変えることで様々なバリエーションの言い換え文を生成している [24]．パラフレーズは質問の意味を変えないまま，異なる語彙を使用した質問を作成するタスクであり，質問を具体化する改訂質問生成のタスクにパラフレーズを適用することはできない．

2.5 質問生成に関する研究

質問生成は，主に教育分野で活用することを目的に，機械読解の逆問題として，パッセージと回答を与えたときに質問を自動生成するタスクとして取り組まれている．Du らは回答を含んだ文を入力すると，自然言語の質問を生成する手法を提案している [19]．また，文中に BIO タグと共参照関係のタグを埋め込むことで，パラフレーズ単位での回答に対する質問生成手法を提案している [18]．Duan らは，パッセージを入力したときに，回答となる文を選択するタスクを質問生成と同時に扱うことで質問生成の性能を向上させる手法を提案している．

質問生成は，パッセージなどの自然文だけでなく，画像を入力したときに，画像に関する質問を生成するタスク [47, 32] や，知識ベースから質問を作成するタスク [58, 20] も行われている．改訂質問生成は，知識源となるパッセージの他に，質問文自体を入力に与えるという違いがある．

第 3 章

オープンナレッジにおける情報要求の言語化

3.1 はじめに

インターネットの普及に伴い、World Wide Web (Web) に蓄積されるデータは増加の一途をたどっており、蓄積される情報もより多様なものとなっている。そのため、大量の情報が蓄積されている Web 上から所望の情報を入手するための、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、言語化することで検索クエリを作成する。検索クエリを作成することができれば、Web から所望の情報を入手することができるが、ユーザが想起できない“疑問”に対しては、情報要求を言語化することができず、クエリを作成することができない。ユーザ個人が考えうる“疑問”の範囲には限りがあるため、自由入力型のクエリ検索では、ユーザは、自身が想起できる情報要求の範囲の情報しか入手することができず、Web 情報に対して大規模なインデックスを持つ Web 検索エンジンを最大限に活用することができないという問題がある。

現在の Web 検索エンジンは、ユーザのクエリ作成を支援し、ユーザが想起できない情報要求に対する Web 検索を実現する手段として、クエリ拡張機能を提供している。クエリ拡張では、検索エンジンに蓄積されているクエリログデータから、ユーザの検索意図を推測し、ユーザが入力したクエリのキーワードの変更、又は新規キーワードを追加した新たなクエリを提示する [3]。ユーザが入力したクエリが単一のキーワードから構成される単純なものであっても、検索エンジンが提示した拡張クエリを選択することで、的確な Web 検索を実行できる。しかしながら、クエリ拡張によって提示される候補は、ユーザの情報要求と必ずしも一致しているとは限らない。最も典型的な例として、同音同字異義語が挙げられる。ユーザがコンピュータ関係の“ウイルス”を検索したいと思ったとしても、検索エンジン側が病気に関する“ウイルス”に関する拡張を行うという場合がある。特に、“RS ウイルス”、“MAC ウイルス”など、一

見コンピュータ関係か病気関係か判断できない候補が含まれる場合がある。同音異義語のように、語には、様々な観点が含まれる、複数の観点からクエリが推薦されることは、情報検索の多様性の観点から見れば好ましいことである。しかし、観点の違いをユーザが認識していない場合、ユーザの混乱を招く結果となる恐れがある。

本研究では、コミュニティ QA (CQA) の質問記事により Web 検索の支援を行う。CQA は、疑問を質問記事として自然言語で記述し、投稿することで他ユーザの回答を得ることができる知識共用サービスである。代表的な CQA の例として、Yahoo!知恵袋^{*1}や、教えて!goo^{*2}が挙げられる。自然言語で記述された質問記事は Web ユーザの“疑問”そのものであるといえる。質問記事を Web ユーザの情報要求とみなすことで、潜在的な情報要求が明確に言語化できる。そこで、質問記事を検索支援に用いることで、情報要求を意識した“疑問ベース”の Web 検索が実現できると考えている。

3.1.1 本章の構成

以下では、本章での取り組みについて説明する。まず 3.2 節で、周期性に着目した、CQA の話題変動の抽出について述べる。次に、3.4 節で、CQA を用いた Web 検索支援として開発した、コンテキスト切り替え型のクエリ拡張システムの概要を説明する。

3.2 単語の出現確率推移による話題変動抽出

インターネットの普及に伴い、Web ユーザは生活に密着した情報に関しても気軽に検索や質問記事の投稿を行うようになった。生活に密着した情報に関しては、アクセスする時期や時節の影響を大きく受けることがわかっている。GoogleTrend^{*3}では、検索クエリの使用頻度から検索のトレンドの推移を可視化するサービスを提供している。図 3.1 は、GoogleTrend で“Christmas”を検索した時の検索トレンドの推移を示している。クエリ“Christmas”は毎年 12 月に検索頻度が高く、その他の期間はあまり検索されない。このように、Web ユーザの情報要求の変化は周期性を持つことがわかる。

^{*1} <http://chiebukuro.yahoo.co.jp/>

^{*2} <http://oshiete.goo.ne.jp/>

^{*3} <http://www.google.com/trends/>

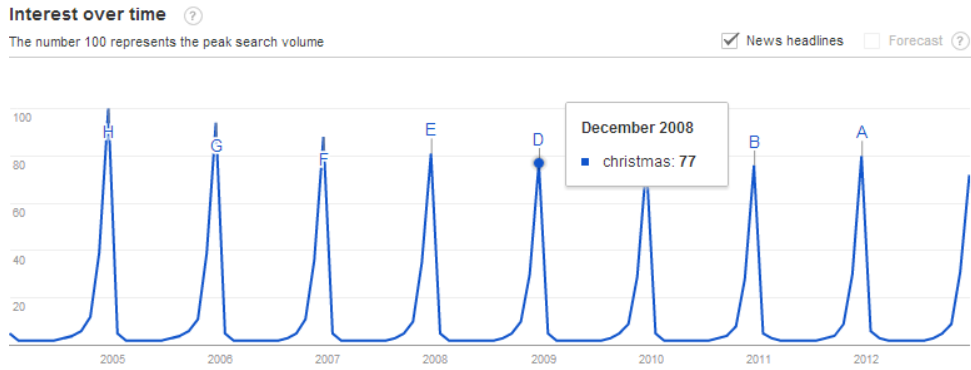


図 3.1 “Christmas” に関する GoogleTrend でのトレンド変化

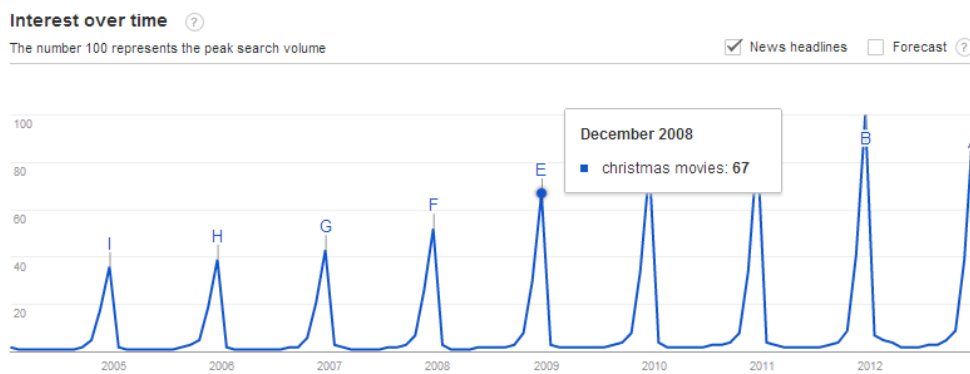


図 3.2 “Christmas Movies” に関する GoogleTrend でのトレンド変化

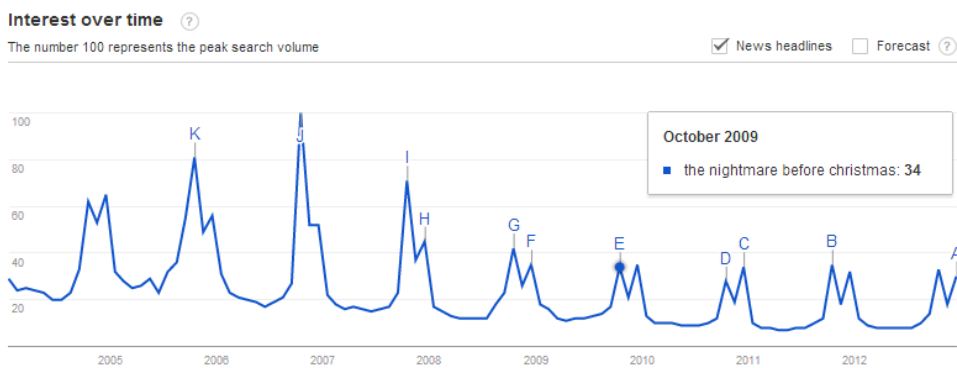


図 3.3 “The Nightmare Before Christmas” に関する GoogleTrend でのトレンド変化

本章では、CQA における質問記事について質問内容の周期性を明らかにする。質問内容に周期性がある場合、それらの周期にあった質問記事を用いて拡張クエリを作成することでよりユーザの状況に適合するクエリが推薦できる。本章ではまず、カテゴリ毎に投稿された質問記事内における単語の出現頻度の推移を追跡する。次に、質問記事をトピックモデルにより“話題”単位にクラスタリングし、話題の変化を追跡する。最終的に、CQA の中から、周期性を持つ話題のみを抽出する手法を提案する。

CQA の質問記事は自然言語で記述されているため、質問記事の話題の変動を確認するには、質問記事中の単語の使用頻度を分析することが最も一般的である。本節では、CQA のカテゴリごとに、質問記事中で使用される単語の出現確率の推移により CQA で周期的に発生する話題を抽出する。

各カテゴリにおいて、CQA 質問記事を MeCab^{*4}により形態素解析し、月ごとの単語の出現頻度をカウントする。月ごとの投稿質問記事数にはバラつきがあるため、ある月の出現頻度を、投稿された総質問記事数により正規化して出現確率とすることで、単語の使用頻度の推移を評価する。カテゴリ C において、1ヶ月間に投稿された質問記事のうち、単語 w の出現確率は以下で与えられる。

$$P_{C,w} = \frac{N_{C,w}}{N_C} \quad (3.1)$$

$N_{C,w}$ 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、 N_C はカテゴリ C にひと月に投稿された総質問記事数である。

3.2.1 分散とバーストによる周期性を持つ単語抽出

本章では、周期的に使用頻度が変動する単語を CQA において発見することを目的としている。周期性を持つ単語は、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、各単語の出現確率の分散を月ごとに計算し、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、ある特定の月に偏在する単語を抽出する。変動係数 C.V は以下の式で与えられる。

$$C.V = \frac{\sqrt{\sigma^2}}{\bar{x}} \quad (3.2)$$

このとき、 $\sqrt{\sigma^2}$ は出現確率の標準偏差、 \bar{x} は平均を示している。変動係数を各単語で計算し、降順に並べたとき、上位の単語を周期性を持つ語として抽出する。

^{*4} <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>

図 3.1 の “Christmas” は、毎年 12 月のみ検索頻度が急上昇し、その他の月では、ほとんど検索されない。このようにある特定の期間のみ、頻度や確率が上昇する現象を、本研究ではバーストと呼ぶ。バーストは、Web 上の様々な情報で発生するとされている。Vlachos ら [66] らは、クエリログから Web クエリのバースト性を検証しており、山家ら [87] 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その周期性を発見している。本研究においても、CQA のバーストを発見するために Vlachos らの手法により、単語の出現確率推移からバーストを抽出する。Vlachos らのバースト抽出手法を以下に示す。

1. データ系列 $t = (t_1, \dots, t_n)$ に対して長さ w の移動平均 MA_w を計算
2. 移動平均の平均 $mean(MA_w)$ と標準偏差 $std(MA_w)$ から閾値 $cutoff = mean(MA_w) + x * std(MA_w)$ を計算
3. i 番目のデータに対して、 $\{ MA_w(i) > cutoff \}$ の場合バーストと判定

3.2.2 単語の出現確率推移による話題遷移抽出結果

本節では、3.2.1 節で提案した、分散とバースト抽出による周期性を持つ単語抽出手法を実際の Yahoo!知恵袋データに適用した実験結果を示す。まず、提案手法の妥当性を評価するために、PC カテゴリにおいて、以下の 3 つの手法で単語抽出を行い、結果を比較する。

- **Freq** :単語の出現確率の平均値を降順に抽出
- **C.V** :単語の出現確率の変動係数を降順に抽出
- **C.V+Burst** :変動係数 (C.V) の結果に、バースト抽出によるバーストの周期性を考慮

期間は、2006 年 1 月から 2008 年 12 月までの 36 ヶ月間で実験を行う。実験で用いたカテゴリは、毎月安定して投稿質問数が多かった PC カテゴリ、旅行カテゴリ、恋愛カテゴリで実験に用いる。3.2.1 節で説明したバースト抽出手法のパラメータを、データ系列 t を 2006 年 1 月から 2008 年 12 月までの 36 ヶ月分の出現確率のデータ、移動平均の長さ w を季節の区切りと対応させ $w = 3$ と設定した。また、しきい値計算のための重み付けパラメータ x は予備実験の結果 $x = 2.5$ とした。本研究では、周期性のある単語を抽出するため、データセット 3 年間で、毎年同じ月にバーストが発生した単語のみを抽出する。

3 つの手法により、抽出できた上位の単語を表 3.1 から表 3.3 に示す。表 3.1 は、PC カテゴリにおいて出現確率 36 ヶ月間の平均が上位の語である。“教える”、“使う” など CQA 一般に使用頻度の高い語の他、PC 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。図 3.4 は、表 3.1 の単語の 36 ヶ月の出現確率の推移である。どの単語もほぼ同じ確率推移をしていることがわかる。2007 年 3 月にすべての単語の出現確率が低下しているがその他は、大きな変動は発生していない。表 3.2 は単語の出現確率の変動が上位の語である。図 3.5 にこれ

らの単語の出現確率推移を示す。“年賀状”は毎年11, 12月ごろに高い出現確率を示すが, その他の月ではほとんど出現していない。“ボーダフォン”は2006年は比較的出現確率が安定して推移しているが, 2006年10月ごろから一気に出現確率が低下し, 2007年以降はほとんど出現しない語となっている。“湿る”は毎年6から8月にかけて出現確率が上昇するが, “年賀状”よりも緩やかな変化となっている。“流出”は, 2006年5月に一度出現確率が急上昇しているが, その他の期間はほとんど出現しない語である。表3.3は変動係数での順位付後, バースト抽出により周期的なバーストが発生している単語のみを抽出した結果である。表3.2から“ボーダフォン”, “流出”が表外となり, “除”^{*5}, “4月”が新たに上位で抽出されている。図3.6に, C.V + Burst法により抽出された単語の出現確率の推移を示す。“湿る”と“除”はほぼ同じ確率推移となっている。また, “4月”は緩やかながら毎年3月から5月に出現確率が上昇している。

変動係数とバースト抽出による単語抽出手法(C.V+Burst)を他のカテゴリにおいても適用した。ここでは, 旅行カテゴリと恋愛カテゴリにおいて抽出できた単語を示す。表3.4は旅行カテゴリにおいて抽出できた単語, 図3.7は, それらの出現確率推移である。“2月”, “4月”という特定の月を示す単語の他, “雪”, “GW”といった語が出現している。出現確率推移では, 期間はズレているが, “年賀状”と同様, 出現確率が急激に上昇するという傾向を示している。恋愛カテゴリでの単語抽出結果を表3.5, 出現確率推移を図3.8に示す。抽出できた単語は, バレンタインデーに関する単語であることがわかる。抽出できた全ての単語が毎年1～3月に出現確率が上昇している。

^{*5} “除く”, “除湿”などの単語の一部が抽出されたものであると考えられる

表 3.1 出現確率平均上位の単語 (Freq)

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

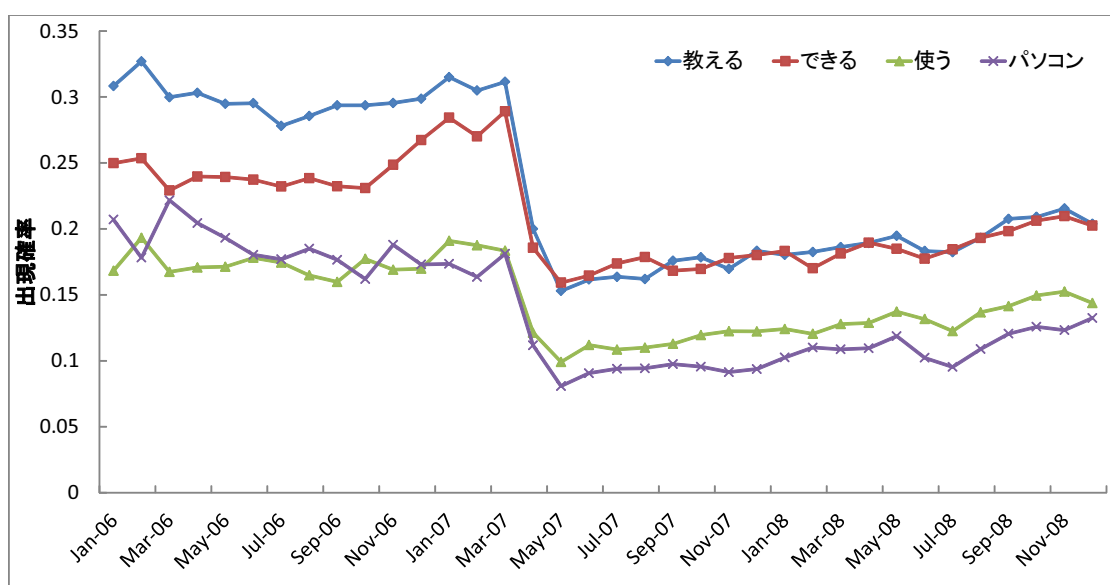


図 3.4 Freq により抽出された単語の出現確率推移

表 3.2 変動係数上位の単語 (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

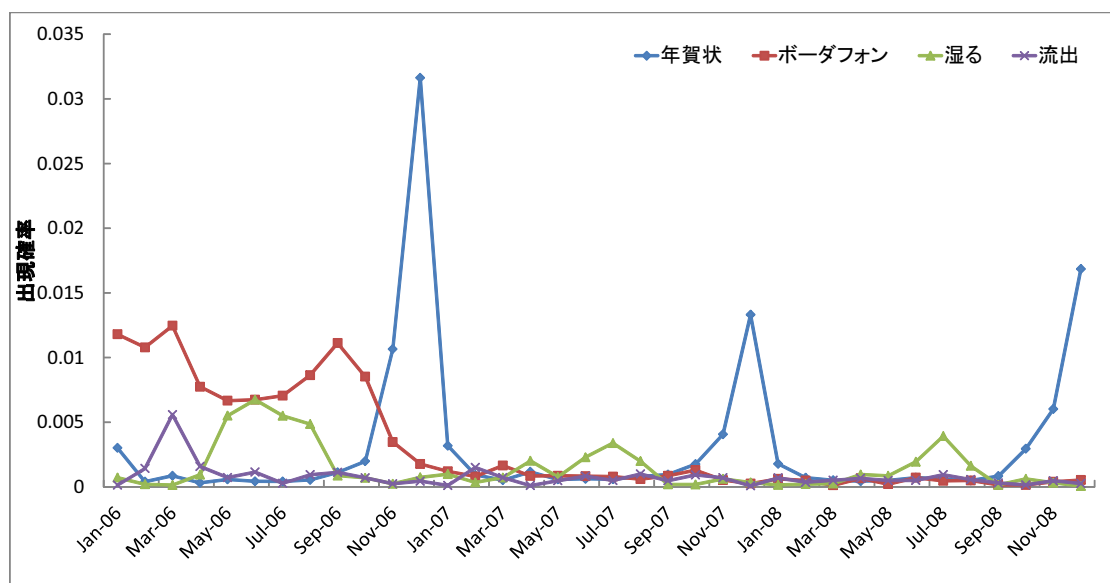


図 3.5 C.V により抽出された単語の出現確率推移

表 3.3 変動係数とバースト抽出による単語 (C.V+Burst)

順位	単語	変動係数	バースト発生月
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	春	0.98	3 月, 4 月

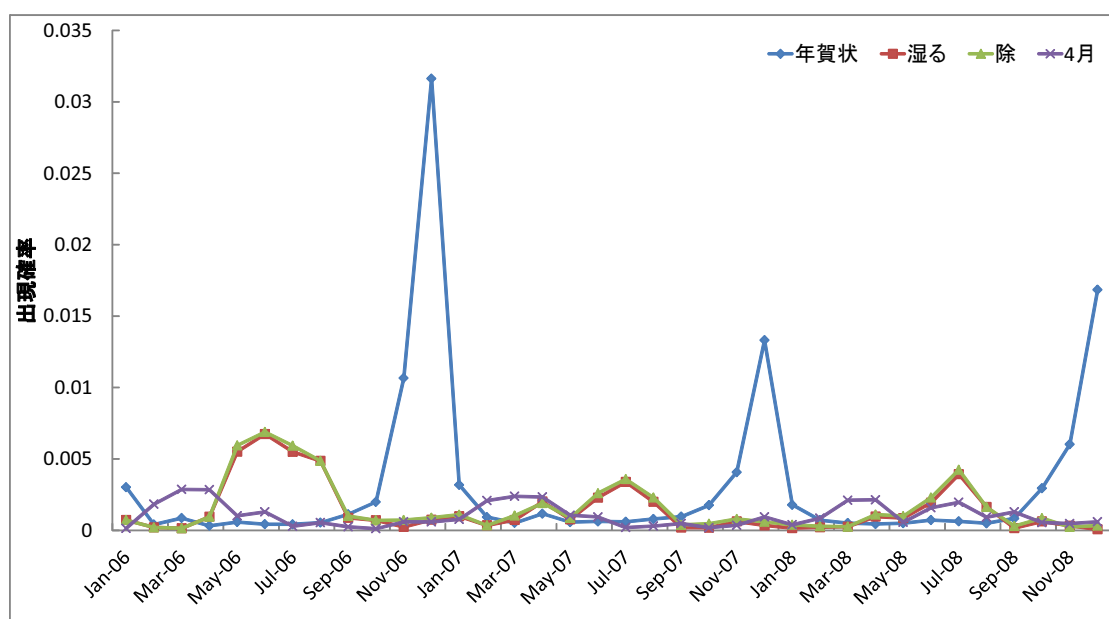


図 3.6 C.V+Burst により抽出された単語の出現確率推移

表 3.4 旅行カテゴリでの単語抽出結果

順位	単語	バースト発生月
1	2月	1月, 2月
2	GW	4月
3	雪	1月, 2月
4	4月	3月, 4月

表 3.5 恋愛カテゴリでの単語抽出結果

順位	単語	バースト発生月
1	チョコ	2月
2	バレンタイン	1月, 2月
3	お返し	2月, 3月
4	義理	2月

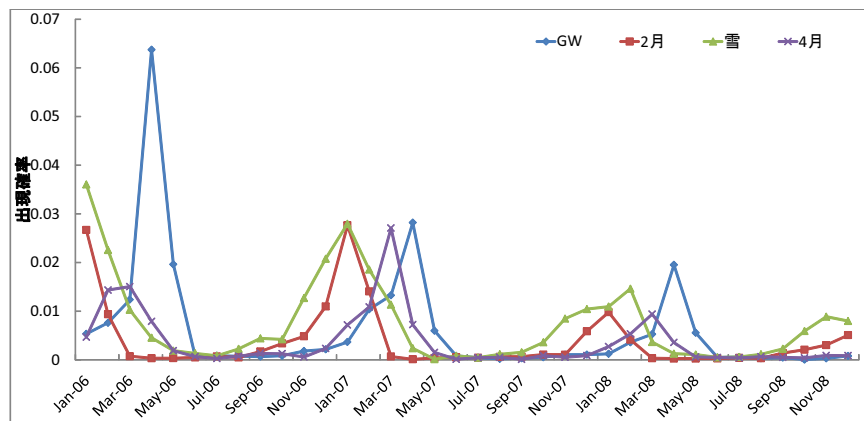


図 3.7 旅行カテゴリでの単語の出現確率推移

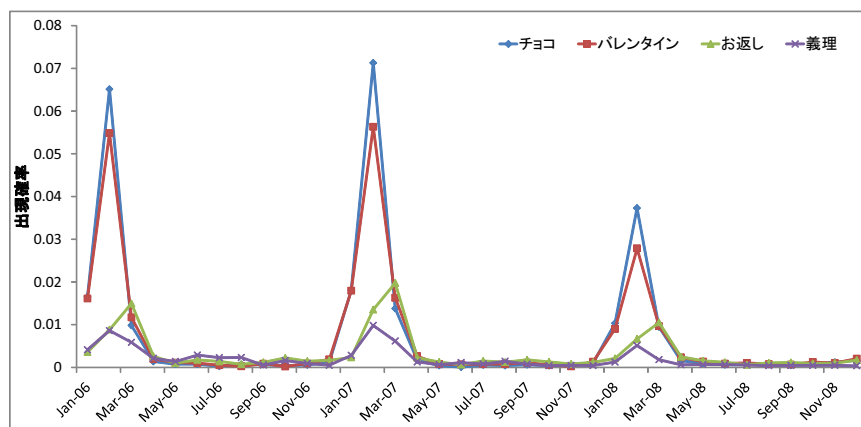


図 3.8 恋愛カテゴリでの単語の出現確率推移

3.3 トピックモデルによる CQA の話題変動抽出

3.2 節では、質問記事を bag of words として見て、bag of words 内の一つの単語のみに注目することで CQA から周期的な話題に関する単語を抽出した。しかし、特定の話題は一つの単語で決定するとは限らない。図 3.1 では、“Christmas” というクエリは、毎年 12 月に多く使用されるクエリであることを示しており、“Christmas” は季節性を持つといえる。また、図 3.2 の “Christmas Movies” も同様に、毎年 12 月に多く利用されるクエリであるが、その他の月ではほとんど使用されておらず、季節性の強いクエリであることがわかる。しかしながら、“Christmas Movies” の一つである図 3.3 の “The Nightmare Before Christmas”^{*6} というクエリはこれらの傾向と異なり、1 年間で 10 月と 12 月の 2 回ピークが存在し、その他の月も比較的多く使用されている、比較的周期性が弱いクエリであることがわかる。このように、季節性を持つキーワードが使用されるクエリであっても、キーワードの組み合わせによっては、季節性が弱いクエリとなる場合が存在する。

CQA においても同様の傾向があり、PC カテゴリでは、“年賀状” という単語が毎年 11、12 月に集中して投稿されることが明らかになった。一方で、“プリンタ” に関する話題は年間を通して数多くの質問記事が投稿されている。しかし、“プリンタ 年賀状” に関する質問記事は、毎年 11、12 月に集中して投稿される。特に、CQA の質問記事は自然言語で記述されているため、Web クエリと違い一つの質問記事に多くの単語が出現する。以上のことより、CQA におけるユーザの情報要求の変化を的確に捉えるためには、単一キーワードに限定せず、キーワード集合からなる“話題”単位で分析することが有用であると考えられる。

本節では、CQA でのユーザの情報要求の変化をより詳細に分析するために、時系列トピックモデルを用いる。トピックモデルは、文書と単語の間には潜在的なトピックがあると仮定するモデルであり、トピックは単語の出現確率分布で表現される。このため、同じ話題で使用される単語は、確率分布内で、近い確率を持つようになるという特徴がある。トピックモデルによって生成される“トピック”をユーザの情報要求の“話題”として扱うことで CQA の質問記事を話題単位で分析することができると考える。

以降は、まず 3.3.1 節で、時系列トピックモデルによる CQA 質問記事の“話題化”と話題変動を追跡するための手法について説明し、3.3.2 で話題変動の例を示す。そして、3.3.3 節で、周波数解析を用いることによる CQA からの周期性を持つ話題の抽出手法について説明し、3.3.4 節で評価実験について詳述する。

^{*6} 1993 年に公開された映画のタイトル

3.3.1 時系列トピックモデルによる話題変動抽出手法

時系列トピックモデル

CQA の話題は、カテゴリにより分類されているが、各カテゴリ内では、より詳細かつ多様な話題が展開されている。そこで、カテゴリ内の話題をより詳細に分析するため、話題の基本単位となるトピックを、トピックモデルによって生成する。トピックモデルでは、文書集合によって作成された文書 - 単語空間を、文書 - トピック空間、トピック - 単語空間に分割する。各トピックは単語の確率分布 (単語分布) で表現されている。本研究では、Blei ら [5] が提案した、Dynamic Topic Model(DTM) を用いる。DTM のグラフィカルモデルを図 3.9 に示す。 z はトピック、 w は単語を表し、 K はトピック数、 N は単語数、 D は文書数である。 α はハイパーパラメータ、 θ は (トピック数 \times 文書数) のトピック比率行列、 ϕ は (単語数 \times トピック数) の単語分布行列である。DTM では、モデル生成に時間情報を用いるため、同一トピックを時間を超えて追跡できるという特徴がある。図 3.9 では、時間分割数 3 の場合を示しており、トピックの単語分布 ϕ が初期状態から 2 回遷移している。時間分割数を TS とした場合、トピックは $\phi_0, \phi_1, \dots, \phi_{TS-1}$ までの TS 個の単語分布を持つことになる。

DTM の各トピックでは、同一の話題で使用する単語が近い確率を持つ。そのため、各トピックの単語分布の遷移を追跡することで、同一話題の変化を調べる。

JS ダイバージェンスによるトピックの話題変動抽出

DTM により生成したトピックの話題遷移を追跡するために、JS ダイバージェンスを用いる。JS ダイバージェンスは、2 つの確率分布同士の類似度を算出する手法であり、確率分布 P と Q から次式で与えられる。

$$JS(P||Q) = \frac{1}{2} \left(\sum_x P(x) \log \frac{P(x)}{R(x)} + \sum_x Q(x) \log \frac{Q(x)}{R(x)} \right)$$

R は確率分布 P と Q の平均であり、 $R = \frac{P+Q}{2}$ である。JS ダイバージェンスは $0 \leq JS$ の値を取り、同じ確率分布 ($P = Q$) のとき、 $JS = 0$ をとる。

DTM では、同一トピック内で、時刻毎に確率分布を生成する。時刻間の確率分布の類似度は、時刻間でどれだけ話題が変化したかを表す、話題の変化量と定義することができる。時刻間の確率分布の類似度が高い場合は、話題がほとんど変化していないことを表し、類似度が低い場合、時間の遷移によって話題が大きく変化したことを示している。

時間分割数 TS の時系列データに対して、時系列の最初の時刻を基準時刻 t_0 とし、 t_0 と $t_0, t_1, \dots, t_{TS-1}$ までの、全ての時刻との JS ダイバージェンスを計算する。計算した JS ダイバージェンス値を時系列上に並べることで、トピック内の話題が時刻毎にどれだけ変化しているのかを分析する。

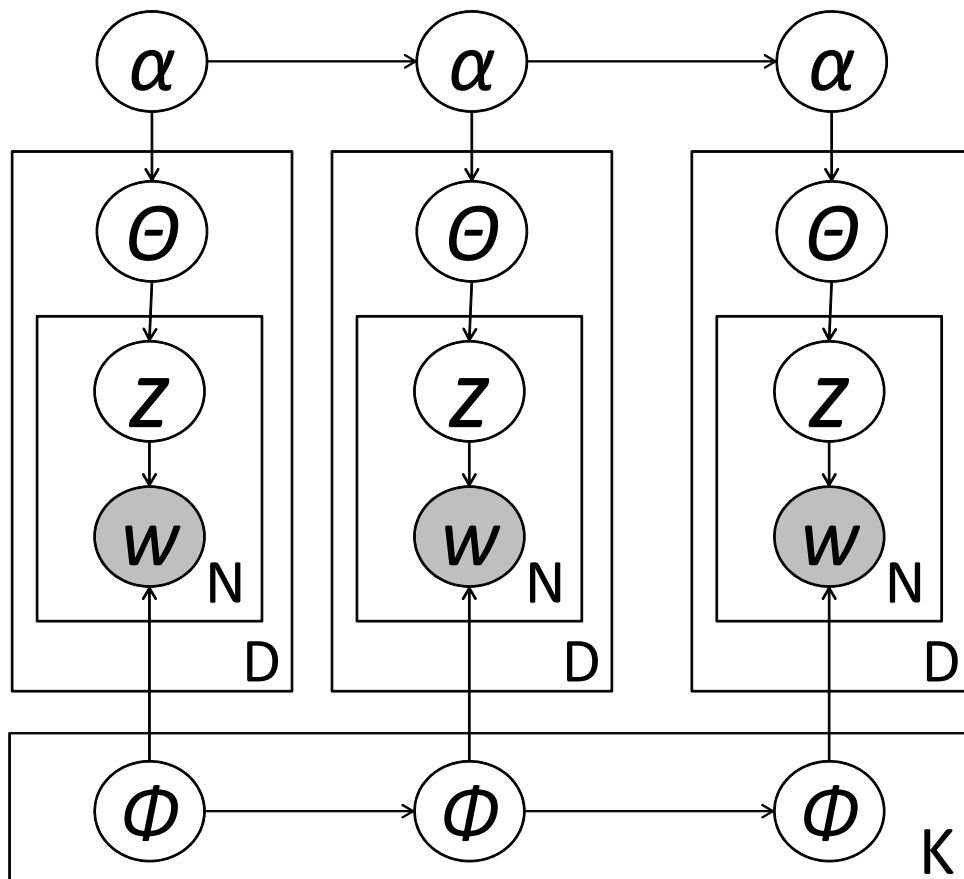


図 3.9 DTM のグラフィカルモデル (時間分割数 3 のとき)

3.3.2 CQA の話題変動抽出結果

3.3.1 節で提案した，話題追跡手法を実際の CQA に適用した結果を示す．実験のデータとして Yahoo!知恵袋データを用いた．期間は 2006 年 1 月から 2008 年 12 月までの 36 ヶ月である．実験に使用したカテゴリは，旅行，PC，経済の 3 カテゴリである．DTM のパラメータは以下の通り設定した．

- トピック数 $K = 200$
- ハイパーパラメータ $\alpha = 0.01$
- 時間分割数 $TS = 36$ (一ヶ月刻み)

実験によって抽出できた話題変動の例をカテゴリ毎に図 3.10 から図 3.12 に示す．図 3.10

はPCカテゴリでの話題遷移，図 3.11 は旅行カテゴリでの話題遷移，図 3.12 が経済カテゴリでの話題遷移を示している．各図の横軸は 2006 年 1 月から 2008 年 12 月までの 36 ヶ月間の時系列を表しており，縦軸は JS ダイバージェンス値である．各グラフは 2006 年 1 月の確率分布との類似度であるため，基準時刻である 2006 年 1 月は全トピックの JS ダイバージェンス値が 0 であり，JS ダイバージェンス値の上昇は 2006 年 1 月から話題が遠くなっていることを示している．各カテゴリのいずれのトピックにおいても，JS ダイバージェンス値は徐々に上昇していることがわかる．しかしながら，JS ダイバージェンス値の上昇のパターンはトピック毎に異なっている．

以下では，カテゴリ毎に話題変動の特徴について，詳細に分析する．

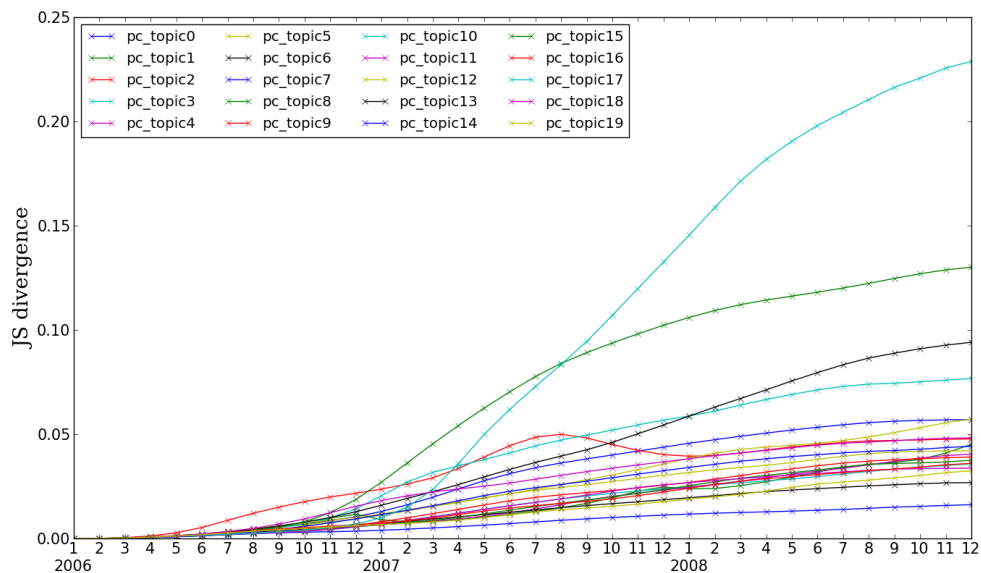


図 3.10 PC カテゴリ全トピックの JS ダイバージェンス (2006 年 1 月基準)

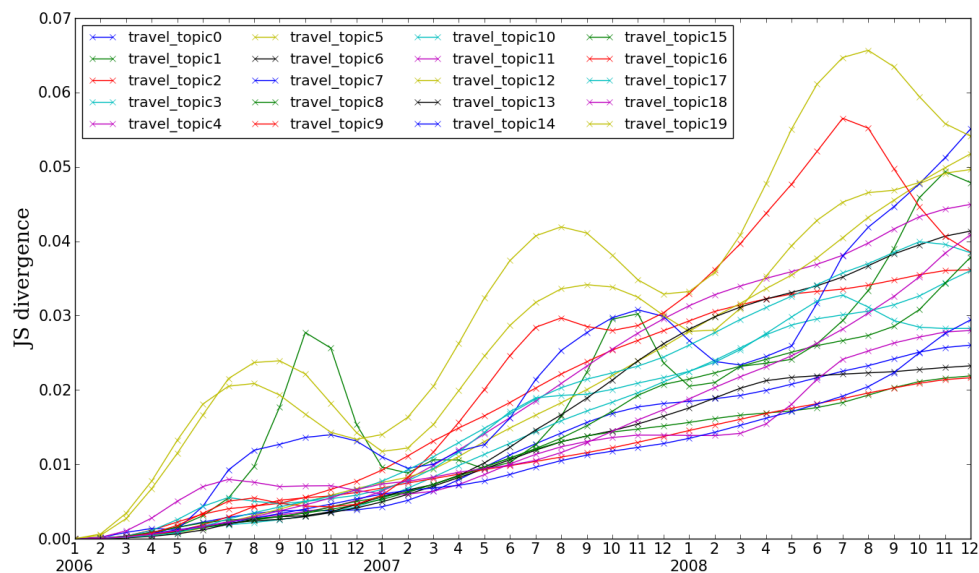


図 3.11 旅行カテゴリ全トピックの JS ダイバージェンス (2006 年 1 月基準)

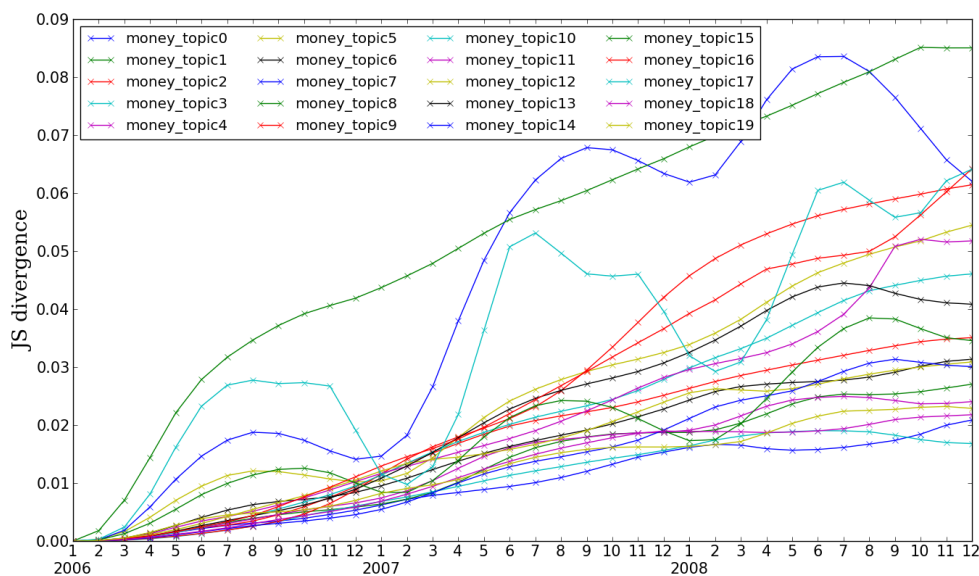


図 3.12 経済カテゴリ全トピックの JS ダイバージェンス (2006 年 1 月基準)

PC カテゴリでの話題変動

PC カテゴリでは，ほとんどのトピックが JS ダイバージェンス値が単調増加のグラフとなっており，JS ダイバージェンスが減少する期間は見られない．図 3.10 の中から，トピック 3 を抜き出したものを図 3.13 に示す．毎月一定の割合で JS ダイバージェンスが増加している．トピックの話題変動の内容を調査するため，各トピックで，時系列上での確率の変動が大きい単語を抽出した結果を表 3.6 に示す．特徴語はトピック全期間の平均確率の上位 5 語，変動語は変動係数 CV の上位 5 語を示している．トピック 3 はパソコン購入に関するトピックであることがわかる．変動語には，“eeepc”，“ミニノートパソコン”などネットブックにカテゴリライズされるパソコンのジャンル名が抽出されている．変動語のトピック 3 内での確率変動の推移を図 3.14 に示す．いずれの語も 2006 年 1 月の時点では，出現確率はほぼ 0 であるが，“ゲーム”，“モニタ”は 2007 年ごろから，“eeepc”，“ミニノートパソコン”，“aspire”は 2008 年ごろから出現確率が上昇している．

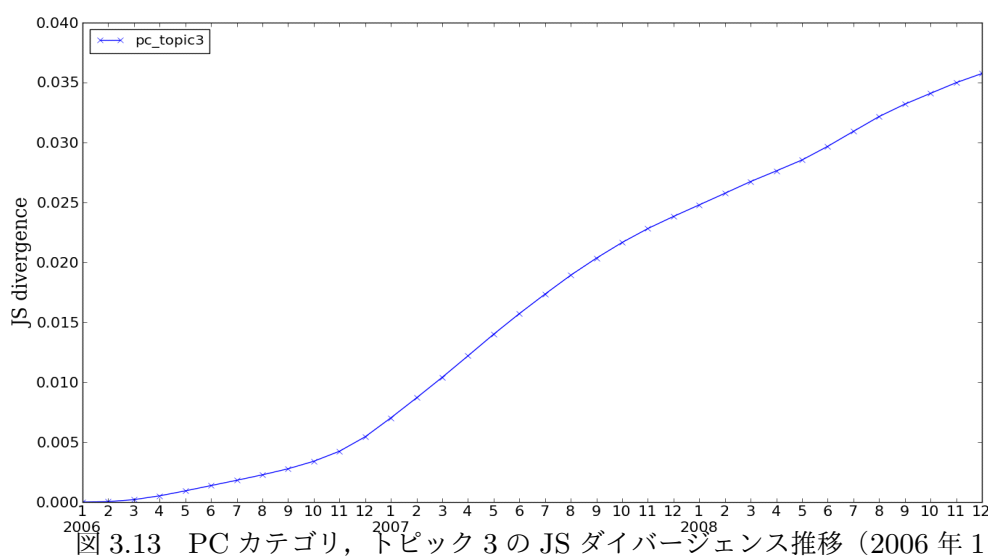


図 3.13 PC カテゴリ，トピック 3 の JS ダイバージェンス推移（2006 年 1 月基準）

表 3.6 PC カテゴリトピック 3 の特徴語と変動語

トピック	特徴語	変動語
3	パソコン 買う ノートパソコン 初心者 新しい	eeepc ゲーム ミニノートパソコン aspire モニター

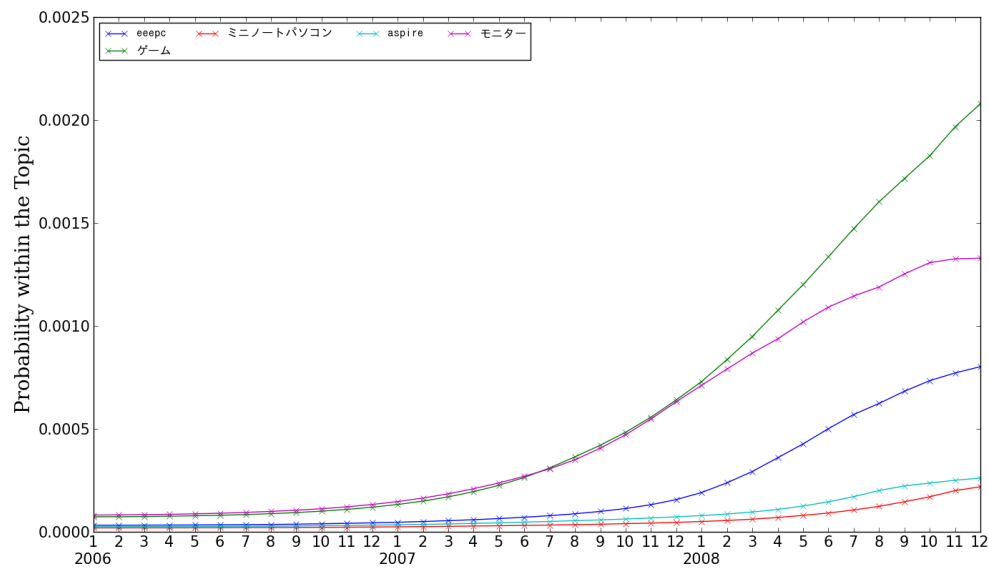


図 3.14 PC カテゴリ，トピック 3 内の単語の確率変動

PC カテゴリでは，ある一時期のみ JS ダイバージェンス値が急上昇する話題変動パターンを発見できた．この話題変動パターンとなった PC カテゴリトピック 2 の JS ダイバージェンス値の推移を図 3.15 に示す．この話題変動では，2007 年 8 月ごろに JS ダイバージェンス値がピークになり，その後は一旦減少した後，緩やかな上昇となっている．トピック 2 の内容語と変動語を表 3.7 に示す．内容語には，“cd”，“音楽”，“ipod” などパソコンで音楽を扱うための用語が抽出されており，変動語には“ニコニコ動画”，“初音ミク”などの語が抽出できた．変動語のトピック 2 内での確率推移を図 3.16 に示す．また，図中に実際に発生した出来事を付与している．“ニコニコ動画”は，ニコニコ動画の実際のサービス開始から発展の過程に対応して出現確率が上昇している．“ニコニコ動画”や“初音ミク”という語において，確率分布の推移とトピック 2 の JS ダイバージェンス推移が対応していることがわかる．

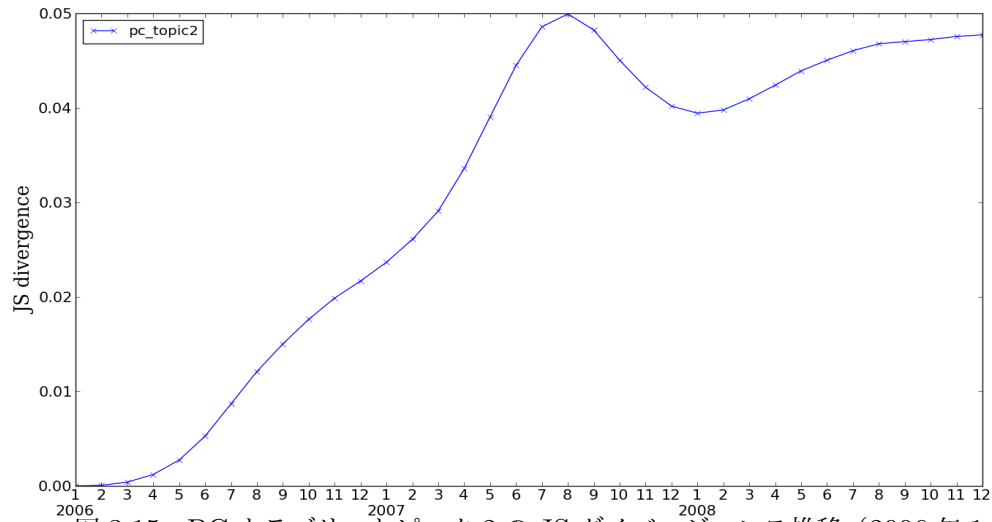


図 3.15 PC カテゴリ, トピック 2 の JS ダイバージェンス推移 (2006 年 1 月基準)

表 3.7 PC カテゴリトピック 3 の特徴語と変動語

トピック	特徴語	変動語
2	cd, 音楽, ipod, mp, ニコニコ動画	iphone, 初音ミク, ボーカロイド, ニコ ニコ動画, dsi

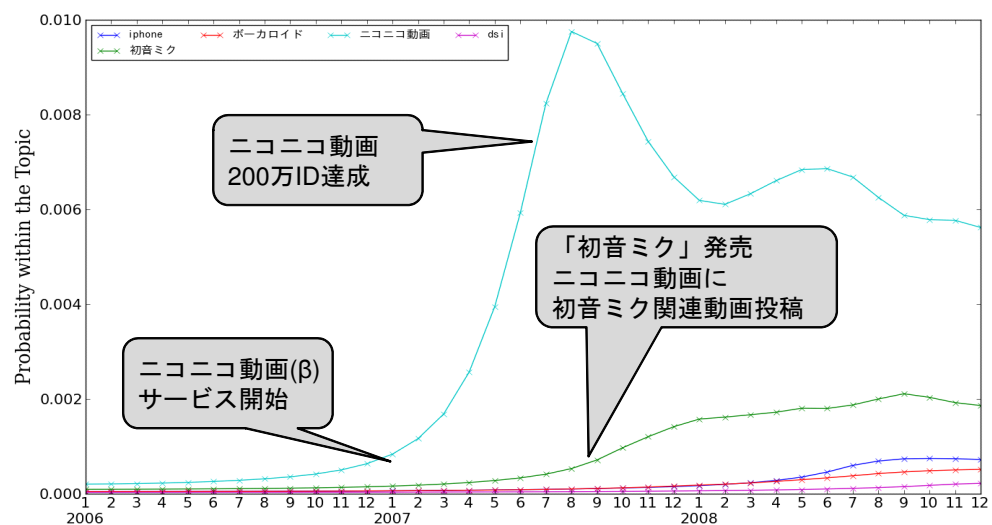


図 3.16 PC カテゴリ, トピック 2 内の単語の確率変動

旅行カテゴリでの話題変動

旅行カテゴリでは、PC カテゴリに見られない話題変動が確認できた。旅行カテゴリトピック1のJSダイバージェンスの推移を図3.17に示す。図から、トピック1においても、JSダイバージェンス値は時間の経過とともに上昇していることがわかる。しかし、JSダイバージェンス値の上昇は単調増加ではなく、増減を繰り返しながら徐々に増加している。JSダイバージェンス値は話題の近さを表しているため、JSダイバージェンス値の増減は、2006年1月と近い話題と遠い話題が交互に発生していることを示している。旅行カテゴリトピック1の内容語と変動語を図3.8に示す。内容語では“京都”，“修学旅行”などの語が抽出されている。変動語では，“紅葉”，“桜”など季節に関連する語が抽出されている。変動語の確率推移を図3.18に示す。“紅葉”は毎年9～11月，“桜”は3，4月に出現確率が上昇するが、他の季節では低い出現確率になっていることがわかる。

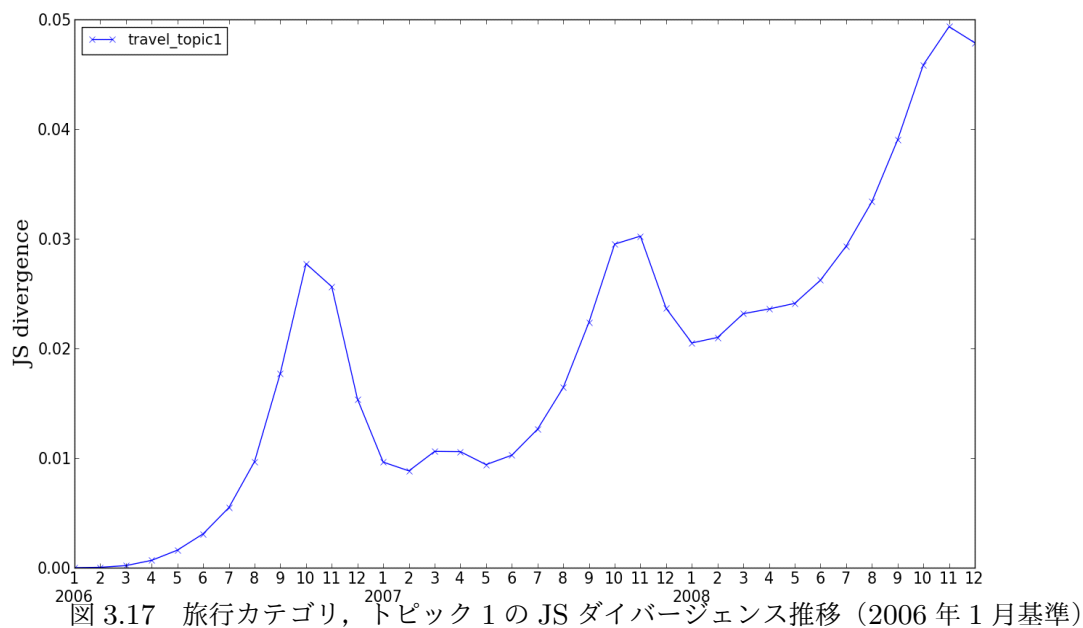


図 3.17 旅行カテゴリ，トピック1のJSダイバージェンス推移（2006年1月基準）

表 3.8 旅行カテゴリトピック 1 の内容語と変動語

トピック	特徴語	変動語
1	京都 見る 修学旅行 桜	紅葉 桜 永観堂 美しい 景色

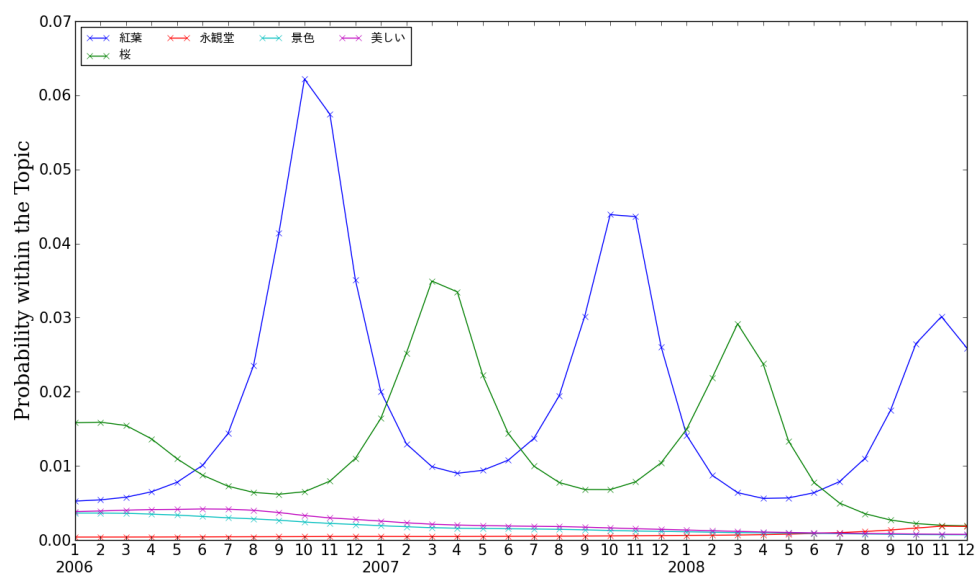


図 3.18 旅行カテゴリ，トピック 1 の単語の確率変動

3.3.3 周波数解析による周期的なトピックの抽出

時系列トピックモデルと、JS ダイバージェンスによりトピックの話題変動パターンを時系列データとして抽出できた。本節では、得られた話題変動パターンから図 3.17 のように周期的な話題変動を発生させるトピックを抽出する手法を説明する。得られた時系列データが周期的な変動を含むデータであることを分析するため、離散フーリエ変換による周波数分析を行う。フーリエ変換は、時系領域で表現されているデータを周波数領域の特徴量に変換するための手法である。 n_0, n_1, \dots, n_{N-1} までの N 個のデータによって表現される時系列データ x から、離散フーリエ変換によって得られる周波数 k のスペクトルは以下で与えられる。

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j \frac{2\pi}{N} kn} \quad (3.3)$$

ここで、指数関数部分はオイラーの公式を用いて実数部と虚数部に分割でき、以下の式に変換できる。

$$\begin{aligned} X[k] &= \sum_{n=0}^{N-1} \left\{ (x[n] \cos \frac{2\pi}{N} kn) - j(x[n] \sin \frac{2\pi}{N} kn) \right\} \\ &= Re[k] + Im[k] \end{aligned} \quad (3.4)$$

各周波数スペクトルの信号の強さは、パワースペクトルを計算することで求めることができる。周波数 k の周波数スペクトル $X[k]$ のパワースペクトル $S(X[k])$ は以下の式により計算できる。

$$S(X[k]) = \sqrt{Re[k]^2 + Im[k]^2} \quad (3.5)$$

パワースペクトルにより、入力した時系列データにどの周波数成分が多く含まれているのかを調べることができる。周波数 k のとき、周期 T は以下の式で与えられる。

$$T = \frac{N}{k} \quad (3.6)$$

離散フーリエ変換により得た、話題変動の周波数領域のパワースペクトル分布により、話題変動を周期的な特徴に基づいて議論することができる。そこで、パワースペクトル分布を素性としてトピックのクラスタリングを行うことで、質問記事の話題を変動のタイプに基づいて分類する。離散フーリエ変換では、サンプリング定理により、周波数 $k = 0, 1, \dots, \frac{N}{2} - 1$ までの $N/2$ 個のパワースペクトルを得ることができる。しかし、 $k = 0$ は周期 $T = \infty$ となり、周期

性に関係しないため除外し, $k = 1, \dots, \frac{N}{2} - 1$ までのスペクトル $S(X[k])$ をクラスタリングの素性とする.

階層的クラスタリングは, 各要素間の距離行列を使用する. ここで, パワースペクトルを正規化し, パワースペクトル分布を確率分布として表現する. パワースペクトルの正規化は以下の式により実行する.

$$P(X[k]) = \frac{S(X[k])}{\sum_{i=1}^{(N/2)-1} S(X[i])} \quad (3.7)$$

正規化により $\sum P(X[k]) = 1$ を満たすので, トピック間の距離を, 式の JS ダイバージェンスによって計算する.

階層的クラスタリングにより, トピックを話題変動の周期的なタイプに基づいて分類する. パワースペクトルの確率分布は一般には指数分布に従うため, 類似度計算を行った場合, どのトピック間も比較的距離が近い分布となってしまう. そのため, クラスタリングのアルゴリズムは, 高い分解能を持つ最長距離法を使用する. 最長距離法はクラスタ間の距離を最長となるように要素対を選択していく手法であり, クラスタ A と B の距離は以下の式で定義される.

$$D(A, B) = \max_{a_i \in A, b_j \in B} d(a_i, b_j) \quad (3.8)$$

3.3.4 CQA からの周期的なトピックの抽出実験

Yahoo!知恵袋の旅行カテゴリに、提案手法を適用し、話題変動のパワースペクトル分布をクラスタリングする。データセットは 3.3.2 節と同様とし、DTM のトピック数を 200 に設定して実験を行った。

結果を図 3.19 に示す。また、各クラスタに割り当てられてトピックの詳細を表 3.9 に示す。旅行カテゴリでは、特徴的な 3 つのクラスタに分離された。まず、中央に 183 個のトピックからなる巨大なクラスタ、その右側に 16 個のトピックからなる小規模なクラスタ、そして、最も左側にトピック 15 のみからなるクラスタが存在している。

以下では、それぞれのクラスタについて、特徴的なトピックを例にしてクラスタの特徴を明らかにする。

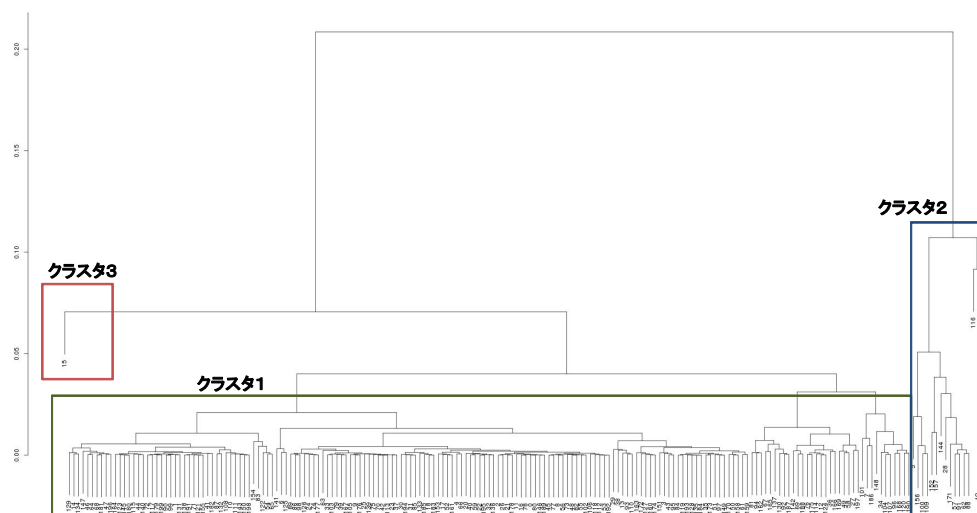


図 3.19 旅行カテゴリ 200 トピックの周波数解析によるクラスタリング結果

表 3.9 旅行カテゴリでの階層的クラスタリング結果（分割数 3）

クラスタ	トピック数	トピック番号
1	183	下記以外の全トピック
2	16	5, 19, 28, 57, 60, 62, 88, 91, 100, 109, 144, 152, 156, 157, 171
3	1	15

単調変化型トピックの周波数解析結果

旅行カテゴリでクラスタ1に属するトピック0の話題変動の結果を図3.20に示す。横軸は2006年1月から2008年12月までの時系列、縦軸はJSダイバージェンス値である。2006年1月の確率分布を基準としているため、2006年1月のJSダイバージェンス値は0である。その後、時間とともにJSダイバージェンスが単調に増加している。これは、2006年1月の確率分布から徐々に確率分布が変化していることを示している。図3.20の話題変動グラフを離散フーリエ変換により周波数のパワースペクトル分布で表現した結果を図3.21に示す。横軸は、 $k = 0, 1, \dots, 17$ までの周波数、縦軸はパワースペクトルのエネルギーである。スペクトルが $k = 0$ から指数関数に従う分布になっている。

トピック0において各月の確率分布の類似度により、クラスタリングした結果を図3.22に示す。各要素は、トピック0内の月を表しており、括弧内は年を表す。“1(2006)”は2006年1月の確率分布を表している。どの月においても自身と近い時期でクラスタを形成している。また、同じ月（1月）であっても、年（2006, 2007, 2008年）が異なっている場合、それぞれ遠いクラスタに属している状態になっている。最後に、旅行カテゴリのトピック0において出現確率の高い単語の例を表3.10に示す。“群馬”、“栃木”など北関東に関連する単語が上位に出現している。

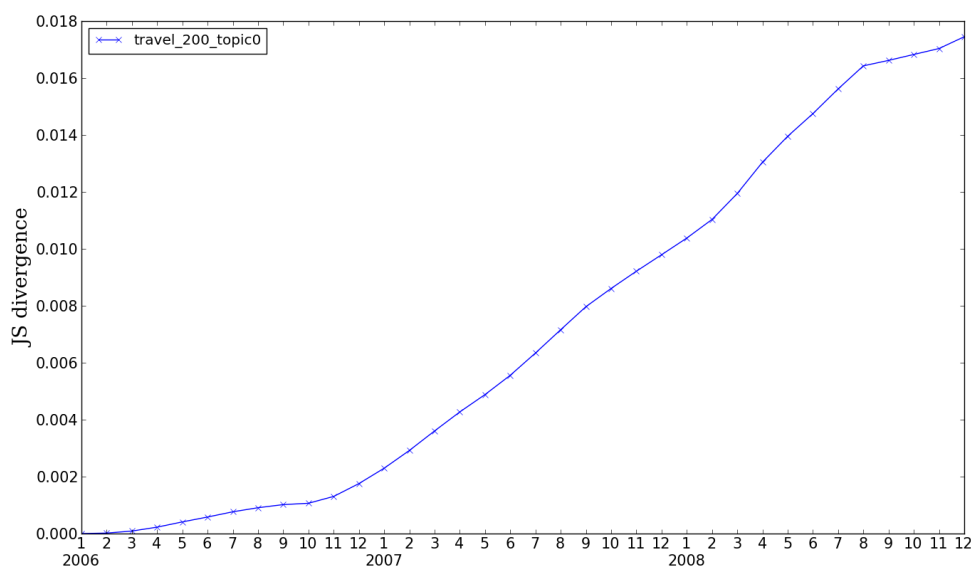


図 3.20 旅行カテゴリ、トピック0の話題遷移

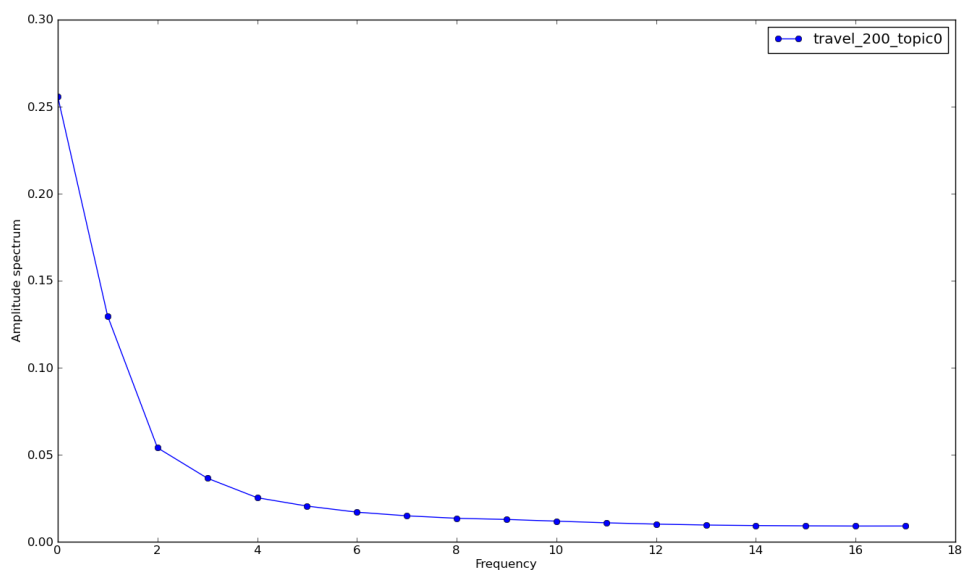


図 3.21 旅行カテゴリ、トピック 0 のパワースペクトル分布

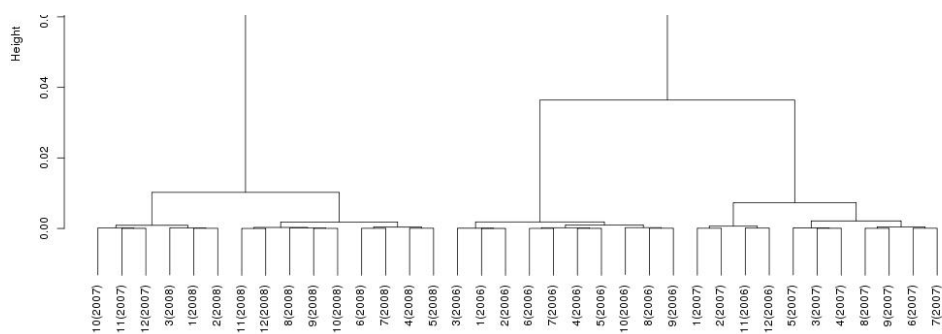


図 3.22 旅行カテゴリ、トピック 0 の各月の単語分布によるクラスタリング結果

周期変化型トピックの周波数解析結果

16 個のトピックから成るクラスタ 2 の例として，トピック 171 の話題変動の結果を図 3.23 に示す．2006 年 1 月から JS ダイバージェンスが徐々に上昇していくが，10 月から低下し，2007 年 2 月まで JS ダイバージェンスが低下した後，再び上昇するパターンを繰り返している．離散フーリエ変換により，図 3.24 のパワースペクトル分布を得る． $k = 3$ の時にスペクトルにピークが立っている．これは，単調変化である図 3.21 とは，大きく異なる特徴である．話題変動は 36 点のデータによって描画されていることから，この話題変動の周期は $T = 36/3 = 12$ である．

トピック 171 の各月の確率分布の類似度から作成したデンドログラムを図 3.25 に示す．2006 年と 2007 年はそれぞれ同じ月と同じクラスタあるいは非常に近いクラスタに属している．トピック 171 において出現確率が上位の単語を表 3.10 に示す．“ヨーロッパ”，“スペイン”など海外に関連する単語の他，“3 月”，“卒業旅行”といった時期を連想する単語が高い出現確率を持っている．

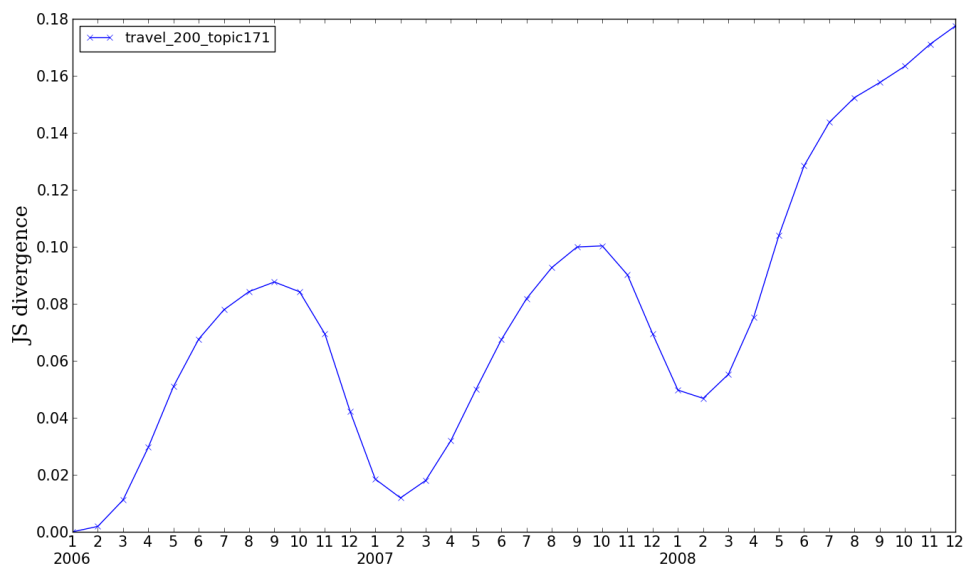


図 3.23 旅行カテゴリ，トピック 171 の話題遷移

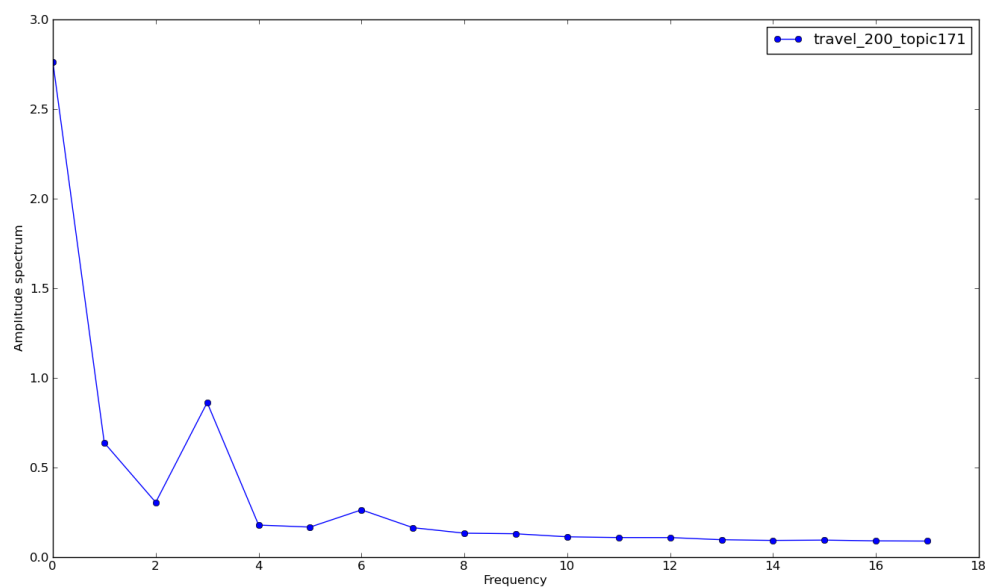


図 3.24 旅行カテゴリ、トピック 171 のパワースペクトル分布

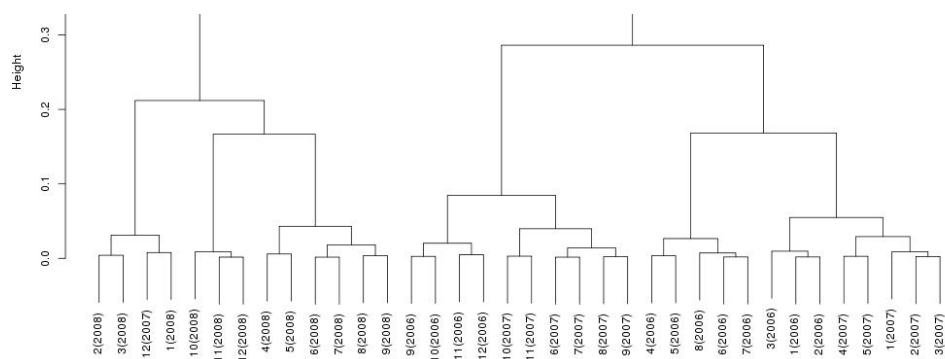


図 3.25 旅行カテゴリ、トピック 171 の各月の単語分布によるクラスタリング結果

バースト型トピックの周波数解析結果

旅行カテゴリ，トピック 15 のみ，他のクラスタから距離のある位置に存在していることがわかる．トピック 15 の話題変動グラフを図 3.26 に示す．2006 年 11 月から 2007 年 4 月にかけて急激な話題変化が見られるが，その後は緩やかな話題変動となっている．トピック 15 の話題変動から離散フーリエ変換により得られたパワースペクトル分布を図 3.27 に示す． $k = 2, 5$ でピークが発生しており，図 3.21，図 3.24 とは異なるパワースペクトル分布となっている．トピック 15 での各月の確率分布の類似度によって作成したデンドログラムは図 3.28 となる．ここで，2006 年 11 月から 2007 年 4 月までの 7 ヶ月は，独自のクラスタを形成していることがわかる．トピック 15 の代表的な単語は表 3.10 に示す通り，“お薦め”，“名所”といった語が上位にきている．

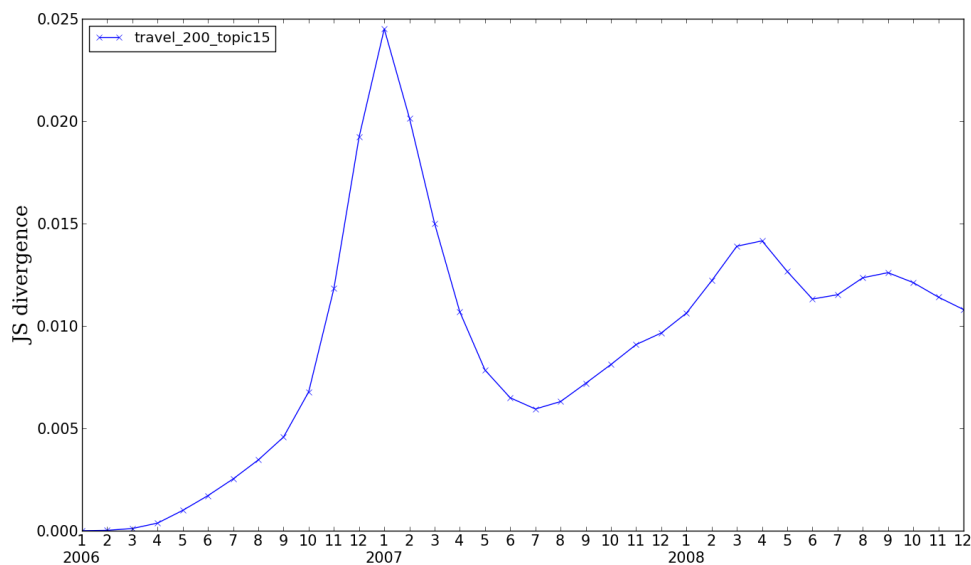


図 3.26 旅行カテゴリ，トピック 15 の話題遷移

表 3.10 旅行カテゴリ各トピックの代表的な単語

トピック	単語
0	群馬，栃木，真ん中，里，湯沢
171	3 月，ヨーロッパ，卒業旅行，スペイン，予定
15	お薦め，名所，新潟，動物園，工事

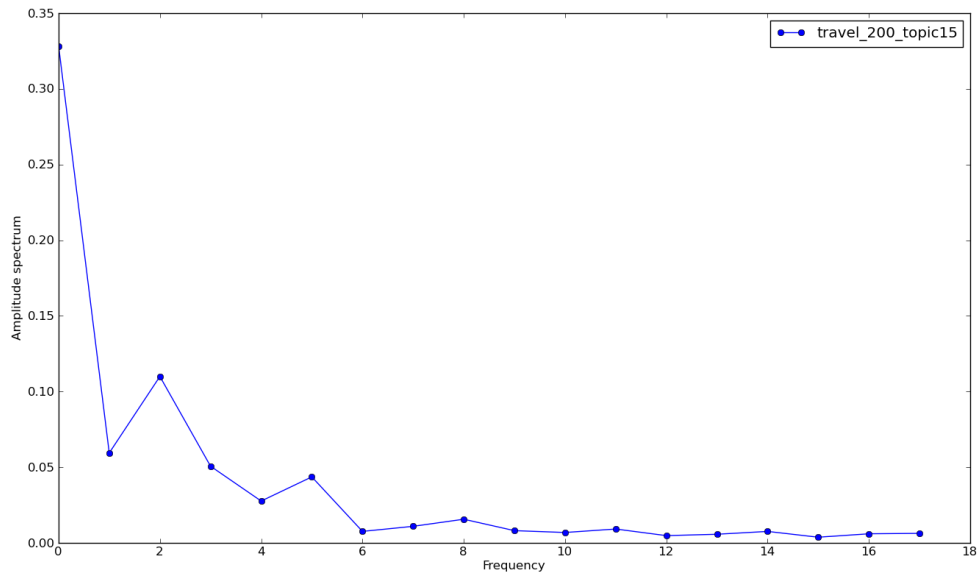


図 3.27 旅行カテゴリ，トピック 15 のパワースペクトル分布

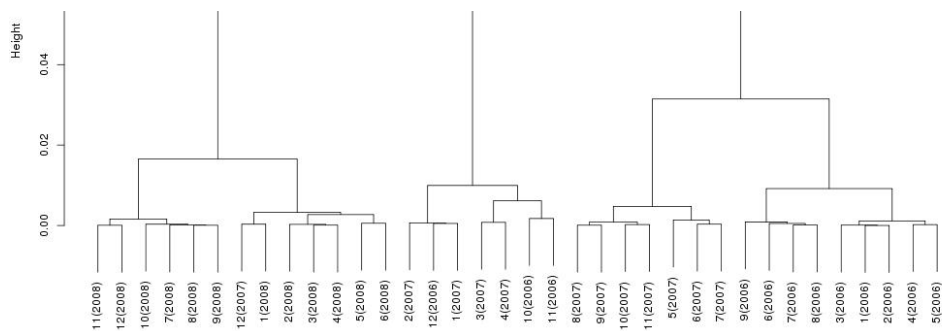


図 3.28 旅行カテゴリ，トピック 15 の各月の単語分布によるクラスタリング結果

その他のカテゴリでのトピックのクラスタ分類結果

“PC”，“経済”，“健康” カテゴリにおいて，旅行カテゴリと同様に階層的クラスタリングを行った．ここで，クラスタ数を3とした時の，各カテゴリで作成できたクラスタ内のトピックと，旅行カテゴリのトピック0, 171, 15 のパワースペクトル分布の類似度を計算し，クラスタの話題変動のタイプを分類した．旅行カテゴリにおいてトピック0が所属するクラスタを“単調変化”クラスタ，トピック171が所属するクラスタを“周期”クラスタ，トピック15が所属するクラスタを“バースト”クラスタとした．分類を行った結果を表3.11に示す．表内の数字は各変動タイプのクラスタに属するトピック数である．スポーツトピックは，3クラスタに分類したところ，2クラスタでトピック171と最も類似度が高くなるトピックが多数となったため，2クラスタが周期型クラスタに分類されている．実験の結果，各話題変動タイプに属するトピックの代表として表3.12に示す単語を抽出できた．

表 3.11 PC，健康，経済カテゴリでのトピック分類結果

カテゴリ	単調型	周期型	バースト型
PC	167	1	32
健康	190	7	3
経済	177	23	0

表 3.12 PC，健康，経済カテゴリでの代表的な単語

カテゴリ	トピック番号	変動タイプ	単語
PC	18	単調変化	mp3, 音楽, 音質
	73	周期	印刷, できる, 年賀状
	105	バースト	ニコニコ動画, ニコ厨
健康	12	単調変化	癌, 細胞, 禁止
	1	周期	寒い, 対策, 外
	169	バースト	下痢, ノロウイルス, 腸
経済	7	単調変化	現金, 振込み, atm
	6	周期	医療費, 入院, 対象

3.4 情報要求を提示するクエリ拡張型 Web 検索システム

本章では、CQA の質問記事を用いた、情報要求の言語化支援を行う Web 検索システムについて説明する。まず始めに、情報要求言語化のための拡張クエリ“質問記事付き拡張クエリ”とタブとタグクラウドを組み合わせたファセット検索型の検索インターフェースを提案する。

実装したシステムのスクリーンショットを図 3.29、システム構成図を図 3.30 に示す。本システムは主に 4 つのブロックで構成されている。

ファセット検索部の検索窓にキーワードが入力されると、キーワードに関連するカテゴリタブ、システムにアクセスした時刻から季節タブが初期状態として設定される。カテゴリ、季節それぞれでタブを選択すると、入力キーワード、選択カテゴリ、季節に関連するキーワード一覧がタグクラウドに表示される。タグクラウド中のキーワードが選択されると、システムは選択されたカテゴリ、時期に投稿された質問記事の中から、入力キーワードと選択キーワードが含まれる質問記事を検索し、CQA クエリブロックに表示する。それぞれの質問記事には、質問記事からのキーワード抽出により作成した Web 検索用のキーワード組が付与されている。

以降は、まずシステムの実装に用いたデータセットについて説明し、次に、ファセット検索ブロックでのタブとタグクラウド、CQA クエリブロックでの CQA クエリの実装方法について詳述する。



図 3.29 クエリ拡張型 Web 検索システムのプロトタイプ画面

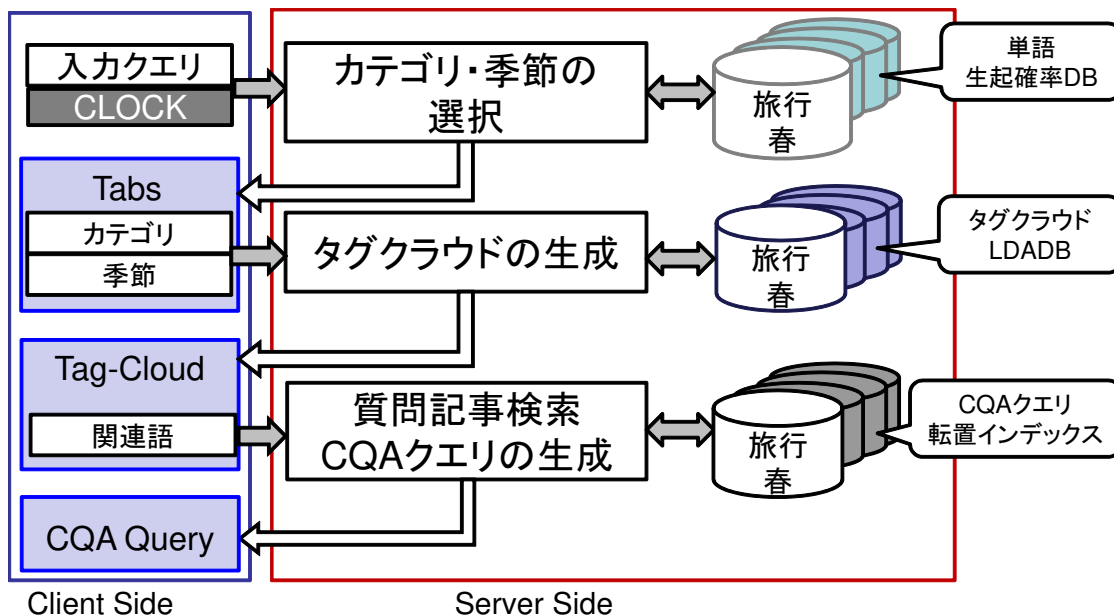


図 3.30 クエリ拡張型 Web 検索システムのシステム構成図

3.4.1 質問記事付き拡張クエリ

本研究では、Web 検索における情報要求の言語化を支援するため、CQA の質問記事と質問記事から作成した拡張クエリをセットで提示する“質問記事付き拡張クエリ（CQA クエリ）”を提案する。CQA クエリの例を図 3.31 に示す。複数のキーワードからなるキーワード拡張クエリには、ユーザにとって未知のキーワードが含まれていることがあるが、CQA クエリでは、質問記事を参照することで、クエリの主題を知ることができるという特徴がある。例では、“ノロウイルス”というキーワードをユーザが知らなかったとしても、質問記事を参照し、“私の友人も感染”などの文章を読むことで、ノロウイルスが人に感染するウイルスであることを把握でき、違和感なく Web 検索でノロウイルスの予防法について調べることができるようになる。このように、自然言語はユーザにとって理解しやすく、疑問が詳細に記述されているため、質問記事を参照することで、拡張クエリから“具体的な情報要求”を表出することができる。

3.4.2 タブとタグクラウドによるファセット検索インターフェース

CQA には大量の質問記事が投稿されている。質問記事は個人的かつ詳細な内容が記述されているため、同じ主題の質問記事であっても、質問記事の細かい内容によって異なる拡張クエ

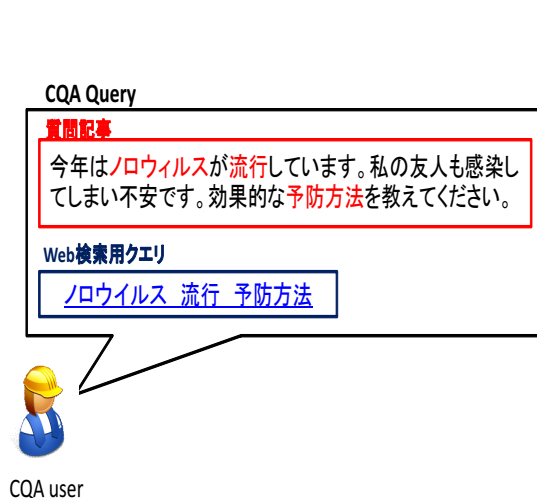


図 3.31 質問記事付き拡張 (CQA クエリ)

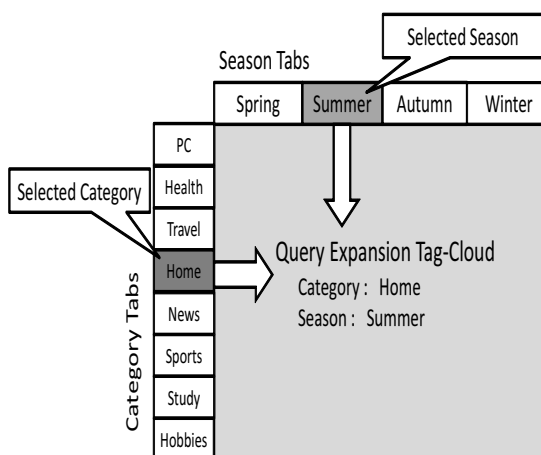


図 3.32 タブ-タグクラウドインターフェース

りが作成される．そのため，大量に生成される CQA クエリの中から，ユーザの興味や状況に近いクエリを提示するための仕組みが必要となる．本研究では，様々な切り口から CQA クエリを提示するファセット検索型のインターフェースを提案する．検索のファセットとして，質問記事が投稿されているカテゴリと投稿時期を使用する．CQA では，質問記事を投稿する際，質問の内容に近い一つのカテゴリを選択する必要がある．そのため，CQA の質問記事はカテゴリによって質問内容の分類がなされていることになる．質問記事の内容が異なれば，そこから作成される拡張クエリのキーワードも異なるため，ファセットとして，カテゴリを切り替えることで，異なる CQA クエリを提示できるようになる．2 つ目のファセットとして，季節を導入する，CQA に投稿される質問記事の話題には周期性があることが，3.3 節での実験の結果分かっている．そこで，質問記事の投稿時期を用いて，質問記事を季節ごとに分類することで，季節ごとに異なる CQA クエリを提示する．

カテゴリと季節の 2 つのファセットを用いた拡張クエリの提示法として，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用いる．インターフェースのイメージを図 3.32 に示す．縦に配置されているのがカテゴリ選択タブ，横に配置されているのが季節タブである．ユーザは，それぞれのタブから興味のあるカテゴリ，季節を選択することでそれらのコンテキストに関連する CQA クエリを入手することができる．インターフェースの中央部は，キーワードのタグクラウドが表示される．タグクラウドには，クエリ拡張で追加されるキーワード一覧が表示される．キーワードが選択されると，入力キーワードと選択キーワードを含む質問記事から作成された CQA クエリが表示される．

以降は，CQA クエリとファセット検索型インターフェースの実装方法と評価について詳述する．

3.4.3 データセット

本システムで使用する CQA コーパスとして、国立情報学研究所の提供する Yahoo!知恵袋^{*7}コーパスを利用する。コーパスには、2004 年から 2008 年の Yahoo!知恵袋のデータが収録されている。ここでは、Yahoo!知恵袋が正式サービスとなった 2006 年から、2008 年までの 3 年分のデータを使用する。Yahoo!知恵袋は数百のカテゴリが 3 階層構造で存在している。すべてのカテゴリ別にデータベースを作成した場合、似たカテゴリの話題が関連カテゴリとして抽出されてしまい、話題の切り替えとして有効に機能しなくなることから、今回はある程度広い話題ごとにカテゴリが分かれるように、複数のカテゴリを集約する。最下層の小カテゴリは、カテゴリによって質問記事数に差があるため、今回はどのカテゴリにおいても比較的多数の質問記事が投稿されてる中階層のカテゴリのデータのみを使用し、中カテゴリの質問記事を階層構造に習った状態で集約していく。例えば、最上位のカテゴリ“インターネット、PC と家電”カテゴリは、“インターネット”、“パソコン”、“AV 機器”、“携帯電話”などのカテゴリを持つが、それらのカテゴリの質問記事を全て一つにまとめて“デジタル・家電”カテゴリという一つのカテゴリとして扱う。また、季節性のために、データセットを期間毎に分割する。2006 年から 2008 年までの各年で同じ月に投稿されたデータを集約する。使用する季節は以下のようになる。

- 春: 各年 3 月から 5 月に投稿されたの質問記事
- 夏: 各年 6 月から 8 月に投稿されたの質問記事
- 秋: 各年 9 月から 11 月に投稿されたの質問記事
- 冬: 各年 12 月から 2 月に投稿されたの質問記事

データセットの一部を表 3.13 に示す。*NavigationCategory* は集約したカテゴリ名でシステムにはこのカテゴリ名が表示される。*Second - LevelCategory* は集約した中カテゴリであり、春～冬はデータセットに含ませる質問記事数である。冬のデータが最も多くなっているが、これは Yahoo!知恵袋が正式サービスを開始した 2006 年 4 月からデータを採集しているためである。作成したシステムでは、*NavigationCategory* を 20 カテゴリ、作成し、それぞれを 4 つの季節に分割したことで、全 80 個のデータセットとなった。

3.4.4 タブによるコンテキスト提示の実装

ユーザがシステムを利用する際、最初に検索クエリとしてキーワードを入力する。また、その際ブラウザ上の時間をサーバに送信する。サーバではユーザが入力したキーワードと時間か

^{*7} <http://chiebukuro.yahoo.co.jp/>

表 3.13 Navigation Category とデータセット例

Navigation Category	Second-Level Categories	春	夏	秋	冬
趣味	アニメ, コミック, 本, おもちゃ, 占い, くじ, 伝統工芸	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	14,9830	16,9239	19,6009	22,0074

ら、関連のあるカテゴリと季節を選択する。サーバ上に保存されているカテゴリと季節の組み合わせからなる 80 個のデータセットから最も関連のあるデータセットを選択する。季節の関連はユーザが入力をシステムに送信した時間から、対応する月の属する季節を選択する。季節は 4 つで固定のため、ここで決定するのは初期状態で選択される季節であり、ユーザはタブ切り替えによって自由に季節を変更できる。カテゴリは、各カテゴリでの入力キーワードの出現確率によって決定する。カテゴリ C において、キーワード w の出現確率 $P_{C,w}$ は以下で計算できる。

$$P_{C,w} = \frac{N_{C,w}}{N_C} \quad (3.9)$$

$N_{C,w}$ はカテゴリ C でのキーワード w が含まれる質問記事数である。 N_C はカテゴリ C での質問記事総数である。出現確率が高いカテゴリは入力キーワードと関連のあるカテゴリであるとみなし、出現確率の降順に最大 10 個のカテゴリをタブに表示する。

3.4.5 タグクラウドの実装

ファセット検索ブロックでタブが選択されると、システムは、ユーザが入力したキーワード、タブにより選択したカテゴリと季節から関連語のタグクラウドを作成する。タグクラウドは、キーワードを表示順と文字サイズによって差別化できるため、情報検索の可視化で幅広く使用されている [26]。本システムでは、タグクラウドの表示順として、入力キーワードに関連

のあるキーワードほど先に表示される。

表示順を決定するための、単語間の関連度を計算する手法を述べる。本実装では、データセットである CQA の特徴を活用し、質問に共に使用される確率が高い語を提示するために統計的手法を用いた確率的トピックモデルを用いたアプローチによる関連度計算を実装する。ここでは、Blei ら [6] が提案した、潜在的ディリクレ配分法 (LDA) を用いる。LDA では、文書-単語行列を潜在的なトピックを用いて、文書-トピック行列、トピック-単語行列に分割する。トピックは単語の確率分布によって定義される。ここで、各トピックでの出現確率の分布が近い単語は潜在的に関連のある単語であるといえる。

LDA では、教師なし学習によって文書-単語行列からトピック集合 Z を推定する必要がある。本実装では、崩壊型ギブスサンプリングを用いた推定を用いる。文書 d , n 番目の単語 $w_{d,n} = v$ のトピック $z_i = k$ の更新式は以下で計算できる。

$$P(z_i = k | Z_{-i}, W) = \frac{N_{k-i}^d + \alpha}{N_{-i}^d + T\alpha} \frac{N_{k-i}^v + \beta}{N_{k-i} + W\beta} \quad (3.10)$$

ここで、 $-i$ は、トピック集合全体から i (d 番目の文書の n 番目の単語) 分を除くことを示す。 N_k^d は、文書 d において、トピック k が割り当てられた回数、 N^d は文書 d において単語が生成された回数、 N_k^v はトピック k において単語 v が出現する回数。 N_k は、トピック k に出現する単語の総数である。 T はトピックの種類数、 W は単語の語彙数である。 α, β はディリクレ分布のハイパーパラメータである。

トピック分布 Z から文書-トピック分布 θ と、トピック-単語分布 ϕ が得られる。文書 d において、トピック k が生成される確率 θ_d^k 、トピック k から単語 w が生成される確率 ϕ_k^w は以下の式で与えられる。

$$\hat{\theta}_d^k = \frac{N_k^d + \alpha}{N^d + T\alpha} \quad (3.11)$$

$$\hat{\phi}_k^w = \frac{N_k^v + \beta}{N_k + W\beta} \quad (3.12)$$

トピックは単語の出現確率分布で表現されているため、データセットに含まれる全ての単語はトピック k での出現確率を持っている。各トピック間で似た出現確率を持つ単語同士は実際の質問記事においても関連して使用される確率が高い語である。そこで、各トピックでの出現確率の類似度を計算することで、単語の類似度を計算できる。各トピックを次元圧縮された行列の基底とみなし、コサイン距離をよって類似度を計算する。

$$\text{sim}(w_1, w_2) = \cos(P_{w_1}, P_{w_2}) = \frac{P_{w_1} \cdot P_{w_2}}{|P_{w_1}| |P_{w_2}|} \quad (3.13)$$

P_{w_1}, P_{w_2} は単語 w_1 と w_2 の各トピックでの出現確率分布である。システムでは類似度を単語間の関連度として、関連度順に関連語を提示する。また、タグクラウドでの文字の大きさは

その関連語を選択した際に検索される CQA クエリの数によって決定する．検索される質問記事数に比例し文字のサイズは大きくなる．LDA による検索の結果，潜在的に関連があっても，質問記事で入力キーワードと同時に使用されていない関連語については今回はタグクラウドに表示しない．

3.4.6 質問記事検索と拡張クエリ作成の実装

システムはタグクラウドから選択されたキーワードを基に関連する質問記事を提示する．また，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．クエリ尤度モデルはクエリ Q が文書 D を発生させる確率 $P(D|Q)$ の最大尤度を発見することにより検索を行う情報検索手法である．この手法では，ベイズの定理により $P(D|Q)$ を以下の式に変形して考える．

$$P(D|Q) = \frac{P(Q|D)P(D)}{P(Q)} \quad (3.14)$$

$P(Q)$ は文書に対して独立な確率を持つため，定数項として除外できる．また $P(D)$ は文書に対して事前知識が必要となるため省略可能であり，最終的に $P(D|Q) \propto P(Q|D)$ として考えることができる． $P(Q|D)$ は文書 D がクエリ Q を表現する確率である．つまり，文書 D に対して尤もらしいキーワードの組み合わせによって作成されるクエリ Q を発見することができる．これは，本実装において，質問記事からキーワード組を抽出し拡張クエリを作成する際に利用可能である．そのため，本実装ではクエリ尤度モデルを用いた質問記事検索を実装する．

クエリ尤度モデルでは $P(Q|D)$ が最も高くなる文書が最もクエリに適合する文書となる． $P(Q|D)$ は複数個の単語の組み合わせ \mathbf{w} によって表現される．

$$P(Q|D) = \prod_{w \in Q} P(w|\theta_D)^{c(w,Q)} \quad (3.15)$$

θ_D は文書 D を構成する言語モデルである．本実装ではユニグラム言語モデルを用いた． $c(w, Q)$ はクエリ Q における単語 w の出現頻度である． $P(w|\theta_D)$ は，言語モデルが単語 w を生成する確率であるが，本実装では，逆頻度要素を持たせるため，ディリクレスムージングを用いた以下の計算を行った．

$$P(w|\theta_D) = \frac{c(w, D)}{|D| + \mu} + \frac{\mu}{|C|(|D| + \mu)} \sum_{D \in C} \frac{c(w, D)}{|D|} \quad (3.16)$$

$c(w, D)$ は文書 D における単語 w の出現頻度である． $|D|$ は文書 D における総単語数， $|C|$ は文書コレクション C における文書総数である． μ はパラメータであり今回は $\mu = 1$ としている．クエリ尤度モデルを用いて，ユーザが入力したキーワードとタグクラウドから選択したキーワードが含まれる文書をランキング化する．また，質問記事 D_{que} から最も特徴的な拡

張クエリ Q_{ex} を作成するため、 $P(Q_{ex}|D_{que})$ が最も高くなる単語を 3 語抽出する。その際、ユーザが入力した単語とタグクラウドから選択した単語は優先的に抽出される。

3.5 評価実験

本節では、実装したシステムに対しての評価実験について説明する。評価実験として、ファセットを切り替えた際に提示されたクエリによって検索できる Web ページの多様性に対して評価を行った。

以降は、多様性評価のための実験手法について説明し、次にカテゴリと多様性評価、季節と多様性評価を説明する。

3.5.1 キーフレーズ抽出による Web 検索の多様性評価

本研究で実装したシステムでは、質問記事の違いにより多様な Web 検索を実現している。多様性の評価では、拡張クエリにおいて検索される Web ページが異なる内容について記述されたものかを調べる必要がある。例えば、異なる URL の Web ページは異なる内容を扱うと仮定し、URL の重複を見ることで多様性を評価する手法 [61] などが知られている。

本研究では、拡張クエリから検索された Web ページからキーフレーズを抽出し、キーフレーズの種類数を比較する。理由は、同様の話題を扱う Web ページは、使用される単語も同様であり、単純に URL の異なりを調べただけでは多様なクエリ拡張が行えていると判断できない場合があると考えたためである。キーフレーズ抽出を行うことで、異なる話題の拡張クエリから検索される Web ページはより多くのキーフレーズを抽出できると考えられ、異なる URL で同様の話題を扱う Web ページが多様性に貢献しないと判断できる。

キーフレーズ抽出による多様性評価実験手法を、既存の検索エンジンのクエリ拡張を用いて実験評価する手順を以下に示す。

1. 作成した拡張クエリを Web 検索 API^{*8}に送信し、Web ページ検索結果を入手
2. 入手した Web ページ検索結果からスニペットを抽出
3. キーフレーズ抽出 API^{*9}を用いて、スニペット毎にキーフレーズを抽出
4. スコアが閾値以上のキーフレーズを選択
5. Web ページ上位 10 件で 2 回以上出現したキーフレーズをカウント

API のスコアは 0~100 に設定されている。API のスコアは相対値であり、必ずスコア 100 のキーフレーズが存在する。そのため、どの Web ページでも最低 1 個はキーフレーズを抽出

^{*8} <http://developer.yahoo.co.jp/webapi/search/>

^{*9} <http://developer.yahoo.co.jp/webapi/jlp/keyphrase/v1/extract.html>

できる。一般語や日付など、内容を表すキーワードに相応しくないキーワードのスコアを調査したところ 30 以下だったため、今回はより特徴的な語に絞ることを含めて、API のスコアは 50 に設定した。1 回しか出現しないキーワードはその Web ページ固有のものである可能性が高いため、複数ページで使用されたものを話題を示すキーワードと設定した。スニペットのみを用いるのは一つの Web ページに複数の話題が記述されている可能性があるため、クエリ周囲の文章のみのスニペットを用いることで、クエリ周辺の話題のみを抽出するためである。

評価手法の妥当性を評価するために、商用の検索エンジンである bing^{*10}、Google^{*11}、Yahoo!JAPAN^{*12} の拡張クエリについて、評価実験を行う。評価は、“ウィルス”と“mac”の2つの入力語で行っている。この語は、複数の意味を持つ単語であり、推薦される拡張クエリに違いが出やすいと判断したためである。bing で入手できる拡張クエリ数が最大 8 件なので、8 個の拡張クエリでのキーワード抽出数を比較する。検索エンジンによって推薦された“ウィルス”の拡張クエリを表 3.14 に、“mac”の拡張クエリを表 3.15 に示す。また、各検索エンジンの拡張クエリによって検索された Web ページから抽出されたキーワード数の一覧を表 3.16 に示す。各検索エンジンでキーワードの抽出数に違いがでていることがわかる。“ウィルス”において、bing ではコンピュータ関連のクエリが 9 個推薦されている。Google も同様に 9 個がウイルスセキュリティ関連のクエリである。Yahoo!JAPAN は 4 個がコンピュータに関連するクエリであるが、他の 3 個は病気関連、1 個はテレビドラマに関する語が提示されている。キーワード数では、Yahoo!JAPAN が最も多くのキーワードを抽出でき、bing が最も抽出できたキーワード数が少ない。“mac”では、bing は飲食店関係が 4 個、その他、アニメ、Web サービス、イベントホール、セキュリティソフトに関するクエリが推薦されている。Google と Yahoo!JAPAN は全てコンピュータ関連のクエリが推薦されているが、Google はコンピュータの mac 製品に関するキーワードが大半なのに対し、Yahoo!JAPAN は“windows”や“dvd”など mac 製品以外に対するキーワードも含まれている点異なる。抽出キーワード数は、bing が最も多く、Yahoo!JAPAN、Google の順で抽出できたキーワード数が減少している。

既存手法との比較によるカテゴリとクエリの多様性に関する評価

複数のカテゴリから拡張クエリを作成することで、より多様な Web 検索を実現できることを、既存の拡張クエリ提供サービスとの比較により評価する。実験では、カテゴリごとにキーワードの関連語を出力するプロトタイプシステム^{*13}を用いて評価を行っている。実験では、

*10 <http://www.bing.com/?cc=jp>

*11 <http://www.google.co.jp/>

*12 <http://www.yahoo.co.jp/>

*13 プロトタイプシステムでのデータセットは 2004 年 4 月から 2005 年 9 月までのデータを用いている

表 3.14 入力語“ウイルス”の拡張クエリ

bing	Google	Yahoo!JAPAN
ウイルスバスター	ウイルスバスター	ウイルスソフト 無料
ウイルスバスター 2011	ウイルス	ウイルスチェック
ウイルスソフト	ウイルスソフト	ウイルスソフト
ウイルスバスター 2010	ウイルスバスター 2011	ウイルスバスター
ウイルスバスター 無料	ウイルスミス	ウイルス性胃腸炎
ウイルス対策ソフト	ウイルスチェック	ベートーベンウイルス
ウイルスミス	ウイルス対策ソフト	ロタウイルス
ウイルス対策ソフト フリー	ウイルスセキュリティ	ノロウイルス

表 3.15 入力語“mac”の拡張クエリ

bing	Google	Yahoo!JAPAN
マクドナルド	mac	mac book
マック	macbook pro	mac book air
マクドナルド クーポン	macbook air	mac Windows
マクドナルド メニュー	macbook	mac OS
マクロス F	mac アドレス	mac book pro
マクロミル	mac mini	mac DVD コピー
幕張メッセ	mac os	MAC アドレス
マカフィ	mac air	mac air

表 3.16 各検索エンジンのキーフレーズ抽出数

入力クエリ	bing	Google	Yahoo!JAPAN
ウイルス	66	82	91
mac	60	43	58

1 語の入力クエリから 30 個の拡張クエリを作成する．入力するクエリは各カテゴリで出現率の高かった語の中で，10 カテゴリ全てに存在する 10 語を用いる．入力クエリとそれに代表されるカテゴリを表 3.17 に示す．比較対象として Yahoo!関連検索ワード API^{*14} によって 抽出した拡張クエリを用いる．以下の cqa と yahoo の 2 つの手法を比較する．

- **cqa** :各カテゴリの第一段階拡張クエリ上位 3 語を 10 カテゴリにまとめた 30 個

^{*14} <http://developer.yahoo.co.jp/webapi/search/assistsearch/v1/webunitsearch.html>

表 3.17 入力語とそのカテゴリ

カテゴリ	入力語	カテゴリ	入力語
seiji	中国	health	検査
renai	友達	travel	東京
pc	ソフト	tv	番組
auCTION	メール	baseball	選手
bukuro	質問	kotoba	日本語

表 3.18 入力語“ソフト”の拡張クエリ

cqa	yahoo
ソフト フリー	ソフトバンク
ソフト 読み上げる	DS ソフト
ソフト シェア	マイクロソフト
ソフト ゲーム	wii ソフト
ソフト os	フリーソフト
ソフト 使う	PSP ソフト
ソフト ハード	PS3 ソフト
ソフト 使い捨て	ソフトバンクホークス
ソフト コンタクトレンズ	解凍ソフト

● **yahoo** :Yahoo!関連検索ワード API によって推薦された拡張クエリ 30 語

例として，入力語“ソフト”の拡張クエリの一部を表 3.18 に示す．作成したは前節のキーフレーズ抽出手順に従い，キーフレーズが抽出される．

各入力語に対して，cqa と yahoo の 30 個の拡張クエリにより検索した Web ページから抽出したキーフレーズの合計を図 3.33 に示す．横軸は各入力語，縦軸は Web ページから抽出できたキーフレーズの合計数である．“ソフト”以外の入力語に対して，提案法が既存手法を上回っている．

次に，キーフレーズ数の増加の推移のグラフを図 3.34 と図 3.35 に示す．図 3.34 が“日本語”，図 3.35 が“ソフト”の結果である．横軸は拡張クエリ，縦軸は抽出したキーフレーズ数の累計である．“日本語”では，クエリ数が多くなるに従い，提案法 (cqa) と既存手法 (yahoo) のキーフレーズ抽出数の差が，大きくなっていくことがわかる．一方で，“ソフト”は，提案法，既存手法共にクエリ数に比例して同様にキーフレーズ数が増加している．

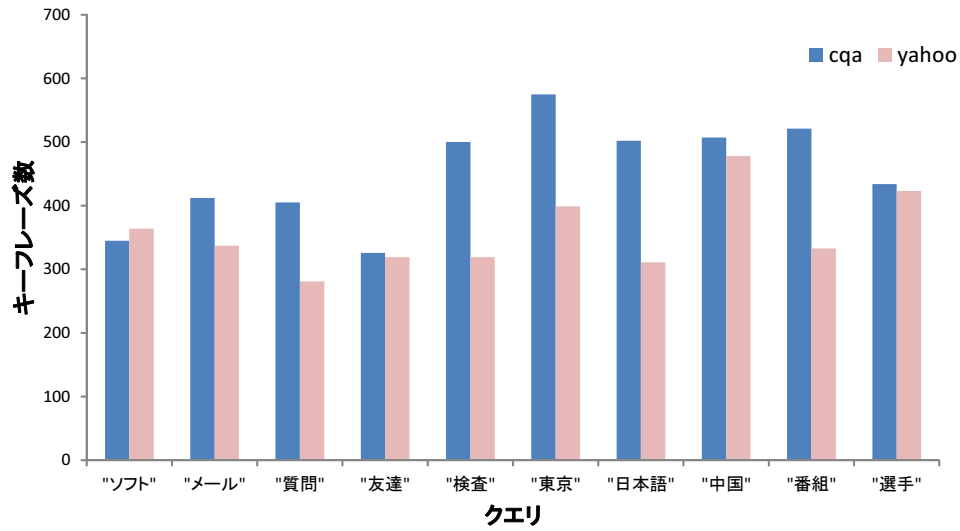


図 3.33 提案法 (cqa) と既存手法 (yahoo) のキーワード抽出数

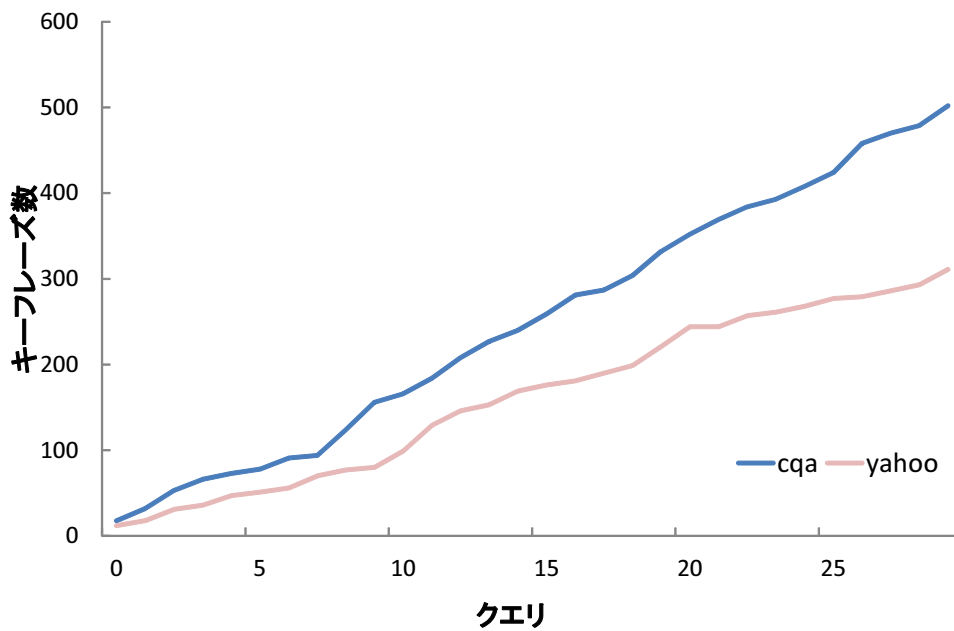


図 3.34 “日本語”でのキーワード抽出数の推移

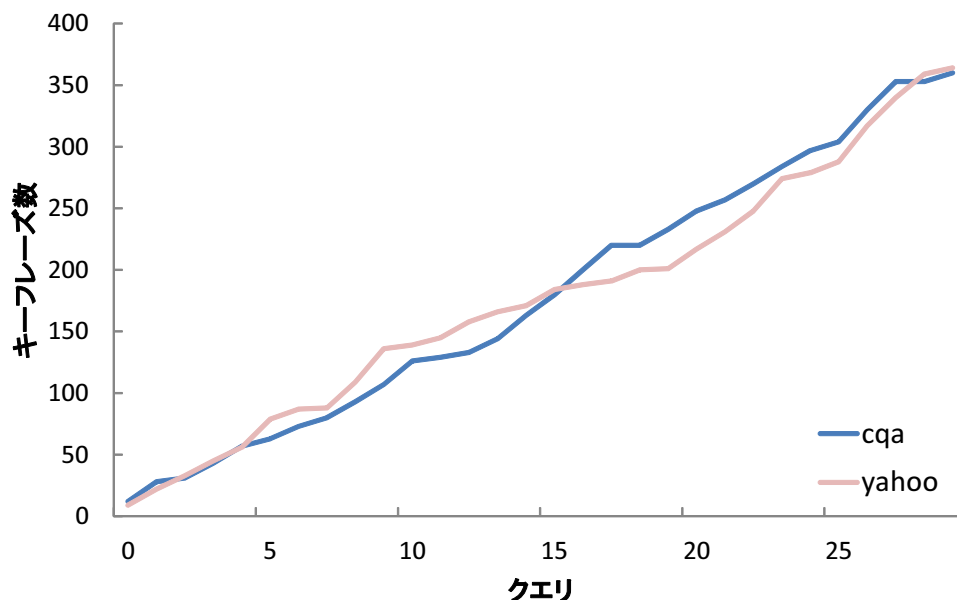


図 3.35 “ソフト”でのキーフレーズ抽出数の推移

表 3.19 全クエリに対する提案法と既存手法の比較

	cqa	yahoo
キーフレーズ抽出数が比較手法を上回ったクエリ数	22	8
平均キーフレーズ抽出数	350.2	312.2

カテゴリ数と多様性の評価

表 3.17 に示した実験用の入力クエリは，提案法において，用意した 10 カテゴリ全てで拡張クエリが推薦されている．しかし，実際には，10 カテゴリ全てのカテゴリにおいて拡張クエリが作成されるとは限らない．そこで，より一般的な入力クエリに対して，拡張クエリを作成し，キーフレーズ抽出数について評価を行う．入力クエリには 2004 年と 2005 年の Yahoo!JAPAN の検索ワードランキング^{*15} の総合ランキングから “ANA”，“dell”，“hotmail” などの企業名やサービス名，“天気” や “地図” など一般生活に関連する語などを幅広く 30 語抽出したものを使用する．各入力クエリに対して，30 個の拡張クエリを作成する．今回の入力クエリでは，抽出できる関連カテゴリは 10 カテゴリとは限らない．入力クエリ q に対して， N_q 個の関連カテゴリが抽出できたとき，1 つのカテゴリ c から抽出する拡張クエ

^{*15} <http://picks.dir.yahoo.co.jp/new/review2004/>
<http://picks.dir.yahoo.co.jp/new/review2005/>

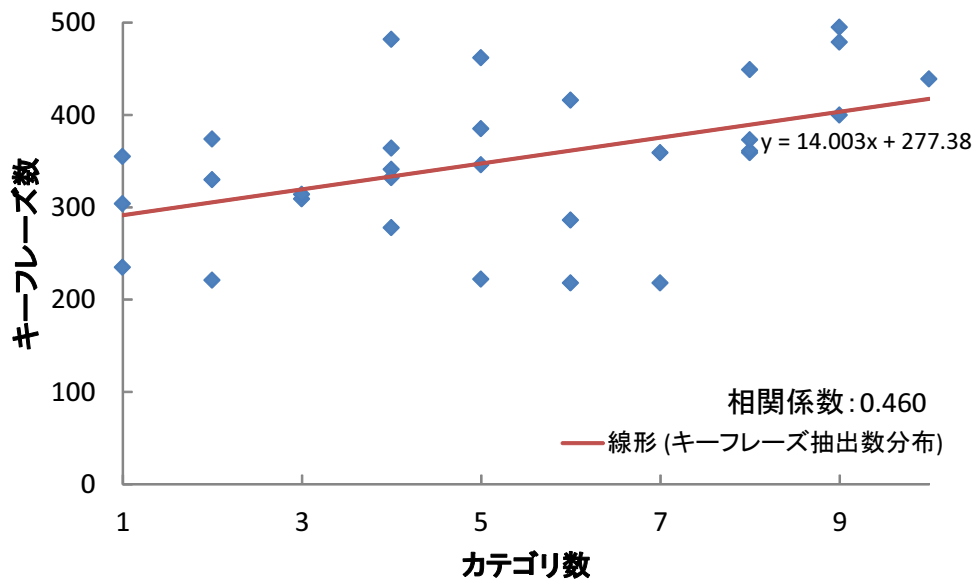


図 3.36 抽出カテゴリ数とキーワード数の相関

り数 $n_{q,c}$ は $n_{q,c} = \lfloor \frac{30}{N_q} \rfloor$ 端数は最も関連の低いカテゴリの拡張クエリを切り捨てることにより調整する。

全入力で拡張クエリを作成し、キーワード抽出を行った結果を表 3.19 に示す。全 30 クエリの内 22 クエリにおいて、提案法が既存手法よりも多くのキーワードを抽出できている。また、平均値においても提案法が多くのキーワードを抽出できている。

次に、各クエリにおいて、抽出できたカテゴリ数とキーワード抽出数の関連を図 3.36 に示す、横軸が抽出できたカテゴリ数、縦軸がキーワードの抽出数である。また、回帰分析を行い、最小二乗法による回帰直線を付与した。使用した入力クエリによる抽出カテゴリ数は 1 から 10 まで幅広く分布した。“ANA”、“dell” は抽出カテゴリ数は 1、“hotmail” は 2 であるのに対し、“天気” は抽出カテゴリ数 9、地図は 7 と固有名詞よりも一般的な語の方がより多くのカテゴリを抽出できていた。1 カテゴリでも 300 程度のキーワードを抽出できているが、回帰分析の結果、抽出カテゴリが多くなるにつれ、キーワード抽出数が増加する傾向が見られた。相関係数は 0.46 であり、これは中程度の正の相関があるといえる。特に、カテゴリ抽出数が 8 以上のクエリについて、多くのキーワードが抽出できていた。

3.5.2 季節の違いによるクエリ拡張の評価

3.3 節の実験で、CQA の質問記事の内容の話題変動は、季節性を持つことがわかっている。本節では、実際のシステムにおいて、データセットを季節ごとに分割した際に、ことなる拡張

クエリが推薦されるのかを評価する。3.2 節の実験で家電カテゴリでは“年賀状”，旅行カテゴリは“桜”，恋愛カテゴリでは“チョコ”が季節性を持つ語であることを明らかにしている。そこで，これらの語を入力クエリとしてシステムに投稿した際，システムに出力された拡張クエリを季節ごとに比較する。

表 3.20 にシステムの出力結果をまとめたものを示す。入力クエリはシステムに最初に入力したキーワードである。それぞれの入力語で最も関連のあるカテゴリでの季節ごとのタグクラウドの出力キーワード数をまとめている。出力語はタグクラウドに表示されたキーワードの総数である。ユニーク語は，タグクラウドに出力された語のうち，他の季節では出力されず，その季節のみで出力されるユニークな語が提示される数であり，ユニーク率はその割合である。家電カテゴリでの“年賀状”が冬，旅行カテゴリでの“桜”が春，恋愛カテゴリでの“チョコ”が冬に最も多くのユニーク語を持つことがわかる。ユニークキーワード数の推移をグラフにまとめたものを図 3.37 に示す。横軸は季節，縦軸はタグクラウドのキーワードを示しており，棒グラフがタグクラウドが出力する総出力語数，折れ線グラフがユニーク語数である。ユニークキーワード数はバーストしている月に対応した季節が最も多く，その次の季節は極端に少なくなっている。またバーストする季節に向かうに連れ徐々にユニークなキーワードが増加していることがわかる。

表 3.21 に，実システムでの CQA クエリの例を示す。例では，入力クエリとして“桜”をシステムに入力し，旅行カテゴリ，春のタブを選択，出力されたタグクラウドから“花見”を選択したときの例である。タグクラウドから“花見”を選択した時点で，CQA の旅行カテゴリで春に投稿された質問記事から“桜”と“花見”が含まれる質問記事が検索される。一つ目の CQA クエリの例では，アメリカ，ワシントン DC の桜についての質問記事であり，もう一方の CQA のクエリは花見をしながら食事ができるレストランを探す質問記事となっている。同じキーワードが含まれる質問記事であっても，質問の主題は異なっている。また，質問記事から作成される Web 検索用のキーワード組（クエリ）も異なるものとなっている。

3.6 考察

3.6.1 クエリ拡張におけるカテゴリと多様性

3.5.1 節の実験では，検索された Web ページのスニペットからキーフレーズを抽出することで多様性を評価できるのかを検証した。異なる話題に関するクエリを提供している検索エンジンが，多くのキーフレーズを抽出できていたことから，キーフレーズが多様性の評価に有効に機能していると考えられる。提案法と既存のクエリ拡張の比較では，ほとんどの場合において

表 3.20 タグクラウドの出力例とユニークキーワード数

入力クエリ [カテゴリ]		出力語	ユニーク語	ユニーク率	キーワード例
年賀状 [デジタル・家電]	春	35	5	0.14	黒, 送る, 用紙サイズ
	夏	26	6	0.23	レーザ, プリント
	秋	52	14	0.26	dpi, スキャン, 画像
	冬	62	23	0.37	レフィルインク, 筆ぐるめ, プレビュー
桜 [旅行]	春	45	37	0.82	ブルーシート, 花見, カラオケ
	夏	13	13	1.00	宮津, 船島
	秋	21	18	0.86	紅葉, 祭り
	冬	22	15	0.68	梅林, 庭, 河津
チョコ [恋愛]	春	57	21	0.37	お返し, ブランド, ルイヴィトン
	夏	28	8	0.29	あげる, バレンタイン, 渡す
	秋	27	4	0.14	手作り, お菓子, ワイン
	冬	69	30	0.43	本命, ネクタイ, ラブレター

提案法の方がより多くのキーフレーズを抽出できていたことから、提案法は多様性を持ったクエリ拡張が行われていると考えることができる。

図 3.36 より、展開されるカテゴリが多ければ多いほど、より多様性を展開できると考えられる。抽出カテゴリ数と、キーワード抽出数の相関実験では、やや強い正の相関となったが、実験で使用した 30 クエリに表 3.17 の全カテゴリで出現する代表的な語を含めると、相関係数は 0.589 となりより強い相関となる。そのため、データ数をさらに増やすことでより相関があることが確認できるのではないかと考えられる。また、図 3.36 では、単一のカテゴリのみで拡張クエリを作成してもある程度のキーフレーズが抽出できた。これは、CQA の性質によるものであると考えられる。CQA では、一問一答の形式で質問のやりとりを行うため、個々の質問記事は完結した内容で具体的な内容が記述されていること、CQA 内では、他人の質問、回答を閲覧できるため、同様の内容の質問が先に存在している場合、質問を投稿する必要がなくなる。膨大な数に及ぶ同カテゴリ内での質問記事も内容が細かく異なっているため、提示される関連語にも多様性が発生したのではないかと考えられる。

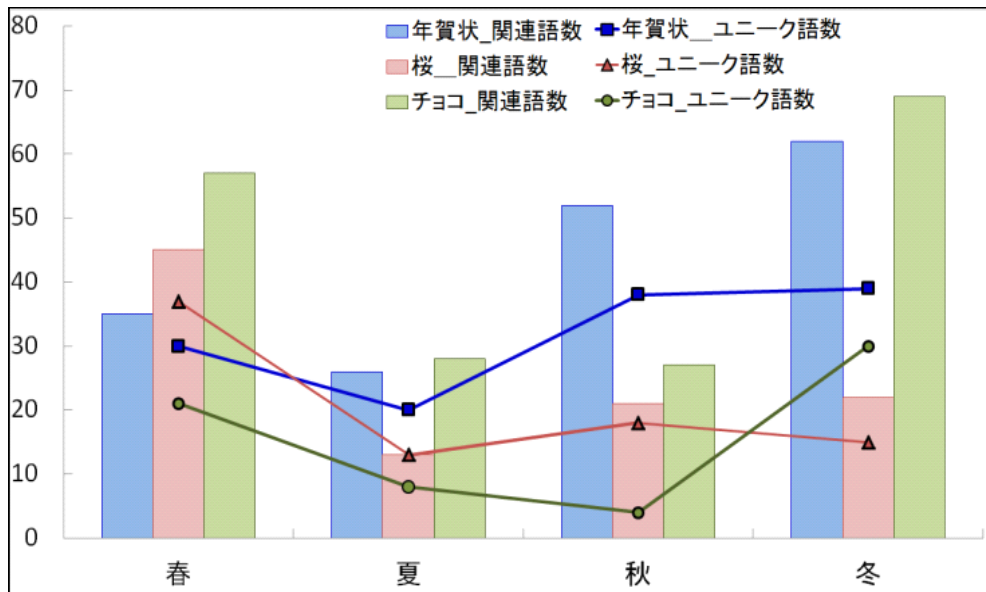


図 3.37 ユニークキーワード数の季節ごとの推移

表 3.21 CQA クエリの例

タグクラウド			CQA クエリ	
選択クエリ	季節	カテゴリ	質問記事本文	キーワード組
桜 花見	春	旅行	アメリカ、ワシントン D.C. のポトマック川周辺に日本から輸入された桜があるそうですが、花見行ったことある人いますか？	桜 花見 アメリカ
			室内でお花見（桜）出来るレストラン等、ご存知の方教えてください。	桜 花見 レストラン

3.6.2 CQA の季節性とクエリ拡張への影響

システムの出力結果では、3.2 節でバーストが発生していた時期に、最もユニークなキーワードが提示されていた。バーストした季節は投稿される質問記事数も多いため、より詳細な内容の質問記事が多いためであると考えられる。このことから、システムは CQA の季節性を反映できていると考えることができる。バーストが発生した季節以外にも、多くのユニーク語が出力されており、恋愛カテゴリでの“チョコ”は2月にバーストする冬のキーワードであるが、春にも多くのユニーク語が出力されている。“お返し”、“ブランド”というキーワードか

らこれは、3月のホワイトデーに関するキーワードであると考えられる。また、家電カテゴリの“年賀状”に関しては、“dpi”、“スキャン”などのユニーク語が出力されているが、これは、年賀状作成のための準備としてプリンタを秋に購入することが多いからではないかと考えられる。このように、バースト以外の季節によっても、システムでは季節に応じた独特な拡張クエリを提示できることを示しており、拡張のコンテキストとして季節を導入したことは妥当であると考えることができる。

バースト以外の季節においても多くのキーワードが出力されているケースもあるが、他の季節と重複したものが多いという傾向があった。これは、質問の内容がごく一般的なものであることを示しており、季節による差別化が行われているといえる。また、更に投稿質問数が少ない場合は少数の質問記事によりタグクラウドに出力されるキーワードが影響を受けていることがわかる。

3.7 まとめ

本章では、Web 検索における情報要求言語化支援のための、クエリ拡張手法を提案した。コミュニティ QA (CQA) の質問記事を外部リソースとして、拡張クエリを実装しキーワード抽出を行った質問記事とセットで提示する“質問記事付き拡張クエリ”により、キーワード組の背後にある疑問の根拠を提示することにより、ユーザ自身では思いつくことのできない情報要求を言語化している。

CQA の質問記事のメタ情報である投稿カテゴリと投稿時期を検索のコンテキストとして、タブとタグクラウドを組み合わせたユーザインターフェースにより多様なクエリ拡張を提供するシステムの実装を行った。実験により、異なるカテゴリから拡張クエリを作成することで、既存のクエリログを用いた拡張クエリよりも多様な Web ページを検索できることを明らかにした。また、季節ごとにデータセットを分割し、異なる拡張クエリを提供することで季節特有の情報要求に対するクエリが推薦されることを明らかにした。

第 4 章

クローズドナレッジにおける情報要求の言語化

4.1 はじめに

人工知能がユーザの質問について回答する質問応答技術の必要性は，スマートフォンやスマートスピーカー，ロボットといったデバイスの普及に伴いますます高まっている．近年，機械読解と呼ばれる質問応答が特に注目を集めている [29]．機械読解とは，システムが自然言語で記述された文書（パッセージ）を読み解き，そのパッセージに関する質問に対して，回答となる情報をパッセージ中から発見，抽出することで回答を可能とする質問応答技術である．

機械読解による質問応答は，深層学習に基づく手法により高い回答精度を達成することができている．例えば，Wikipedia の記事に対して，クラウドソーシングによって質問と回答を作成した機械読解のオープンデータセットである SQuAD [54] では，BiDAF [57] や QANet [73] といったニューラルネットワークによる End-to-End 型の機械読解モデルで人間に匹敵する回答精度を得られることが報告されている．

機械読解は自然言語による質問応答で高い精度を達成できる技術であるが，実際に機械読解を質問応答システムとして使用する場合を想定すると，依然として課題が存在する．従来の機械読解の回答精度は，SQuAD などのオープンデータセットを用いた実験によって得られたものである．しかしながら実際の質問応答システムでは，オープンデータセットの様に，回答を特定できるような明確な質問が必ずしも入力されるとは限らない．回答が複数個考えられる曖昧な質問や，回答するための必要な情報が不足している短い質問が入力されたとき，現在の機械読解では十分な回答精度を得ることは難しい．

質問者が自身の質問意図を明確にできず、曖昧な質問をしてしまう場面は、人同士の質問応答でも発生する。例えば、コールセンターの質問者とオペレーターとのインタラクションでは、不慣れな質問者が、曖昧であったり、意図を特定できない質問をオペレーターに問い合わせってしまう場面がある。オペレーターは、マニュアル等の外部知識と質問者の質問を照らし合わせ、質問者の質問意図を推測し、質問者に確認する作業を行う。このときオペレーターは、質問者の質問をオペレーターが回答可能な具体的な質問に訂正し質問者に聞き返すことで、質問意図を確認する。本研究では、オペレーターと同様に、質問者の質問意図を推測し確認する仕組みを機械読解による質問応答システムに導入することで、曖昧な質問に対する自動応答を実現する。

質問応答において曖昧な質問が入力されたとき、質問者の質問意図の明確化を目的とするタスクとして「改訂質問生成 (Specific Question Generation : SQG)」が提案されている [85]。改訂質問生成は、機械読解による質問応答において、パッセージと入力質問が与えられたとき、入力質問をパッセージの内容に基づいて、具体的な質問になるように書き換える。本研究では、改訂質問生成を用いた機械読解による質問応答手法を提案する。改訂質問生成を用いた質問応答を図 4.1 に示す。改訂質問生成により生成した具体化した質問である改訂質問 (Specific Question : SQ) の候補を質問者に提示する。質問者は改訂質問の中から自身の質問意図に近い質問を選択することで、質問応答システムから精度の高い回答を得ることができる。

改訂質問生成を実行可能にするモデルとして、本研究では「改訂質問生成モデル (Specific Question Generation Model: SQGM)」を提案する。改訂質問生成モデルは、生成モデルである Encoder-Decoder モデル [4] をベースに、機械読解モデル [57] とコピー機構 [11] を組み合わせた End-to-End 型のニューラルネットワークである。改訂質問生成モデルに入力質問と質問対象のパッセージを入力すると、改訂質問生成モデルは入力質問とパッセージの内容を機械読解モデルにより読み解き、関連する情報を抽出する。抽出した情報をアテンションによって注視しながら、生成モデルによって改訂質問を生成する。この際、コピー機構によって入力質問やパッセージ中の重要な単語や表現を改訂質問にコピーすることにより、パッセージの内容に即した改訂質問を生成している。本研究では改訂質問生成モデルの提案に加えて、文圧縮によって擬似的に短い質問を作成することによる学習データの自動生成手法についても提案する。

4.1.1 本章の構成

4.1.2 問題定義

本節では、本章が取り組むタスク、関連する用語についての定義を行う。

Passage: CBS provided digital streams of the game via CBSSports.com, and the **CBS Sports apps** on tablets, Windows 10, Xbox One and other digital media players (such as Chromecast and Roku). Due to Verizon Communications exclusivity, streaming on smartphones was only provided to Verizon Wireless customers via the **NFL Mobile service**. The ESPN Deportes Spanish broadcast was made available through WatchESPN.

Question: What app did viewers use to watch the game on their smartphones?

Answer of worker #1: **NFL Mobile service**

Answer of worker #2: **the CBS Sports apps**

Answer of worker #3: **NFL**

Specific Question #1: What app did viewers use to watch the game on their smartphones of Verizon Wireless?

Coresponding system answer: **NFL Mobile service**

Specific Question #2: What app did viewers use to watch the game on their tablets?

Coresponding system answer: **CBS Sports apps**

図 4.1 改訂質問生成の概要

[定義 1] 改訂質問生成 (Specific Question Generation : SQG) は、入力質問とパッセージを入力とし、改訂質問 (SQ) を出力するタスクである。改訂質問生成を用いた質問応答では、質問者は複数の改訂質問の候補から自身の質問意図に合った改訂質問を選択することができる。

[定義 2] 機械読解 (Reading Comprehension : RC) は、入力質問の単語トークン系列 q と、パッセージの単語トークン系列 x を入力とし、パッセージ中から回答の始点 s と終点 e を推定するタスクである。機械読解の出力となる回答はパッセージから、推定した回答範囲を抜き出した単語トークン系列となる。

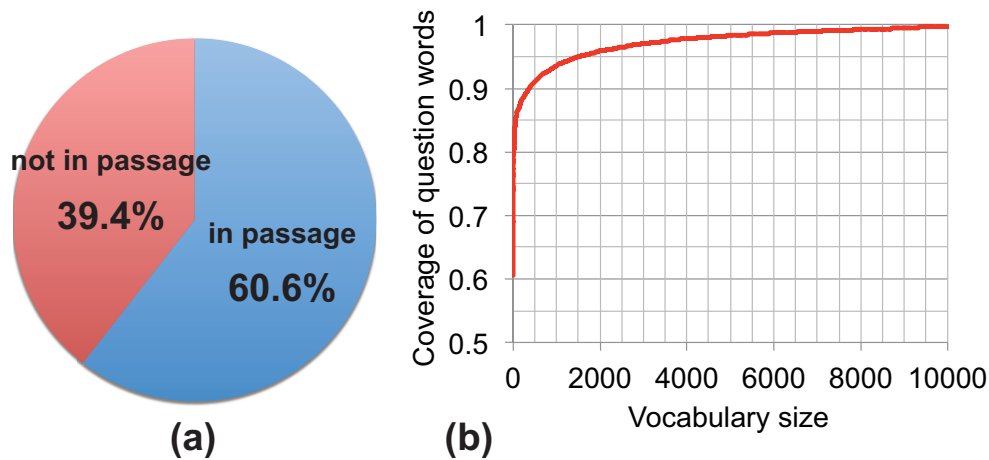


図 4.2 SQAD データセットの質問文中の単語について. (a) 質問文中の単語がパッセージに含まれる割合 (b) 質問中の単語をカバーするために必要な語彙数の分布

[定義 3] 改訂質問生成モデル (Specific Question Generation Model : SQGM) は, 入力質問の単語トークン系列 q と, パッセージの単語トークン系列 x を入力とし, 改訂質問の単語トークン系列 y を出力するニューラルネットワークモデルである. ここで, 各トークン系列の単語は改訂質問生成が扱う語彙集合 (語彙数: V) に含まれるものとする.

[定義 4] 入力質問は, 自然言語で記述された文である. 入力質問は形態素解析等により, 単語トークン列に分割し, one-hot ベクトルの系列である $q = \{q_1, q_2, \dots, q_J\}$ として表す. one-hot ベクトルは, 単語辞書のインデックスに対応する次元のみ 1, それ以外の次元を 0 とする V 次元のベクトルである.

[定義 5] パッセージは, 自然言語で記述された文である. パッセージは数百語程度の単語から構成され, 入力質問と同様に one-hot ベクトルで表される単語トークン列 $x = \{x_1, x_2, \dots, x_T\}$ とする. ここで, パッセージは, 入力質問の回答となる情報を含むものとする.

[定義 6] 改訂質問 (Specific Question : SQ) は, 入力質問の内容が具体化された文 $y = \{y_1, y_2, \dots, y_K\}$ である. 改訂質問は入力質問をパッセージ内容に沿って回答が一意に決定できるように具体化した質問である.

4.2 改訂質問生成による機械読解モデル

本研究で提案する改訂質問生成と機械読解による質問応答を図 4.3 に示す. 入力質問とパッセージを入力すると, 機械読解モデルにより, 回答がパッセージから抽出される. 同時に, 改訂質問生成モデルにより, 入力質問とパッセージから複数の改訂質問の候補が生成される. 質問者は機械読解が出力した回答に満足できない場合, 改訂質問の候補のうち, いずれか 1 つを選択する. 選択した改訂質問とパッセージを機械読解モデルに入力することで, 新たな回答を

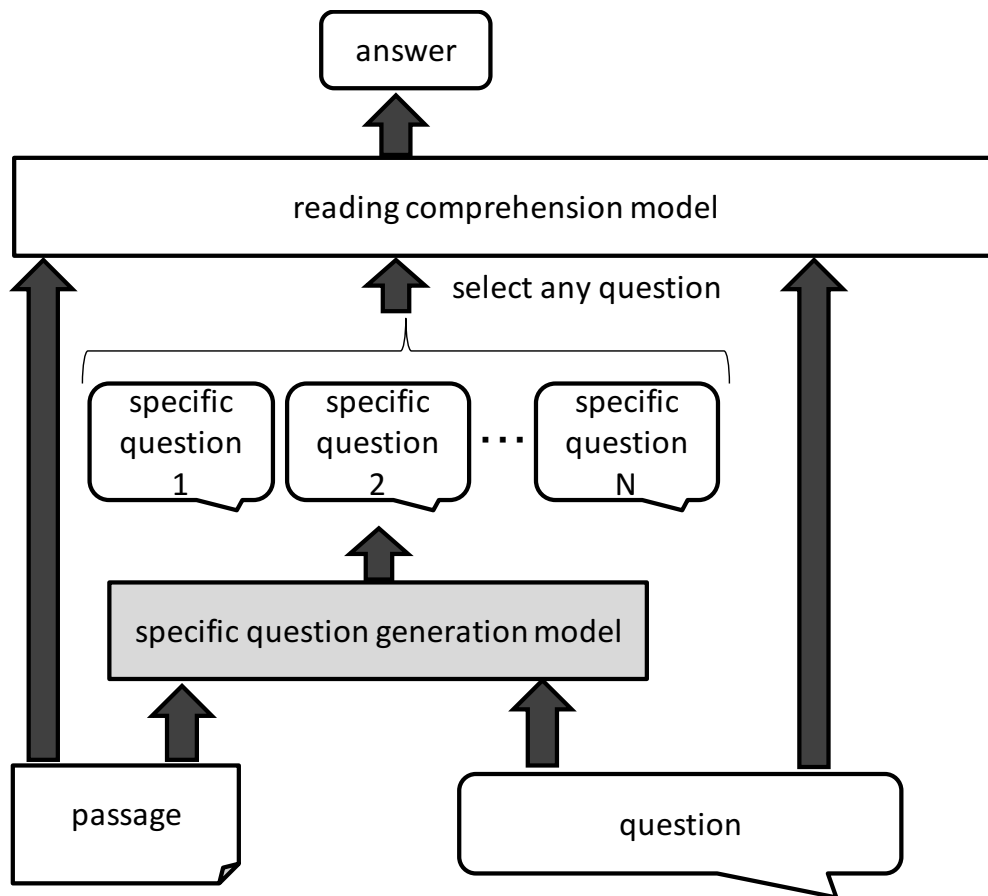


図 4.3 改訂質問生成と機械読解による質問応答.

得ることができる。本研究では、機械読解は任意の既存モデルを使用することを想定する。以降は、改訂質問生成モデルについて詳述する。

本研究が提案する改訂質問生成モデルの構成を図 4.4 に示す。改訂質問生成モデルはアテンション付きのエンコーダ・デコーダモデルをベースに、機械読解モデルと 2 つのコピー機構を組み合わせたモデルである。[85] のモデルでは、コピー機構はパッセージからコピーを行う 1 つのコピー機構となっていたが、本研究で提案するモデルは入力質問からもコピーを行う。

改訂質問生成モデルは以下の 3 層によって構成される。

1. **Encoding layer** : パッセージと入力質問の単語トークン系列を、学習済みの単語埋め込みモデルにより連続値のベクトルに変換する。また、パッセージと入力質問のコンテキストを考慮したベクトル系列を作成する。
2. **Matching layer** : パッセージと入力質問の単語間の相互関係を捉えることにより、入力質問に依存した、パッセージのベクトル系列をモデリングする。
3. **Decoding layer** : アテンション付きリカレントニューラルネットワークと 2 つのコピー

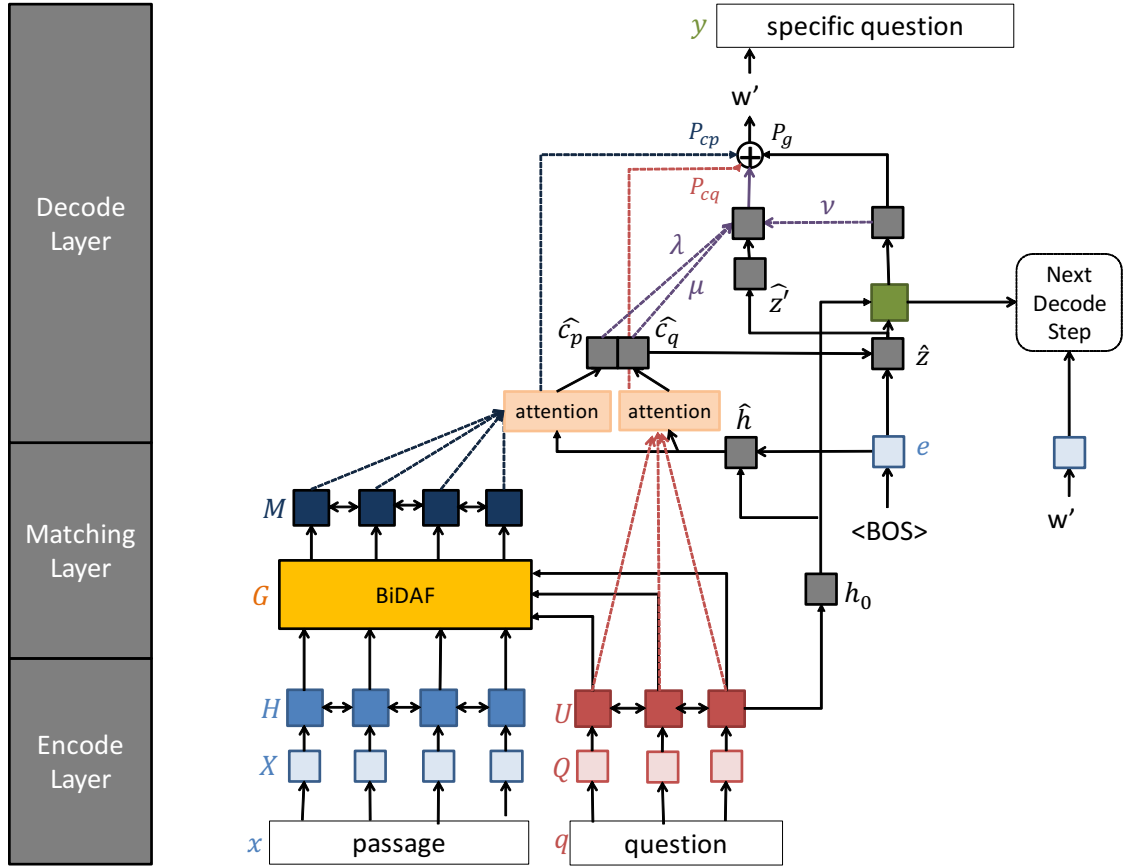


図 4.4 改訂質問生成モデル (specific question generation model : SQGM).

機構により，改訂質問を構成する単語トークン系列を生成する．

以降は各層について詳述する．

4.2.1 Encoding layer

Encoding layer の入力として， V 次元で表現される one-hot のベクトル系列で表されたパッセージ $x = \{x_1, \dots, x_T\}$ と入力質問 $q = \{q_1, \dots, q_J\}$ を与える．はじめに，one-hot ベクトルから， v 次元で表現される連続値ベクトルに変換する．単語の変換には学習済みのベクトル変換行列 $W_e \in \mathbb{R}^{v \times V}$ を用いる．変換されたベクトルは 2 層のハイウェイネットワーク [60] を通して，最終的にパッセージと入力質問それぞれのベクトル系列 $X \in \mathbb{R}^{v \times T}$ と $Q \in \mathbb{R}^{v \times J}$ を得る．

連続値のベクトル系列に変換されたパッセージと入力質問を，リカレントニューラルネットワーク (RNN) に入力する．本研究では，RNN に 1 層の GRU [14] を使用する．ここで，GRU のパラメータはパッセージと入力質問で共有している．また，最初に入力したベクトルの影響が小さくなることを防ぐために，双方向の GRU (Bi-GRU) を使用する．Bi-GRU によ

り、パッセージのコンテキスト行列 $H \in \mathbb{R}^{2d \times T}$ と入力質問のコンテキスト行列 $U \in \mathbb{R}^{2d \times J}$ を得る。ここで、 d は隠れ状態の次元数である。

エンコーダの RNN の状態は、デコード時の RNN の初期状態を求めるためにも使用する。Decoding layer の RNN の初期状態として与えるベクトル $h_0 \in \mathbb{R}^{2d}$ は、入力質問のコンテキスト行列にアテンションを適用した以下の式により計算する。

$$h_0 = \sum_{j < J} \alpha_j U_j, \quad (4.1)$$

ここで、 $\alpha_j = \text{softmax}_j(U_j^\top U_j)$ であり、 U_j^\top は、入力質問のコンテキスト行列の最終状態を表している。

4.2.2 Matching layer

Matching Layer では、Encoding Layer でエンコードしたパッセージと入力質問を照合し、パッセージ中から、入力質問に関連する領域を発見する。パッセージと入力質問の関連性を取得するモデルとして、本研究では、Bi-directional attention flow (BiDAF) [57] を使用する。BiDAF は、パッセージと入力質問の双方向からアテンションを計算し、最終的に入力質問に依存したパッセージのコンテキスト行列を作成する。

BiDAF ではまず、パッセージに関するコンテキスト行列 H と入力質問に関するコンテキスト行列 U により、類似度行列 $S \in \mathbb{R}^{T \times J}$ を以下の式により計算する。

$$S_{tj} = w_s^\top [H_t; U_j; H_t \odot U_j], \quad (4.2)$$

ここで、 $w_s \in \mathbb{R}^{6d}$ は学習パラメータであり、 $[\cdot]$ は行成分のベクトルの連結を表している。次に、類似度行列を基に、パッセージから入力質問へのアテンションと、入力質問からパッセージへのアテンションの2方向のアテンションを計算する。

パッセージから入力質問へのアテンションでは、パッセージ中の単語について、入力質問の単語で重み付けしたベクトルを計算する。パッセージの t 番目の単語についてのアテンションベクトル $\check{U}_t \in \mathbb{R}^{2d}$ は次式によって計算する。

$$a_t = \text{softmax}_j(S_t), \quad (4.3)$$

$$\check{U}_t = \sum_j a_{tj} U_j, \quad (4.4)$$

入力質問からパッセージへのアテンションでは、入力質問のいずれかの単語に強く関連する単語に重みをつけたベクトル \check{h} を、パッセージの系列長 T 分だけ並べた行列 $\check{H} \in \mathbb{R}^{2d \times T}$ を次式により計算する。

$$b = \text{softmax}_t(\max_j(S)) \quad (4.5)$$

$$\check{h} = \sum_t b_t H_t \quad (4.6)$$

パッセージと入力質問の双方向のアテンションベクトルは、パッセージの各単語に対して、アテンションのベクトルを連結した次式により計算する。

$$G = [H; \check{U}; H \odot \check{U}; H \odot \check{H}] \in \mathbb{R}^{8d \times T}. \quad (4.7)$$

最終的に、双方向のアテンションベクトル G を1層の Bi-GRU に入力し、入力質問を考慮したパッセージのコンテキスト行列 $M \in \mathbb{R}^{2d \times T}$ を得る。

4.2.3 Decoding layer

Decoding layer では、Encoding layer と Matching layer の情報に基づいて、改訂質問を生成する。Decoding layer は、RNN 言語モデルに基づく生成モデルと、パッセージと入力質問のコピー機構を組み合わせたネットワークで構成される。

■RNN 言語生成モデル 生成モデルは、アテンション付きの1層 GRU とソフトマックス層により構成される。改訂質問の単語列 $y = \{y_1, \dots, y_s\}$ を入力したとき、生成モデルによって出力される、改訂質問の次の単語の確率分布 $P_g(y_{s+1}|y_{\leq s}, x, q)$ は次式によって表される。

$$P_g(y_{s+1}|y_{\leq s}, x, q) = \text{softmax}(W_g h_{s+1} + b_g), \quad (4.8)$$

ここで、 $W_g \in \mathbb{R}^{2d \times V_g}$ と $b_g \in \mathbb{R}^{V_g}$ は学習パラメータを表している。 V_g は、生成モデルで生成する単語の語彙数を表しており、 $V > V_g$ である。生成モデルから生成される単語は高頻度の単語のみとし、低頻度の単語はコピー機構により抽出的に生成することで、生成モデルのサイズを削減し、学習速度を向上させる。 $h_{s+1} \in \mathbb{R}^{2d}$ は、GRU の $s+1$ 番目の隠れ状態であり、 $h_{s+1} \leftarrow \text{GRU}(h_s, \hat{z}_s)$ によって更新される。

GRU の入力ベクトル \hat{z}_s は、生成モデルが出力した一つ前の単語 y_s に基づき、以下のように決定する（ここで、最初の入力は文の始端トークン $\langle BOS \rangle$ とする）。 y_s は、Encoding layer と同様に、one-hot ベクトル化した後、単語埋め込み層と2層のハイウェイネットワーク層によって連続値のベクトル $e_s \in \mathbb{R}^v$ となる。次に、アテンションを行うためのクエリとなるベクトル $\hat{h}_s \in \mathbb{R}^{2d}$ を、 e_s と1つ前の GRU の隠れ状態 h_s を用いて次式により作成する。

$$\hat{h}_s = f(W_d[e_s; h_s] + b_d). \quad (4.9)$$

パッセージのアテンション $\alpha_{st} \in \mathbb{R}^T$ と、入力質問のアテンション $\beta_{sj} \in \mathbb{R}^J$ はクエリベクトルを用いて次式により計算する。

$$\alpha_{st} = \text{softmax}_t(M_t \cdot \hat{h}_s), \quad (4.10)$$

$$\beta_{sj} = \text{softmax}_j(U_j \cdot \hat{h}_s). \quad (4.11)$$

最終的に、GRU の入力 $\hat{z}_s \in \mathbb{R}^{v+4d}$ は次式により得る。

$$\hat{c}_{ps} = \sum_t \alpha_{st} M_t, \quad (4.12)$$

$$\hat{c}_{qs} = \sum_j \beta_{sj} U_j, \quad (4.13)$$

$$\hat{z}_s = [e_s; \hat{c}_{ps}; \hat{c}_{qs}], \quad (4.14)$$

ここで、 $W_d \in \mathbb{R}^{(v+4d) \times 2d}$ と $b_d \in \mathbb{R}^{2d}$ は学習パラメータである。また、 f は活性化関数を表しており、本研究では PReLU[27] を使用する。

■コピー機構 コピー機構は、文生成の際に、入力単語の一部をコピーすることにより、文章や対話などの文脈に一貫性を持たせた文を生成するための機構である。コピー機構を導入することで、入力質問やパッセージの内容に沿った改訂質問を生成する。要約や翻訳などでコピー機構が用いられる際はコピー元のテキストは1つであったが、本研究では入力質問とパッセージの両方をコピー元とする。

アテンションの重みを言語モデルの単語の生起確率と捉え [11]、コピー機構による単語の生成確率分布を次式により得る。

$$P_{cp}(y_{s+1}|y_{\leq s}, x, q) = \sum_t \mathbb{1}(y_{s+1} = x_t) \alpha_{(s+1)t}, \quad (4.15)$$

$$P_{cq}(y_{s+1}|y_{\leq s}, x, q) = \sum_j \mathbb{1}(y_{s+1} = q_j) \beta_{(s+1)j}, \quad (4.16)$$

ここで、 $\mathbb{1}(y_s = x_t)$ は、 $y_s = x_t$ のとき 1、それ以外のときは 0 となる関数である。 $\mathbb{1}(y_s = q_j)$ も同様である。

■言語生成モデルとコピー機構の統合 RNN 言語生成モデルとコピー機構では、それぞれで単語の生起確率の分布を得られる。最終的な出力単語を決定するための単語の確率分布 $P(y_{s+1}|y_{\leq s}, x, q)$ は、各々の確率分布の重み付き和によって計算される。

$$\begin{aligned} P(y_{s+1}|y_{\leq s}, x, q) &= \lambda_s P_g(y_{s+1}|y_{\leq s}, x, q) \\ &\quad + \mu_s P_{cp}(y_{s+1}|y_{\leq s}, x, q) \\ &\quad + \nu_s P_{cq}(y_{s+1}|y_{\leq s}, x, q), \end{aligned} \quad (4.17)$$

ここで、 λ_s, μ_s, ν_s は $\lambda_s, \mu_s, \nu_s \in [0, 1]$ 、 $\lambda_s + \mu_s + \nu_s = 1$ の値をとる重みパラメータであり、以下の式に示す softmax 層の出力 $\gamma_s \in \mathbb{R}^3$ により決定する。

$$\hat{z}'_s = \text{Highway}_2(\hat{z}_s), \quad (4.18)$$

$$\gamma_s = \text{softmax}(W_c \hat{z}'_s + b_c), \quad (4.19)$$

$$\lambda_s = \gamma_{s0}, \quad \mu_s = \gamma_{s1}, \quad \nu_s = \gamma_{s2},$$

ここで、 $\text{Highway}_2()$ は 2 層のハイウェイネットワークを示している。また、 $W_c \in \mathbb{R}^{(v+4d) \times 3}$ と $b_c \in \mathbb{R}^3$ は学習パラメータである。

Decoding layer が生成する確率分布をもとに、単語系列を生成することにより改訂質問を生成する。ここで、生成時にビームサーチを導入し、ビーム幅 b にしたがって、生成範囲を探索することで、複数の改訂質問の候補を生成する。

4.2.4 改訂質問生成モデル学習

改訂質問生成モデルの学習は、誤差関数 L を最小にするようにパラメータを更新することで行う。誤差関数 L は、負の対数尤度を用いた以下の式で計算する。

$$L = -\frac{1}{N} \sum_i \sum_s \log P(y_{s+1}^{(i)} | y_{\leq s}^{(i)}, x^{(i)}, q^{(i)}), \quad (4.20)$$

ここで、 N はミニバッチのサイズであり、 i はミニバッチの i 番目のサンプルを示すインデックスである。

4.2.5 文圧縮による学習データ作成

改訂質問生成モデルの学習には、入力となるパッセージと入力質問と、教師データとなる改訂質問を用意する必要がある。本研究では、SQuAD[54] などの機械読解コーパスから改訂質問生成モデルの学習データを作成する。機械読解コーパスの質問を、回答のための情報が十分に含まれている改訂質問とみなし、教師データとする。その上で、機械読解コーパスの質問から重要な情報を欠落させた質問を機械的に作成し、これを入力質問とする。

本研究では、文の依存構造と整数計画法を用いた教師なしの文圧縮手法 [21] により、英語等にも適用可能な学習データの作成手法を提案する。文圧縮ではまず、入力文の依存構造を取得する。日本語では、係り受け解析を行った結果である文節間の依存構造を用いる。英語では単語間の依存関係を取得する。次に、解析結果から取得した文の構成単位（単語）について、文の先頭から順にインデックス番号を付与し、このインデックス番号を基に整数計画問題の定式化を行う。文圧縮の整数計画問題は以下の式により定義される。

$$\begin{aligned} & \underset{a_i \in \{0,1\}}{\text{maximize}} && \sum_{i \leq L} w_i a_i \\ & \text{subject to} && \sum_{i \leq L} a_i \leq l \\ & && a_{\text{parent}(i)} - a_i \geq 0 \quad (1 \leq i \leq L) \\ & && a_{\text{argmax}(w_i)} = 0 \end{aligned} \quad (4.21)$$

ここで、 $a_i = 1$ のとき、文中の i 番目の単語は選択される。 $a_i = 0$ のとき i 番目の単語は選択されず、文圧縮によって削除される。 L は文圧縮前の文の長さ（単語数）であり、 l は文圧縮後の文の最大単語数である。 $\text{parent}(i)$ は、依存構造上で i 番目の単語の親ノードに当たる単

語を示している． w_i は単語重みを表し，本研究では $w_i = \log(F/F(a_i))$ で計算する．このとき， F は学習コーパス上の全単語の出現頻度の合計， $F(a_i)$ は i 番目の単語の出現頻度である．

文圧縮では， $\sum_{i \leq L} a_i \leq l$ の制約によって圧縮後の文の長さを決定する．本研究では，長さを $3 \leq l \leq L - 2$ の範囲で変化させた時のそれぞれの長さで質問を圧縮する． $a_{\text{parent}(i)} - a_i \geq 0$ の制約は依存構造で，親ノードの単語が選択されていないとき，子ノードの単語も選択されないようにするためのものである． $a_{\text{argmax}(w_i)} = 0$ の制約は単語重みが最大の語を敢えて選択しないようにするための制約である．本研究では，回答のための情報が不足している質問を生成するために，質問中で重要と思われる単語を除外するように文圧縮を行っている．

文圧縮は，依存構造を取得できる言語であれば適用可能である．英語の場合は単語間の依存構造によって文圧縮を行うが，日本語の場合でも文節間の依存構造（係り受け関係）を用いることで同様に文圧縮を適用できる．

4.3 評価実験

本節では，改訂質問生成に関する評価実験について述べる．本研究では，**1. 質問の復元実験**と，**2. 改訂質問による機械読解実験**の2つの実験を行うことで提案手法の有効性について検証する．

本節では，まず評価実験に使用する質問について説明し，次に2つの実験について詳述する．

4.3.1 質問セット

本研究では，機械読解のオープンデータセットを用いた実験を行う．機械読解のオープンデータセットには，パッセージと質問，そして質問に対する回答が含まれる．本研究では，機械読解データセットの質問から，以下に示す3種類の質問を用意する．

1. **オリジナル質問 (Original Questions : OQs)** 無加工の機械読解のデータセットの質問．
2. **圧縮質問 (Compressed Questions : CQs)** オリジナル質問に対して，文圧縮によって機械的に短縮した質問．
3. **要約質問 (Abstracted Questions : AQs)** オリジナル質問に対して，人手によって，最低限の質問意図が特定できるまで短縮した質問．オリジナル質問から修飾語等の単語を削除していき，最終的に英語の基本5文型の範囲に収まるまで質問を短縮することで要約質問を作成している．このとき，短縮の結果質問意図は判別可能だが，回答が一意に定まらなくなる質問が作成されることも許容する．

各質問の利用に関する概要を図4.5に示す．質問改訂モデルの学習には，機械読解データ

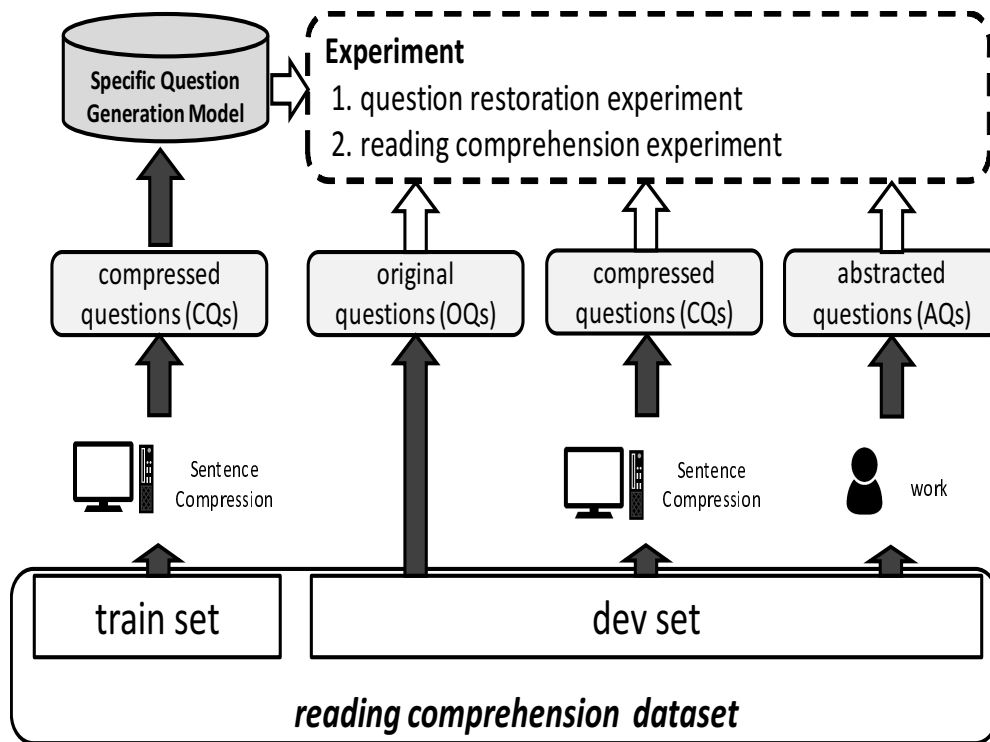


図 4.5 実験で使用する質問セットの概要。

セットの学習セットの質問を利用する。学習セットのオリジナル質問から圧縮質問を作成し、圧縮質問からオリジナル質問を復元するように改訂質問生成モデルを学習する。ここで、要約質問は改訂質問生成モデルの学習では使用しない。評価実験では、公開されている機械読解の開発セットをテストセットとして使用する。オリジナル質問から圧縮質問、要約質問を作成し各実験で使用する。

4.3.2 改訂質問生成モデルによる質問の復元実験

本節では、質問の復元実験について述べる。本実験ではまず、圧縮質問（CQs）および要約質問（AQs）を改訂質問生成モデルに入力し、改訂質問（SQs）を得る。次に、生成した改訂質問とオリジナル質問（OQs）の類似度を計算する。オリジナル質問と類似度が高い改訂質問を生成することができれば、改訂質問生成モデルは、入力質問に不足した情報を適切に補った改訂質問を生成できるモデルであるといえる。

以降はまず、実験で作成したデータセットについて述べる。次に改訂質問生成モデルの作成方法について説明した後、評価尺度と比較モデルについて詳述し、最後に実験結果を示す。

表 4.1 質問の復元実験で使用する機械読解データセット. N はパッセージ数, 質問数を示しており L はパッセージまたは質問の単語長である,

	train		dev		
	OQs	CQs	OQs	CQs	AQs
SQuAD					
N. passage	18,896	18,896	2,067	2,067	278
N. questions	87,599	1,235,000	10,570	32,208	350
L. passage	142.8	142.8	144.5	144.5	140.8
L. question	13.7	7.9	13.3	7.7	6.6
NewsQA					
N. passages	3,072	3,072	1,900	1,900	—
N. questions	10,510	69,959	3,000	16,569	—
L. passage	237.9	237.9	297.5	144.5	—
L. question	10.4	6.3	10.7	6.3	—

データセット

質問の復元実験には, 機械読解のオープンデータセットである SQuAD [54] ^{*1} と NewsQA [64] ^{*2} を使用する. SQuAD は Wikipedia の記事をもとに作成された, 機械読解の基本的なデータセットの一つである. NewsQA はニュース記事によって作成された機械読解データセットであり, SQuAD の質問よりも短く, 抽象的な質問が多いという特徴を持つ.

データセットの詳細を表 4.1 に示す. ここで, 各圧縮質問は文圧縮により, 単語長の制約を 3 から $L - 2$ (L は入力文の最大単語長) まで変化させることで, 1 オリジナル質問に対して, 複数の圧縮質問を作成している. 文圧縮の際に必要な依存構造は, 言語解析器 spaCy^{*3}によって取得した. テストセットについては, オリジナル質問が長いほど多くの圧縮質問を生成することから, 評価の偏りが発生するため, 10 個以上の圧縮質問が生成された場合は, 10 個になるようにランダムに抽出を行った. また, NewsQA には学習セットとテストセットの区別が存在しないため, 学習セットとパッセージが重複しない 3000 質問をランダムで抽出して作成している.

改訂質問生成モデル学習

質問改訂モデルの Encoding layer に入力する, 学習済みの単語埋め込みモデルとして, Wikipedia の全記事と SQuAD, NewsQA の全パッセージ, 質問を結合したコーパスから作成した, 300 次元の fastText [7] のモデルを使用する.

質問改訂生成モデルの学習は, 2 つの GPU (Quadro P6000) を使用した. 学習のためのミ

^{*1} <https://rajpurkar.github.io/SQuAD-explorer/>

^{*2} <https://datasets.maluuba.com/NewsQA>

^{*3} <https://spacy.io/>

ニバッチ数は96に設定し、隠れ状態の次元数 d は128とした。生成モデルの出力語彙数 V_g は5000とし、SQuADとNewsQAのそれぞれの質問で出現回数上位5000単語（質問の全語彙の90%をカバー）を生成する。学習の際のドロップアウト [59] の確率は、ハイウェイネットワークの入力層のみ0.2とし、残りの層を0.5に設定した。パラメータの最適化にはAdamを使用し、学習の反復回数はSQuADが5回（全64,300学習ステップ）、NewsQAが20回（全14,560学習ステップ）である。

評価尺度

改訂質問とオリジナル質問の類似度を計算指標として、精度、再現率、F値、意味的類似度を用いる。精度は、改訂質問の全単語のうち、オリジナル質問の単語が含まれている割合を示しており、再現率は、オリジナル質問の全単語に対する、改訂質問の単語の網羅度を示している。F値は、精度と再現率の調和平均である。意味的類似度は、改訂質問とオリジナル質問の文ベクトルを作成したときのコサイン距離により計算する。ここで、文ベクトルの作成にはuniversal sentence encoder [12] を用いる。

オリジナル質問と比較する改訂質問は、改訂質問生成モデルのビームサーチのビーム幅を1としたときに生成される、1つの改訂質問である。

比較モデル

質問の復元実験において、提案モデル（**proposed**）と比較するモデルとして、提案モデルから、一部の層を除外したモデルを使用する。まず、コピー機構を除外したモデルとして、Decoding layer のすべてのコピー機構を除外したモデル（**w/o copy**）、パッセージからのコピーのみを除外したモデル（**w/o p_copy**）、質問からのコピーのみを除外（**w/o q_copy**）した3つのモデルを用意する。ここで、質問からのコピーを除外したモデル（**w/o q_copy**）は、[85] の提案モデルとほぼ同一のモデルとみなすことができる。次にコピー機構はそのままにし、Matching layer のBiDAFモデルを除外したモデル（**w/o BiDAF**）を用意する。本モデルでは、デコード時のアテンションとして、提案モデルで使用したBiDAFのモデリング層 M の代わりに、Encoding layer のパッセージのコンテキスト行列 H を使用する。

実験結果

質問の復元実験における実験結果を表4.2に示す。ここで、各モデルのCQs, AQsに該当するカラムの結果は、入力圧縮質問または要約質問と、正解であるオリジナル質問を比較した結果である。

SQuAD, NewsQAのどちらのデータセットにおいても、精度を除くすべての項目で提案手法（**proposed**）が他の比較モデルよりも、有意に高いスコアを示している（ t 検定; $p < .01$ ）。精度のみ、BiDAFを使用していないモデルのほうが良い結果となっている。BiDAFを使用し

表 4.2 質問の復元実験に関する実験結果.

	Model	Precision	Recall	F1	Sim
SQuAD	CQs	1.00	0.586	0.715	0.758
	: w/o copy	0.214	0.203	0.205	0.256
	: w/o p_copy	0.346	0.301	0.319	0.361
	: w/o q_copy	0.669	0.539	0.619	0.770
	: w/o BiDAF	0.856	0.678	0.747	0.814
	: Proposed	0.824	0.718	0.758	0.834
	AQs	1.00	0.626	0.754	0.794
	: w/o copy	0.265	0.282	0.260	0.345
	: w/o p_copy	0.365	0.363	0.353	0.436
	: w/o q_copy	0.707	0.570	0.647	0.621
	: w/o BiDAF	0.900	0.650	0.735	0.810
	: Proposed	0.844	0.716	0.759	0.837
NewsQA	CQs	1.00	0.581	0.715	0.715
	: w/o copy	0.281	0.231	0.250	0.232
	: w/o p_copy	0.702	0.597	0.637	0.695
	: w/o q_copy	0.623	0.541	0.568	0.658
	: w/o BiDAF	0.832	0.631	0.711	0.738
	: Proposed	0.805	0.690	0.731	0.766

ていないモデルは、精度が高い一方で、再現率が低くなっている。これは、比較的短い改訂質問しか生成できていないことを表している。しかし、意味的類似度（Sim）では、提案手法のほうが高いスコアになることから、提案手法が最もオリジナル質問に近い改訂質問を生成できているといえる。

実験結果では、コピーを使用していない場合、入力 of 圧縮質問や要約質問よりも低いスコアとなっている。このことから、改訂質問生成において、コピー機構が重要な役割を果たしていることがわかる。SQuAD では、パッセージからのコピーを除外（w/o p_copy）したときのスコアの低下が大きい、これは、表 4.1 に示す通り、SQuAD の質問は比較的長い質問が多く、質問内容もパッセージ内容に沿っていることからパッセージからのコピーが重要な役割を果たしたといえる。一方で、NewsQA では、質問からのコピーを除外（w/o q_copy）したときの影響が大きくなっている。これは、NewsQA の質問が SQuAD ほど長くない（オリジナル質問と圧縮質問の単語長の差が小さい）ことに加え、抽象的な質問が多いため、パッセージからのコピーが SQuAD ほど有効に機能しなかったためであると考えられる。

図 4.6 に、SQuAD の圧縮質問の復元における、欠損単語数による精度、再現率の推移を示す。横軸がオリジナル質問からの欠損単語数、縦軸が精度、再現率を示している。精度はどのモデルにおいても比較的欠損単語数によらず一定値であるが、再現率は、欠損単語数が大きくなるほど低下する。欠損単語数が多いほど、より多くの単語を追加した改訂質問を生成する必

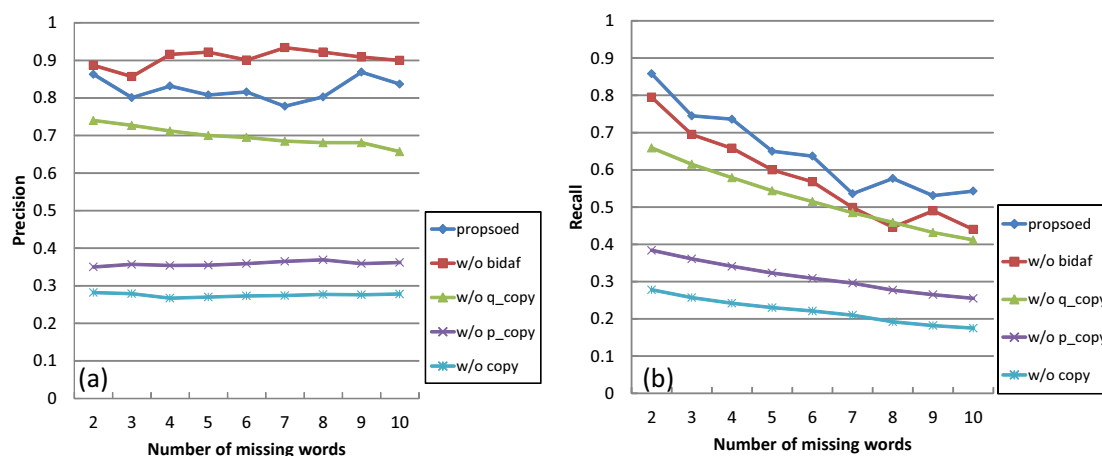


図 4.6 SQuAD の圧縮質問における，単語の削除数毎の質問の復元結果．(a) 精度 and (b) 再現率

表 4.3 改訂質問による機械読解で使用する日本語機械読解データセット．N はパッセージ数，質問数を示しており L はパッセージまたは質問の単語長である．

	train		dev	
	LoQs	ShQs	LoQs	ShQs
jpWiki				
N. passage	1,236	1,236	290	290
N. questions	4,176	4,176	933	933
L. passage	295.3	295.3	300.2	300.2
L. question	18.8	9.7	19.1	9.8

要があるが，提案手法は欠損単語数が 10 単語を超えても，再現率が 5 割を超えており，これは，入力質問が短い場合でも，提案手法は比較的長い改訂質問を生成できることを示している。

4.3.3 改訂質問による機械読解実験

本節では，改訂質問を実際の機械読解モデルに入力したときの，機械読解の回答精度に関する実験を行う．まず，オリジナル質問，圧縮質問，要約質問を機械読解モデルに入力し，出力された回答の正解率を得る．次に，各質問を改訂質問生成モデルに入力し，改訂質問を作成する．作成した改訂質問を機械読解モデルに入力し，回答の正解率を得る．改訂前の質問（オリジナル質問，圧縮質問，要約質問）と改訂質問の機械読解の正解率を比較し，改訂質問による正解率の変化について検証する．また本実験では，日本語の Wikipedia から作成した質問（短文質問，長文質問）を用いた実験も行う．

以降はまず実験で使用するデータセットについて説明する．次に，実験で使用する機械読解モデル，評価尺度について述べ最後に実験結果について詳述する．

データセット

本実験では、英語と日本語の機械読解データセットを使用する。英語の実験には SQuAD を用いる。SQuAD を用いた実験では表 4.1 に示した、SQuAD のテストセットの全 2,067 パッセージ中の 10,570 のオリジナル質問とオリジナル質問から作成した 32,208 の圧縮質問、そしてランダムに抽出した質問から作成した 350 の要約質問を使用する。

日本語の実験として、Wikipedia の日本語版から独自に機械読解のデータセット (jpWiki) を作成した。日本語版 Wikipadia の中からランダムに 600 記事を抽出し、1 記事に対して最大 10 パッセージ程度になるように記事をパッセージ単位に分割した。そして、各パッセージに対し、機械読解用の質問と回答をクラウドソーシングによって作成した。

本実験用の質問として、人手で 1 から作成した短い質問である短文質問も作成する。圧縮質問と要約質問はオリジナル質問から作成した質問である。特に、圧縮質問は文圧縮により機械的に生成した質問であり、非文等も含まれる可能性があるため、実際に質問者が入力する質問と乖離がある可能性がある。そこで、短文質問を作成することで、より実際に質問で改訂質問生成の効果を検証する。意図的に短い質問と長い質問を作成するために、以下の手順により質問と回答を作成した。

1. 作業員 A がある 1 パッセージを読み、質問と回答を作成する。このとき、作業員 A には、できるだけ曖昧な質問を、文字数制限以内 (25 文字) で作成するように指示を行う。
2. 作業員 B は、作業員 A と同じパッセージに対して、作業員 A が作成した質問と回答を閲覧して、回答が同じになり、かつ閲覧した質問よりも具体的になるような質問を作成する。このとき、作業員 B が作成する質問は文字数の下限 (30 文字) を設定し、下限以上の文字数となるようにする。上限は設定しない。

ここで、作業員 A が作成した質問を短文質問 (Short Questions : ShQs)、作業員 B が作成した質問を長文質問 (Long Questions : LoQs) とする。機械読解実験で作成した日本語データセットについて、表 4.3 に示す。作成したデータセットのうち、ランダムに 100 記事を抽出し、抽出した記事に含まれるパッセージと質問、回答をテストセットとし、残りを学習セットとした。学習セットは後述する機械読解モデルおよび改訂質問生成モデルの学習に使用する。なお、jpWiki はデータセットに短文質問と長文質問が含まれるため、機械的に短い質問を生成する文圧縮によるデータ作成は行わない。

機械読解モデル・改訂質問生成モデル

本実験では、学習済みの機械読解モデルを使用する。英語の機械読解モデルとして、Web でオープンに公開されている SQuAD で学習した BiDAF の機械読解モデル [22]^{*4}を使用する。日本語については、[50] が作成した日本語のニュース記事 (jpNews) に関する機械読解データに、表 4.3 の学習セットのデータを加えたデータによって学習した BiDAF モデルを用いる。ここで、機械読解の学習には長文質問を用いた。

英語では 4.3.2 で作成した SQuAD の改訂質問生成モデル (**proposed**) を使用する。日本語の改訂質問モデルとして、jpNews から圧縮質問を作成したものに、jpWiki の学習セットから短文質問を入力、長文質問を教師としたときのデータを加えた学習データ (全 1,673,000 入力質問) を用いて学習を行った。改訂質問生成モデルおよび学習パラメータは 4.3.2 と同様である。

^{*4} <http://allennlp.org/models>

表 4.4 SQuAD における機械読解の回答精度の比較結果、

	OQs			CQs			AQs		
	Success@k	Accuracy	Improvement	Success@k	Accuracy	Improvement	Success@k	Accuracy	Improvement
Input	0.645	0.645	–	0.347	0.347	–	0.445	0.445	–
SQ1	0.663	0.595	0.0520	0.482	0.389	0.207	0.501	0.455	0.108
SQ2	0.684	0.555	0.0811	0.519	0.374	0.194	0.542	0.422	0.134
SQ3	0.698	0.544	0.0869	0.541	0.363	0.189	0.571	0.402	0.134
SQ4	0.710	0.542	0.100	0.558	0.363	0.191	0.580	0.403	0.160
SQ5	0.721	0.522	0.0957	0.571	0.343	0.181	0.589	0.420	0.134

表 4.5 jpWiki における機械読解の回答精度の比較結果.

	LoQs			ShQs		
	Success@k	Accuracy	Improvement	Success@k	Accuracy	Improvement
Input	0.668	0.668	–	0.509	0.509	–
SQ1	0.680	0.584	0.0355	0.628	0.547	0.242
SQ2	0.700	0.553	0.0710	0.694	0.535	0.306
SQ3	0.721	0.507	0.0903	0.729	0.537	0.308
SQ4	0.734	0.473	0.0903	0.744	0.482	0.249
SQ5	0.744	0.437	0.100	0.750	0.423	0.236

評価尺度

機械読解の回答の正解率を測る指標として、完全一致率を使用する。完全一致率は、機械読解モデルが出力した回答と、正解データの回答の文字列が完全一致した質問の割合である。機械読解の評価では、文字列の部分一致の指標（F 値）を用いる場合もあるが、本実験では、改訂質問により、回答の特定性が高まることを検証するため、より厳密に正解を判定できる完全一致率を採用する。

本研究が提案する改訂質問生成では、1つの入力質問に対して、複数の改訂質問を提示する。本実験では、改訂質問生成モデルのビームサーチのビーム幅を5に設定し、5つの改訂質問を生成したときの完全一致率を計測する。複数の改訂質問の完全一致率を評価する指標として、**success@k**, **accuracy**, **improvement**を導入する。success@kは、評価セットの全質問中、入力質問と上位k位までの改訂質問のうちどれか一つでも正解と完全一致となった回答が出力された質問の割合である。accuracyは、入力質問および各順位の改訂質問単体での完全一致率である。improvementは、入力質問の機械読解の回答が不正解のときの、各順位の改訂質問の回答結果のaccuracyである。success@kとimprovementにより、本来の入力質問では完全一致できなかった質問について、改訂質問により完全一致可能となった質問の割合を計測する。

実験結果

改訂質問による機械読解実験の結果を表4.4, 表4.5に示す。SQuADデータにおける実験結果である表4.4では、圧縮質問（CQs）または要約質問（AQs）を機械読解に入力したときの完全一致率（accuracy）は、オリジナル質問（OQs）を入力したときから低下している。各質問から改訂質問を作成した後の機械読解では、圧縮質問と要約質問の第1位の改訂質問（SQ1）が高いaccuracyとなっている。一方で、オリジナル質問を入力したとき、改訂質問はいずれの順位においてもaccuracyが低下している。しかし、入力質問および改訂質問のいずれかで完全一致となるsuccess@kでは、いずれの入力質問の場合でも、入力質問のみのとき

なお、ちり・ほこり・しみの主な原因は、人間、ペット、観葉植物などやその活動である。また、衣服・カーテン・カーペット類の劣化による繊維ごみもある。人やペットの体毛、皮脂、汗、糞尿。また食品の残滓やハネ。他にもインク類のハネこぼれや子供の落書き、区域外から持ち込まれたり舞い込んだりした土埃、花粉などがある。ちり・ほこり・しみ等は、ダニ・ノミ・雑菌類の繁殖培地となり、虫刺され、かぶれ、喘息や、各種炎症・感染症の原因となるほか、カビの発生、腐敗等により悪臭の原因や什器類の劣化の原因となる。地方によっては蚊の発生がマラリアなどの深刻な感染症の原因となるため、観葉植物用の灌水容器、屋外の雨どいや排水設備、防火水槽、エアコン室外機などに溜まった雨水などにボウフラが発生しないよう、とりわけ注意が必要である。電気まわりの汚れ・粉塵などは絶縁不良、トラッキングを引き起こし、火災の原因となるので掃除を行う。コンピュータやテレビ、換気扇など電気製品、作業機械や工作機械などが粉塵に汚染された状態で放置されていると排熱不良や稼働部の消耗、咬合不良などにより故障の原因となるので掃除を行う。

Ground-truth answers: 蚊の発生	
Short and specific questions (ShQ, SQ)	Answers from BiDAF
ShQ1: 何が深刻な感染症の原因となる？	カビの発生、腐敗等
SQ1: 地方によっては何がマラリアなどの深刻な感染症の原因となる？	蚊の発生

■ Copied from passage
 ■ Copied from question
 ■ Generated

図 4.7 jpWiki における改訂質問による機械読解の例。

Following their loss in the divisional round of the previous season 's playoffs , the Denver Broncos underwent numerous coaching changes , including a mutual parting with head coach John Fox (who had won four divisional championships in his four years as Broncos head coach) , and the hiring of Gary Kubiak as the new head coach. (..)

Ground-truth answers: Gary Kubiak, Gary Kubiak, Kubiak	
OQ: who is the head coach of the broncos ?	
Original and specific questions (OQ, SQ)	Answers from BiDAF
SQ1: who is the head coach of the denver broncos ?	John Fox
SQ2: who is the head coach of the broncos ?	John Fox
SQ3: who is the head coach of the denver ?	John Fox
SQ4: who is the new head coach of the broncos ?	Gary Kubiak
SQ5: who is the new head coach of the denver broncos ?	Gary Kubiak

■ Copied from passage
 ■ Copied from question
 ■ Generated

図 4.8 SQuAD における改訂質問による機械読解の例。‘(...)’は省略したパッセージを表している。

から success@5 で 10 ポイント以上の向上が見られた。また、入力質問が完全一致しないときの、各順位での改訂質問の accuracy である improvement では、圧縮質問が順位の高い改訂質問で高いスコアとなった一方で、オリジナル質問と要約質問では、下位の改訂質問のスコアが高くなるという結果となった。

日本語データである jpWiki を用いた実験結果（表 4.5）においても、SQuAD と同様の傾向が見られた。短文質問を入力した際には、短文質問そのものよりも、短文質問から作成した改訂質問（SQ1）の方が accuracy が高くなった。一方で、長文質問から作成した改訂質問では accuracy が低下する結果となった。success@k において短文質問を入力したとき、success@2 で長文質問の accuracy を上回り、success@5 では 75% の正解率となっている。

機械読解の回答精度に関して、圧縮質問（CQs）、要約質問（AQs）、短文質問（ShQs）と、

改訂質問 (SQ1) の accuracy について、正解・不正解に関する McNemar 検定を実施した。圧縮質問と短文質問については、それぞれ $p < .01$, $p < .05$ で SQ1 との accuracy に有意差があることを確認した。一方、要約質問については、SQ1 との accuracy の差について、 $p > .1$ で有意な差があるとは認められなかった。そこで要約質問について、SQ1 と SQ2 のいずれかが正解であるときの accuracy と比較した結果 $p < .05$ で有意差があることが認められた。

図 4.7 は jpWiki の短文質問について、機械読解と改訂質問生成を適用した例である。短文質問を入力したとき、機械読解は比較的正解に近い回答を抽出しているが、改訂質問生成により“マラリア”など特徴的な単語を含んだ改訂質問に変換されたことで、本来の正解となる回答を機械読解が抽出できるようになっている。

図 4.8 は、SQuAD のオリジナル質問を機械読解と改訂質問生成に入力したときの結果である。1 位の改訂質問 (SQ1) では、チーム名が正式名称である“denver broncos”に変化しているが、回答は“John Fox”のまま変化していない。しかし、4, 5 位での改訂質問では“the new head coach”という情報が追加されたことにより機械読解の回答が“Gary Kubiak”に変化している。

4.4 考察

4.4.1 短文質問に対する改訂質問生成に関する考察

英語 (SQuAD)、日本語 (jpWiki) どちらのデータセットにおいても、圧縮質問、要約質問、短文質問を機械読解に入力すると、オリジナル質問、長文質問を入力したときよりも回答精度 (accuracy) が低下している。このことから、機械読解では、質問が短かかったり、質問内に含まれる情報が不十分なとき、十分な性能が発揮できていないといえる。圧縮質問・要約質問 (短文質問) から作成した改訂質問により、機械読解の回答精度が改善されている。これは改訂質問生成モデルが、曖昧な入力質問の質問意図を適切に読み取り、回答に必要な情報を補完できているからであると考えられる。一方、入力質問が短いほど、改訂質問生成による具体化のバリエーションは増加すると考えられる。圧縮質問・要約質問 (短文質問) のように入力質問が短い場合、改訂質問を 1 つ提示するだけでも、入力質問による機械読解よりも高い回答精度が得られることが期待できるが、必ずしも質問者の質問意図どおりに質問が改訂されるとは限らないため、改訂質問生成を用いる場合、2 個以上の改訂質問を提示し、質問者に選択させるほうが望ましいと考えられる。

4.4.2 長文質問に対する改訂質問生成に関する考察

オリジナル質問や長文質問から改訂質問を作成した場合、機械読解の回答精度 (accuracy) は、入力質問をそのまま使用したときよりも低下している。これは、改訂質問が回答に不必要

な情報を追加してしまったり、回答自体を改訂質問に追加してしまっただけが原因であると考えられる。一方で、入力質問を使用したときの機械読解の回答が不正解だったとき、ビームサーチにより生成した下位の改訂質問の方が回答精度の向上に貢献 (improvement) した。上位で生成する改訂質問は、入力質問の質問意図を補強するような改訂質問を生成するが、下位の改訂質問で、異なる回答となる可能性がある改訂質問を提示していると考えられる。

4.5 まとめ

本章では、改訂質問生成を用いた、機械読解による質問応答手法を提案した。改訂質問生成では、入力質問をパッセージの内容を元に具体化し、改訂質問として質問者に提示する。質問者は、曖昧であったり短い質問を質問応答に入力した場合でも、情報が補完された改訂質問の候補から、自身の質問意図に近い改訂質問を選択することで、質問意図に適合する回答を得ることができる。

機械読解コーパスに提案手法を適用した評価実験では、短い質問を入力した際には、改訂質問生成モデルが出力する改訂質問を使用することにより、機械読解の回答精度を改善できることを示した。また、十分に具体的な質問であっても、改訂質問生成が異なる質問意図の改訂質問を生成するため、自身の質問意図に沿った質問に対する回答を選択可能になることを確認した。

今後は、改訂質問の生成精度の向上や、機械読解と改訂質問生成モデルを End-to-End で行うモデルの検討を行う予定である。

第 5 章

考察

5.1 オープンナレッジにおける情報要求の言語化

オープンナレッジにおける情報要求の言語化では，コミュニティ QA の質問記事を言語化された情報要求とみなすことで，多様な情報要求に対する拡張クエリを提示する手法を提案した．評価実験では，特に同音異義語に対しても，提案手法が既存手法よりもより多様な Web 検索を可能にする拡張クエリを提示できることを明らかにしている．提案手法は，カテゴリや季節を切り替えることで，多様な拡張クエリによる検索を可能にしているが，その中には，予測しにくいキーワードが追加されるケースが多く発見された．しかし，拡張クエリと共に提示される質問記事を閲覧することで，拡張クエリの背後にある情報要求がコンテキストを含めて理解できるため，利用者は当初は理解不能だった拡張クエリに対しても，ある程度の理解を持って検索を行えるという利点があった．

情報要求や意図を理解しながら多様な情報を入手できるという点は，近年においては特に非常に重要な観点といえる．Web 検索を使用することで，利用者は自身の所望の情報を入手することができるが，それは，自身が望む情報しか取得しないという危険性を持つ．しかしながら，利用者自身が想起できない情報は，検索クエリを作成することができないため，“利用者が意図しないような”情報を意図して入手することは非常に難しいといえる．しかし，本研究の提案手法は，1つのキーワードから多様な検索を可能にするため，“想起できないクエリは作成できない”という最初の障壁を1段階低くできるものであるといえる．また，提案手法は，コミュニティ QA を使用することでソーシャル性の高いことも特徴といえ，拡張クエリの季節性などに反映されている．

今後の課題として，Web 検索とコミュニティ QA の質問記事の性質の違いに着目する必要があると考えている．コミュニティ QA の質問記事は自然言語で記述でき，文字数制限も比較的ゆるいため，質問記事では，質問者の置かれている状況まで含めた非常に詳細な情報が記載されていることが多い．このような質問記事から拡張クエリを作成した場合，質問の特定性

が高すぎるものが起因し、Web 検索においても有用な結果が得られない場合がある。今後は、質問記事から直接拡張クエリを作成するのではなく、複数の質問記事を集約し、その中から、典型的な質問意図を抽出するなど、質問記事自体を加工することも検討している。

5.2 クローズドナレッジにおける情報要求の言語化

クローズドナレッジに対する情報要求の言語化では、機械読解を対象に、入力質問を具体化する改訂質問生成により、回答の特定性を高める質問を提示する手法を提案した。提示された改訂質問は、オープンナレッジの拡張クエリとは異なり、入力質問をベースとし、回答が変化する可能性の高い単語を対象のパスセージからピンポイントで抽出し追加するという特徴があった。追加される語は、固有表現を始めとする名詞が多いという特徴を持っていた。機械読解は、検索とは異なり、自然言語形式で質問文を入力する。自然言語では、話し言葉において、主語などの情報を落とした文を作成してしまう場合が多いというゼロ代名詞問題 [82] が知られているが、質問応答も、インターフェースとして、音声入力やチャット形式を採用することが多いため、同様の問題が発生しやすいと考えられる。このような問題に対して、提案手法の改訂質問生成では、利用者が無意識のうちに省略してしまった情報を補完することで回答の特定性を高めることができるため、改訂質問生成は質問応答システムと組み合わせることで高い効果を発揮できると考えている。

今後の課題として、改訂質問を出力するための判断を行うための機能が必要であると考えている。本研究で提案した改訂質問生成では、決まった数の改訂質問を生成する。しかし、実際には、回答の特定性を高めるために必要となる改訂質問の数は、入力とパスセージによって異なる。必要以上の選択肢を利用者に提示することは、利用者の混乱につながるため、入力に応じて、改訂質問生成が必要であるか、また必要な場合の改訂質問の数はいくつであるかを判断する機能を追加する必要があると考えている。

5.3 オープン/クローズドにおける情報要求の言語化の違いと今後の展望

本論文では、オープンナレッジとクローズドナレッジでは、情報アクセスの手段が異なると仮定し、オープンナレッジでは Web 検索、クローズドナレッジでは機械読解を対象に情報要求の言語化支援技術の研究に取り組んだ。オープンナレッジでは、コミュニティ QA の多様な質問を知識源とすることで、多様な情報要求に対する Web 検索用の拡張クエリを生成した。一方でクローズドナレッジでは、機械読解に入力する文書中の回答の特定性を高めるための改訂質問を生成した。

今後の研究課題として、それぞれで取り組んだ課題の融合が考えられる。情報アクセスに基

づいて、Web 上の情報にアクセスしたいと考えると、まず、検索エンジンを使って所望の情報が含まれる Web ページを検索する。次に、ヒットした Web ページ内を探索し、所望の情報が含まれている箇所を発見する。これは、1 回の情報アクセスにおいてオープン/クローズドのどちらの知識源にもアクセスにすることを意味している。

本論文では、クローズドナレッジにおける情報アクセスでは、オープンナレッジのようにコミュニティ QA の多様な知識を用いずに、機械読解にされる文章の中の知識のみを用いている。これは、クローズドナレッジにおいては、探索範囲の文章に含まれない情報は、回答抽出の役に立たずノイズとなってしまう可能性が高いことを仮定している。しかしながら、コミュニティ QA のような大量の知識は、本論文が対象とする機械読解をはじめ、深層学習の自然言語処理の応用には非常に効果が高いことが知られている。このようなオープンナレッジの知識を使いつつ、クローズドナレッジの範囲に適用するためには、例えば、大塚ら [86] の研究のように、生成モデルの学習過程あるいは、処理過程で単語や情報に関する制約を与えることで、必要な知識のみを活用できる可能性があるのではないかと考えている。

オープンナレッジにおける情報要求の言語化では、コミュニティ QA の質問記事からの拡張クエリはクエリ尤度モデルによる単語の重要度に基づいて生成される。本手法を採用した理由は、まず対象がオープンナレッジである Web 検索であったため、質問記事の内容を表すキーワード抽出を重視したためである。また、クローズドナレッジでの情報要求の言語化に適用した深層学習は、処理速度やハードウェアリソースの観点から、Web 検索のような大規模検索には適していない [50] という課題があったためである。オープンナレッジにおける情報要求言語化支援のシステムでは、作成した拡張クエリとともに、拡張クエリを作成した基となる質問記事を一緒に提示していた。しかし、質問記事には質問投稿者の個人的背景等の Web 検索等に一般化する上では不要な情報も含まれており、拡張クエリと見比べても可読性がよくない質問記事が多く含まれていた。そこで、質問記事から質問の主題を推測し、要約する形で質問を生成・提示できれば図 1.1 で示したような情報要求のみを残したシンプルな質問を提示できる。多様な質問記事から観点を抽出して要約を行うことは、自然言語処理技術においても重要かつ困難な課題として知られていた、しかしながら、近年では、BERT[16] をはじめとする深層学習モデルが、比較的複雑な文章であっても高い精度で読み解くことを可能にしており、それらを活用した高精度の翻訳モデル [76] も登場しており、これらを応用することで、質問記事から質問の主題のみを抽出した簡潔な質問文が生成することが期待できると考えている。また、先に述べたクローズドナレッジの制約を加えることで、オープン/クローズドをまとめて考慮した情報要求の言語化支援手法が実現できるのではないかと考えている。

以上をまとめると、本論文では、オープンナレッジとクローズドナレッジそれぞれにおいて、別のタスクを設定し、それぞれで課題解決の手法を提案したが、オープンナレッジでは、情報要求の知識源であるコミュニティ QA、クローズドナレッジでは、高い言語理解性能を有する深層学習モデルと今後重要となるキーコンポーネントを用いており、今後はそれらを融合

させ、オープン/クローズドを横断するような情報アクセスにおいても適用可能な情報要求の言語化支援技術が実現できる可能性を提示できたものと考えている。

第 6 章

結論

6.1 本研究のまとめ

本研究では、情報アクセスにおける利用者自身の“何を知りたいか”という情報要求に着目した、情報要求の言語化支援技術に取り組んだ。情報要求の言語化では、利用者が入力した曖昧な検索クエリや質問文を、より具体的な内容にした自然言語の文を提示することで、利用者に自身の情報要求に近い候補を選択させる手法を採用した。利用者が Web 検索のような膨大な情報から情報を取得する場合と、マニュアルのような特定のテキストから情報を取得する場合では、利用者に提示すべき情報要求は異なると考え、前者をオープンナレッジ、後者をクローズドナレッジとして、異なるアプローチで言語化された情報を生成し、利用者に提示する手法を提案した。

オープンナレッジにおける情報要求の言語化では、Web 検索を対象に、多様な情報要求に対して検索を実行するための拡張クエリの作成手法を提案した。コミュニティ QA の質問記事から拡張クエリを作成することで、多様な拡張クエリを提示可能にしている。ここで、拡張クエリを作成したコミュニティ QA の質問記事を言語化された情報要求とみなし、拡張クエリとセットで提示することで、利用者が意図しなかった拡張クエリに対しても、付属している質問記事を閲覧することで、質問のコンテキストを理解できるようになる。拡張クエリの Web 検索結果から取得した Web ページのユニーク語数を調べることで、検索結果の多様性を評価する実験を行った。実験の結果、コミュニティ QA から作成した拡張クエリは、キーワードの関連語から作成した拡張クエリよりも多様な Web 検索結果を得られることを明らかにした。また、コミュニティ QA の出現単語の頻度について調査し、コミュニティ QA に投稿される質問記事には季節性があることを明らかにした。そして、時系列の変化を追跡できるトピックモデルを用いて、カテゴリごとに投稿される質問記事の中から、季節的な周期性をもつトピックが抽出できることを示した。調査結果を元に、コミュニティ QA の質問記事をカテゴリと投稿時期で分割し、それぞれを異なるデータセットとして、拡張クエリを生成するための

モデルを学習し、カテゴリと季節を切り替えることで、異なる拡張クエリが提示されるタグクラウド型の Web 検索システムを実装した。

クローズドナレッジにおける情報要求の言語化では、深層学習による質問応答技術である機械読解を対象に、入力された質問文をより具体化する改訂質問生成について取り組み、改訂質問生成を実現するための深層学習モデルを提案した。提案モデルは、深層学習の文生成モデルに機械読解の機能とコピー機構を組み合わせたモデルであり、入力された質問文とパッセージを読み解き、質問文に不足している情報をパッセージから抽出することで、入力された質問文をパッセージの内容に即して具体化した改訂質問を生成している。同時に、提案モデルを学習するための学習データの自動作成手法についても提案した。提案手法では、機械読解の学習コーパスをベースとして、コーパスに含まれる質問文を、情報の不足がない教師データとなる質問文とみなす。そして、文の意味を変えずに文の長さを短縮する文圧縮によって、機械読解コーパスの質問文を短くする。文圧縮後の質問文を入力、文圧縮前の質問文を教師データとして、質問を復元するような学習を行うことで、不足している情報を保管する改訂質問の生成モデルの学習を行った。評価実験では、まず提案モデルが正しく不足している情報を保管できているかを評価するために、質問の復元実験を行った。復元実験では、機械読解のオープンデータセットに対して、質問文から意図的に情報を欠落させた質問を作成し、この質問を入力したときに正しく復元できた質問の調査した。調査の結果、提案モデルは多くの質問で8割～9割程度、質問を復元できており、必要な情報を補完できていることを明らかにした。生成した改訂質問を機械読解に入力したときの、機械読解の回答精度の比較実験では、短い質問を入力したとき、機械読解の回答精度は低下するが、改訂質問生成によって生成した質問を再度機械読解に入力することで、機械読解の回答精度が改善されることを明らかにした。

謝辞

本論文は、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および博士後期課程に在籍中の研究成果をまとめたものです。また、本論文は日本電信電話株式会社 NTT メディアインテリジェンス研究所との共同研究による成果を含みます。本論文の作成にあたり、多数の皆様からのご指導とご支援をいただきました。

博士前期課程および博士前期課程で主指導教員としてご指導をいただきました、図書館情報メディア系の佐藤哲司 教授には、研究指導のみならず社会の先輩として、博士号を取得することの意義や、学位取得後の心構えや考え方、キャリアについてもご教示いただきました。謹んで感謝を申し上げます。また、副指導教員である同系の芳鐘冬樹 教授、手塚太郎 准教授には、私のスケジュールの希望に合わせてゼミを開催いただき、その中で研究や論文についての議論を行うことができました。深く感謝申し上げます。更に、学位論文審査委員としてご指導いただいた、松本紳 教授、加藤誠 准教授ならびにシステム情報系の宇津呂武仁 教授にもここに感謝申し上げます。ご指導いただいた内容を今後も活かしながら、今後の研究を進めてたいと考えております。

本論文に関わる共著者である、国立情報学研究所の神門典子 教授には、研究内容のみならず英語プレゼンテーションの練習など様々な面でご支援いただきました。また、図書館情報メディア系の関洋平 准教授にも論文についての詳細な添削やコメントにより論文の精度向上に尽力くださり、投稿した論文を採録させることができました。改めまして深く感謝申し上げます。

本論文の作成にあたり、日本電信電話株式会社の皆様にも大変お世話になりました。特に松尾義博グループリーダー（現 NTT アドバンステクノロジー株式会社）、富田準二グループリーダーには本研究の推進に関するご理解とご支援をいただきました。また、飯塚哲也 NTT メディアインテリジェンス研究所所長をはじめ、NTT メディアインテリジェンス研究所企画部、NTT 知的財産センタの皆様には、本研究を進めるにあたっての共同研究契約および業務

調整等でご理解とご支援をいただきました。本研究を進めるにあたり、共著者として研究を共に進めてまいりました、西田京介 特別研究員、浅野久子 主幹研究員、斎藤いつみ 研究員に感謝申し上げます。業務と学業の両立にあたっては、NTT メディアインテリジェンス研究所社会言語処理グループをはじめ、社会知識処理プロジェクト、心理情報処理プロジェクトの皆様には多大なご支援をいただきました。ここに感謝申し上げます。また、入社から指導者として私が研究者として独り立ちできるよう積極的にご指導くださり、博士号取得も勧めてくださった平野徹 氏（現株式会社デンソー）にも感謝申し上げます。

研究を進めるにあたり、友人の支えは必要不可欠でした。博士前期課程の研究室の同期メンバーである渡邊直人 氏、香川雄一 氏をはじめ佐藤研究室のメンバー一同に感謝申し上げます。また、輪島幸治 氏、苅米志帆乃 氏には、同じ社会人の博士後期課程在籍者として仕事のスケジュールを調整し、テレビ会議でのゼミで夜遅くまで議論に参加いただいたこと感謝申し上げます。

最後に、私を博士前期課程まで進学させてくれたと同時に、社会人になった後に博士後期課程に在籍して博士号を取得することについて応援してくださった両親をはじめ家族、親戚一同に心より感謝申し上げます。ありがとうございました。

たくさんの方々のご支援により、博士論文をまとめることができました。
ありがとうございました。

参考文献

- [1] A. Adamic, Jun Zhang, Eytan Bakshy, and Mark S. Ackerman. Knowledge Sharing and Yahoo Answers: Everyone Knows Something. In *Proceedings of the 17th International Conference on World Wide Web(WWW2008)*, pp. 665–674, 2008.
- [2] Wasi Uddin Ahmad, Kai-Wei Chang, and Hongning Wang. Context attentive document ranking and query suggestion. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR2019)*, p. 385394, 2019.
- [3] Hiteshwar Kumar Azad and Akshay Deepak. Query expansion techniques for information retrieval: A survey. *Information Processing & Management*, Vol. 56, No. 3, pp. 1698–1735, 2019.
- [4] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In *International Conference on Learning Representations (ICLR)*, 2015.
- [5] David M. Blei and John D. Lafferty. Dynamic topic models. *Proceedings of the 23rd International Conference on Machine learning(ICML'06)*, pp. 113–120, 2006.
- [6] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent Dirichlet Allocation. *The Journal of Machine Learning Research*, Vol. 3, pp. 993–1022, 2003.
- [7] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics (TACL)*, Vol. 5, pp. 135–146, 2017.
- [8] Eric Brill, Susan T. Dumais, and Michele Banko. An analysis of the askmsr question-answering system. In *Proceedings of the 2002 Empirical Methods in Natural Language Processing (EMNLP2002)*, 2002.
- [9] Christian Buck, Jannis Bulian, Massimiliano Ciaramita, Wojciech Gajewski, Andrea Gesmundo, Neil Houlsby, and Wei Wang. Ask the right questions: Active question reformulation with reinforcement learning. In *Proceedings of the Sixth International*

- Conference on Learning Representations (ICLR2018)*, 2018.
- [10] Huanhuan Cao, Daxin Jiang, Jian Pei, Qi He, Zhen Liao, Enhong Chen, and Hang Li. Context-Aware Query Suggestion by Mining Click-Through and Session Data. In *Proceedings of the 14th ACM international conference on Knowledge discovery and data mining(KDD2008)*, pp. 857–883, 2008.
- [11] Ziqiang Cao, Chuwei Luo, Wenjie Li, and Sujian Li. Joint copying and restricted generation for paraphrase. In *Association for the Advancement of Artificial Intelligence (AAAI)*, pp. 3152–3158, 2017.
- [12] Daniel Cer, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St. John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, Yun-Hsuan Sung, Brian Strope, and Ray Kurzweil. Universal sentence encoder. *CoRR*, Vol. abs/1803.11175, , 2018.
- [13] Nan Yang Bowen Du Weifeng Lv Ming Zhou Chuanqi Tan, Furu Wei. S-net: From answer extraction to answer synthesis for machine reading comprehension. In *Proceedings of the Thirty-Second Conference on Association for the Advancement of Artificial Intelligence (AAAI2018)*, pp. 5940–5947, 2018.
- [14] Junyoung Chung, Çağlar Gülçehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *CoRR*, Vol. abs/1412.3555, , 2014.
- [15] Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, Shijin Wang, and Guoping Hu. Cross-lingual machine reading comprehension. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP 2019)*, pp. 1586–1595, 2019.
- [16] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, June 2019.
- [17] Li Dong, Jonathan Mallinson, Siva Reddy, and Mirella Lapata. Learning to paraphrase for question answering. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP2017)*, pp. 875–886, 2017.
- [18] Xinya Du and Claire Cardie. Harvesting paragraph-level question-answer pairs from wikipedia. In *Proceedings of the 56th Association for Computational Linguistics (ACL2018)*, pp. 1907–1917, 2018.

- [19] Xinya Du, Junru Shao, and Claire Cardie. Learning to ask: Neural question generation for reading comprehension. In *Proceedings of the 55th Association for Computational Linguistics (ACL2017)*, pp. 1342–1352, 2017.
- [20] Hady ElSahar, Christophe Gravier, and Frédérique Laforest. Zero-shot question generation from knowledge graphs for unseen predicates and entity types. In *Proceedings of the 2018 North American Association for Computational Linguistics (NAACL)*, pp. 218–228, 2018.
- [21] Katja Filippova and Michael Strube. Dependency tree based sentence compression. In *International Natural Language Generation Conference (INLG)*, pp. 25–32, 2008.
- [22] Matt Gardner, Joel Grus, Mark Neumann, Oyvind Tafjord, Pradeep Dasigi, Nelson F. Liu, Matthew E. Peters, Michael Schmitz, and Luke Zettlemoyer. Allennlp: A deep semantic natural language processing platform. *CoRR*, Vol. abs/1803.07640, , 2018.
- [23] Jiafeng Guo, Xueqi Cheng, Gu Xu, and Huawei Shen. A Structured Approach to Query Recommendation with Social Annotation Data. In *Proceedings of the 19th ACM international conference on Information and knowledge management (CIKM2010)*, pp. 619–628, 2010.
- [24] Ankush Gupta, Arvind Agarwal, Prawaan Singh, and Piyush Rai. A deep generative framework for paraphrase generation. In *Proceedings of the Thirty-Second Conference on Association for the Advancement of Artificial Intelligence (AAAI2018)*, pp. 5149–5156, 2018.
- [25] Fred.X Han, Di Niu, Kunfeng Lai, Weidong Guo, Yancheng He, and Yu Xu. Inferring search queries from web documents via a graph-augmented sequence to attention network. In *The World Wide Web Conference (WWW2019)*, p. 27922798, 2019.
- [26] Yusef Hassan-montero and Victor Herrero-solana. Improving Tag-Clouds as Visual Information Retrieval Interfaces. *International Conference on Multidisciplinary Information Sciences and Technologies(InScit2006)*, pp. 25–28, 2006.
- [27] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *IEEE International Conference on Computer Vision (ICCV)*, pp. 1026–1034, 2015.
- [28] Marti Hearst. Design Recommendations for Hierarchical Faceted Search Interfaces. In *Proceedings of the ACM SIGIR Workshop on Faceted Search*, 2006.
- [29] Karl Moritz Hermann, Tomáš Kočiský, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. Teaching machines to read and comprehend. In *International Conference on Neural Information Processing Systems (NIPS)*, pp. 1693–1701, 2015.

- [30] Sepp Hochreiter and Jrgen Schmidhuber. Long short-term memory. *Neural Computation*, Vol. 9, No. 8, pp. 1735–1780, 1997.
- [31] Aaron Jaech and Mari Ostendorf. Personalized language model for query auto-completion. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL2018)*, pp. 700–705, 2018.
- [32] Unnat Jain, Svetlana Lazebnik, and Alexander G. Schwing. Two can play this game: Visual dialog with discriminative question generation and answering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR2018)*, pp. 5754–5763, 2018.
- [33] Jyun-Yu Jiang and Wei Wang. Rin: Reformulation inference network for context-aware query suggestion. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM2018)*, pp. 198–206, 2018.
- [34] Yimin Jing, Deyi Xiong, and Zhen Yan. BiPaR: A bilingual parallel dataset for multilingual and cross-lingual reading comprehension on novels. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP2019)*, pp. 2452–2462, 2019.
- [35] Tomasz Jurczyk, Michael Zhai, and Jinho D. Choi. Selqa: A new benchmark for selection-based question answering. In *Proceedings of the IEEE 28th International Conference on Tools with Artificial Intelligence*, pp. 820–827, 2016.
- [36] Makoto P. Kato, Tetsuya Sakai, and Katsumi Tanaka. When do people use query suggestion? a query suggestion log analysis. *Inf. Retr.*, Vol. 16, No. 6, p. 725746, 2013.
- [37] Jonathan Koren, Yi Zhang, and Xue Liu. Personalized Interactive Faceted Search. In *Proceedings of the 17th International Conference on World Wide Web(WWW2008)*, pp. 447–486, 2008.
- [38] Zhen-Zhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. Albert: A lite bert for self-supervised learning of language representations. *arXiv*, Vol. abs/1909.11942, , 2019.
- [39] Ruirui Li, Liangda Li, Xian Wu, Yunhong Zhou, and Wei Wang. Click feedback-aware query recommendation using adversarial examples. In *The World Wide Web Conference (WWW2019)*, p. 29782984, 2019.
- [40] Jimmy Lin. An exploration of the principles underlying redundancy based factoid question answering. *ACM Transactions on Information Systems (TOIS)*, Vol. 25, No. 2, p. 6, 2007.

- [41] Yuan Lin, Song Jin, Hongfei Lin, Yunlong Ma, and Kan Xu. Social Annotation in Query Expansion a Machine Learning Approach. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in information retrieval(SIGIR2011)*, pp. 405–414, 2011.
- [42] Qiaoling Liu, Eugene Agichtein, Gideon Dror, Yoelle Maarek, and Idan Szpektor. When Web Search Fails, Searchers Become Askers: Understanding the Transition. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval(SIGIR2012)*, pp. 801–810, 2012.
- [43] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schuetze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.
- [44] Yajie Miao, Chunping Li, Jie Tang, and Lili Zhao. Identifying New Categories in Community Question Answering Archives: A Topic Modeling Approach. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management(CIKM2010)*, pp. 1673–1676, 2010.
- [45] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems (NIPS2013)*, pp. 3111–3119. 2013.
- [46] Guangtao Wang Jing Huang Xiaodong He Bowen Zhou Ming Tu, Kevin Huang. Select, answer and explain: Interpretable multi-hop reading comprehension over multiple documents. In *Proceedings of the Thirty-Fourth Conference on Association for the Advancement of Artificial Intelligence (AAAI2020)*, 2020.
- [47] Nasrin Mostafazadeh, Ishan Misra, Jacob Devlin, Margaret Mitchell, Xiaodong He, and Lucy Vanderwende. Generating natural questions about an image. In *Proceedings of the 54th Association for Computational Linguistics (ACL2016)*, pp. 1802–1813, 2016.
- [48] Kosuke Nishida, Kyosuke Nishida, Masaaki Nagata, Atsushi Otsuka, Itsumi Saito, Hisako Asano, and Junji Tomita. Answering while summarizing: Multi-task learning for multi-hop QA with evidence extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL2019)*, pp. 2335–2345, 2019.
- [49] Kyosuke Nishida, Itsumi Saito, Kosuke Nishida, Kazutoshi Shinoda, Atsushi Otsuka, Hisako Asano, and Junji Tomita. Multi-style generative reading comprehension. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL2019)*, pp. 2273–2284, 2019.
- [50] Kyosuke Nishida, Itsumi Saito, Atsushi Otsuka, Hisako Asano, and Junji Tomita.

- Retrieve-and-read: Multi-task learning of information retrieval and reading comprehension. In *ACM International Conference on Information and Knowledge Management (CIKM)*, pp. 647–656, 2018.
- [51] Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP2014)*, pp. 1532–1543, 2014.
- [52] Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL2018)*, pp. 2227–2237, 2018.
- [53] Pranav Rajpurkar, Robin Jia, and Percy Liang. Know what you don’t know: Unanswerable questions for SQuAD. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL2018)*, pp. 784–789. Association for Computational Linguistics, July 2018.
- [54] Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. Squad: 100,000+ questions for machine comprehension of text. In *Empirical Methods in Natural Language Processing (EMNLP)*, pp. 2383–2392, 2016.
- [55] Joseph Reisinger and Marius Pasca. Fine-Grained Class Label Markup of Search Queries. In *Proceedings of The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT2011)*, 2011.
- [56] Christian Sengstock and Michael Gertz. CONQUER: A System for Efficient Context-Aware Query Suggestions. In *Proceedings of the 20th International Conference on World Wide Web(WWW2011)*, pp. 265–268, 2011.
- [57] Minjoon Seo, Aniruddha Kembhavi, Ali Farhadi, and Hannaneh Hajishirzi. Bidirectional attention flow for machine comprehension. In *International Conference on Learning Representations (ICLR)*, 2017.
- [58] Iulian Vlad Serban, Alberto García-Durán, Çağlar Gülçehre, Sungjin Ahn, Sarath Chandar, Aaron C. Courville, and Yoshua Bengio. Generating factoid questions with recurrent neural networks: The 30m factoid question-answer corpus. In *Proceedings of the 54th Association for Computational Linguistics (ACL2016)*, pp. 588–598, 2016.
- [59] Nitish Srivastava, Geoffrey E. Hinton, Alex Krizhevsky, Ilya, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, Vol. 15, No. 1, pp. 1929–1958, 2014.
- [60] Rupesh Kumar Srivastava, Klaus Greff, and Jürgen Schmidhuber. Highway networks.

- CoRR*, Vol. abs/1505.00387, , 2015.
- [61] Markus Strohmaier, Mark Kroell, and Christian. Koerner. Intentional Query Suggestion: Making user goals more explicit during search. *Proceedings of the 2009 workshop on Web Search Click Data(WSCD'09)*, pp. 68–74, 2009.
- [62] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to sequence learning with neural networks. *Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS)* , pp. 3104–3112, 2014.
- [63] Robert S. Taylor. Question-negotiation and information seeking in libraries. *College and Research Libraries*, Vol. 29, No. 3, pp. 178–194, 1968.
- [64] Adam Trischler, Tong Wang, Xingdi Yuan, Justin Harris, Alessandro Sordoni, Philip Bachman, and Kaheer Suleman. NewsQA: A machine comprehension dataset. In *Workshop on Representation Learning for NLP (RepL4NLP@ACL)*, pp. 191–200, 2017.
- [65] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, undefinedukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS2017)*, p. 60006010, 2017.
- [66] Michail Vlachos, Christopher Meek, Zografoula Vagena, and Dimitrios Gunopoulos. 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.
- [67] Michael Völske, Ehsan Fatehifar, Benno Stein, and Matthias Hagen. Query-task mapping. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR2019)*, p. 969972, 2019.
- [68] Colin Wilkie and Leif Azzopardi. Query length, retrievability bias and performance. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management (CILM2015)*, pp. 1787–1790, 2015.
- [69] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. Xlnet: Generalized autoregressive pretraining for language understanding. In *Advances in Neural Information Processing Systems (NIPS2019)*, pp. 5753–5763. 2019.
- [70] Zhilin Yang, Peng Qi, Saizheng Zhang, Yoshua Bengio, William Cohen, Ruslan Salakhutdinov, and Christopher D. Manning. HotpotQA: A dataset for diverse, explainable multi-hop question answering. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP2018)*, pp. 2369–2380,

- 2018.
- [71] Zhijun Yin, Milad Shokouhi, and Nick Craswell. Query Expansion Using External Evidence. *31th European Conference on IR Research(ECIR2009)*, LNCS 5478/2009, pp. 362–374, 2009.
 - [72] Soungwoong Yoon, Adam Jatowt, and Katsumi Tanaka. Intent-Based Categorization of Search Results Using Questions from Web Q&A Corpus. In *Proceedings of the 10th International Conference on Web Information Systems Engineering(WISE2009)*, pp. 145–158, 2009.
 - [73] Adams Wei Yu, David Dohan, Minh-Thang Luong, Rui Zhao, Kai Chen, Mohammad Norouzi, and Quoc V. Le. QANet: Combining local convolution with global self-attention for reading comprehension. In *International Conference on Learning Representations (ICLR)*, 2018.
 - [74] Oleg Zendel, Anna Shtok, Fiana Raiber, Oren Kurland, and J. Shane Culpepper. Information needs, queries, and query performance prediction. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR2019)*, p. 395404, 2019.
 - [75] Zheng-Jun Zha, Linjun Yang, Tao Mei, Meng Wang, Zengfu Wang, Tat-Seng Chua, and Xian-Sheng Hua. Visual query suggestion: Towards capturing user intent in internet image search. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, 2010.
 - [76] Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter J. Liu. Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. *arXiv*, 2019.
 - [77] Tom Chao Zhou, Chin-Yew Lin, Irwin King, Michael R. Lyu, Young-In Song, and Yunbo Cao. Learning to suggest questions in online forums. In *Proceedings of the Twenty-Fifth Conference on Association for the Advancement of Artificial Intelligence (AAAI2011)*, pp. 1298–1303, 2011.
 - [78] 高田夏希, 大島裕明, 田中克己. Web と QA コンテンツの相互補完に基づくソーシャルサーチ. Web とデータベースに関するフォーラム 2010(WebDB Forum 2010), 2A-3, 2010.
 - [79] 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. 4, No. 2, pp. 74–87, 2011.
 - [80] 堀憲太郎, 大石哲也, 長谷川隆三, 藤田博, 越村三幸. Wikipedia からの拡張クエリ生成による Web 検索とその評価. 人工知能学会研究会資料, No. SIG-SWO-A803, pp. 13-1–13-7, 2008.
 - [81] 今井良太, 戸田浩之, 関口裕一郎, 望月崇由, 鈴木智也, 今井桂子. Web 検索サービスに

- おける多義的なクエリ推薦手法. 日本データベース学会論文誌, Vol. 9, No. 1, pp. 1–6, 2010.
- [82] 今村賢治, 東中竜一郎, 泉朋子. 対話解析のためのゼロ代名詞照応解析付き述語項構造解析. 自然言語処理, Vol. 22, No. 1, pp. 3–26, 2015.
- [83] 水野淳太, 村田祐一, 勝屋久. ユーザの嗜好を反映したクエリ拡張を用いた情報検索・推薦システムの開発. 楽天研究開発シンポジウム 2009, 2009.
- [84] 廣嶋伸章, 戸田浩之, 松浦由美子, 片岡良治. 概念ベースに基づく Web 検索のクエリタイプ判定手法とその評価. 情報処理学会論文誌 トランザクション データベース, Vol. 3, No. 3, pp. 33–45, 2010.
- [85] 大塚淳史, 西田京介, 斉藤いつみ, 浅野久子, 富田準二. 質問の意図を特定するニューラル質問生成モデル. 日本データベース学会和文論文誌, Vol. 17-J, No. 6, pp. 1–8, 2019.
- [86] 大塚淳史, 別所克人, 西田京介, 浅野久子, 富田準二. ニューラル翻訳モデルを用いた回答文書検索のための拡張クエリ生成手法. 日本データベース学会和文論文誌, Vol. 16-J, No. 4, pp. 1–8, 2019.
- [87] 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, Vol. 2, No. 3, pp. 130–140, 2009.
- [88] 村田真哉, 戸田浩之, 松浦由美子. 検索結果中のアクセス集中サイトを利用したクエリ拡張法の提案. *DBSJ Letters*, Vol. 6, No. 4, pp. 45–48, 2008.

業績一覧

学術雑誌論文

- 大塚淳史, 西田京介, 斉藤いつみ, 西田光甫, 浅野久子, 富田準二. 問い返し可能な質問
応答: 読解と質問生成の同時学習モデル. 日本データベース学会和文論文誌, vol.18-J,
no.16, pp.1-8. 2020.
- 大塚淳史, 西田京介, 斉藤いつみ, 浅野久子, 富田準二, 佐藤哲司. 質問意図の明確
化に着目した機械読解による質問応答手法の提案. 人工知能学会論文誌, vol.34, no.5
A-J14, pp.1-12, 2019.
- 大塚淳史, 西田京介, 斉藤いつみ, 浅野久子, 富田準二. 質問の意図を特定するニューラ
ル質問生成モデル. 日本データベース学会和文論文誌, vol.17-J, no.6, pp.1-8. 2019.
- 大塚淳史, 別所克人, 西田京介, 浅野久子, 松尾義博. ニューラル翻訳モデルを用いた回
答文書検索のための拡張クエリ生成手法. 日本データベース学会和文論文誌, vol.16-J,
no.4, pp.1-8. 2018.
- 大塚淳史, 関洋平, 神門典子, 佐藤哲司. 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. 日本データベース学会論文誌, Vol.11, No.1,
pp.1-6. 2012.
- 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, Vol.4, No.3,
pp.1-11. 2011.

国際会議論文

- Atsushi Otsuka, Kyosuke Nishida, Itsumi Saito, Hisako Asano, Junji Tomita. Spe-
cific Question Generation for Reading Comprehension. Proceedings of the AAAI
2019 Reasoning and Complex QA Workshop, pp.12-20. 2019.
- Kyosuke Nishida, Itsumi Saito, Kosuke Nishida, Kazutoshi Shinoda, Atsushi
Otsuka, Hisako Asano and Junji Tomita. Multi-style Generative Reading Com-

- prehension. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL2019), pp.2273-2284. 2019.
- Kosuke Nishida, Kyosuke Nishida, Masaaki Nagata, Itsumi Saito, Atushi Otsuka, Hisako Asano and Junji Tomita. Answering while Summarizing: Multi-task Learning for Multi-hop QA with Evidence Extraction. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL2019), pp.2273-2284. 2019.
 - Kyosuke Nishida, Itsumi Saito, Atsushi Otsuka, Hisako Asano, Junji Tomita. Retrieve-and-Read: Multi-task Learning of Information Retrieval and Reading Comprehension. Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM2018), pp.647-656. 2018.
 - Toshiki Tomihira, Atsushi Otsuka, Akihiro Yamashita, Tetsuji Satoh. What Does Your Tweet Emotion Mean?: Neural Emoji Prediction for Sentiment Analysis. Proceedings of the 20th International Conference on Information Integration and Web-based Applications & Services (iiWAS2018), pp.289-296. 2018.
 - Atsushi Otsuka, Kyosuke Nishida, Katsuji Bessho, Hisako Asano, Junji Tomita. Query Expansion with Neural Question-to-Answer Translation for FAQ-based Question Answering. Companion of the The Web Conference 2018 on The Web Conference 2018, pp.647-656. 2018.
 - Atsushi Otsuka, Toru Hirano, Chiaki Miyazaki, Ryuichiro Higashinaka, Toshiro Makino, Yoshihiro Matsuo. Utterance Selection Using Discourse Relation Filter for Chat-oriented Dialogue Systems. Dialogues with Social Robots Lecture Notes in Electrical Engineering, 427, pp.355-365. 2016.
 - Atsushi Otsuka, Toru Hirano, Chiaki Miyazaki, Ryo Masumura, Ryuichiro Higashinaka, Toshiro Makino, Yoshihiro Matsuo. Discourse Relation Recognition by Comparing Various Units of Sentence Expression with Recursive Neural Network. Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation (PACLIC 29), pp.6372. 2015.
 - Atsushi Otsuka, Yohei Seki, Noriko Kando, Tetsuji Satoh. QAque: Faceted Query Expansion techniques for Exploratory Search using Community QA Resources. Workshop in Community Question Answering on the Web(CQA2012) (Proceedings of WWW '12 Companion), pp.799-806. 2012.

国内会議論文

- 大塚淳史, 西田京介, 斉藤いつみ, 西田光甫, 浅野久子, 富田 準二. 問い返し可能な質問応答: 読解と質問生成の同時学習モデル. 第 11 回データ工学と情報マネジメントに関するフォーラム, A3-3, pp.1-8. 2019.
- 西田京介, 斉藤いつみ, 西田光甫, 篠田一聡, 大塚淳史, 浅野久子, 富田準二. 回答スタイルを制御可能な生成型機械読解. 言語処理学会年次大会発表論文集. 2019.
- 西田光甫, 西田京介, 永田昌明, 大塚淳史, 斉藤いつみ, 浅野久子, 富田準二. 抽出型要約との同時学習による回答根拠を提示可能な機械読解. 言語処理学会年次大会発表論文集. 2019.
- 光田航, 東中竜一郎, 大塚淳史, 片山太一, 齋藤邦子, 富田準二. 雑談対話におけるキーワードベース大規模対話ルールの構築. 言語処理学会年次大会発表論文集. 2019.
- 斉藤いつみ, 西田京介, 大塚淳史, 西田光甫, 浅野久子, 富田準二. クエリ・出力長を考慮可能な文書要約モデル. 言語処理学会年次大会発表論文集. 2019.
- 片山太一, 大塚淳史, 光田航, 齋藤邦子, 富田準二. 相手の発話を深掘りするための質問生成技術. 第 32 回人工知能学会全国大会論文集, 4G1-03. 2018.
- 西田京介, 斉藤いつみ, 大塚淳史, 浅野久子, 富田準二. 情報検索とのマルチタスク学習による大規模機械読解. 言語処理学会年次大会発表論文集, 2018.
- 大塚淳史, 別所克人, 西田京介, 浅野久子, 松尾義博. 答えを先読みする文書検索手法の提案. 第 9 回データ工学と情報マネジメントに関するフォーラム論文集, 4G-03. 2017.
- 大塚淳史, 片山太一, 杉山弘晃, 東中竜一郎, 松尾義博. 対話シナリオ構築におけるユーザ回答候補の推定. 第 31 回人工知能学会全国大会論文集, 1B3-2. 2017.
- 別所克人, 大塚淳史, 西田京介, 浅野久子, 松尾義博. クリックログの正解尤度の推定と検索への適用. 第 31 回人工知能学会全国大会論文集, 2O4-3. 2017.
- 大塚淳史, 別所克人, 平野徹, 東中竜一郎, 浅野久子, 松尾義博. 文構造を考慮した発話理解に基づく自然文検索. 第 30 回人工知能学会全国大会論文集, 4B4-2. 2016.
- 別所克人, 東中竜一郎, 大塚淳史. 雑談対話における話題継続願望判定の検討. 言語・音声理解と対話処理研究会, 74. 2015.
- 大塚淳史, 平野徹, 宮崎千明, 東中竜一郎, 牧野俊朗, 松尾義博. Recursive AutoEncoder を用いた文間の接続関係推定. 第 29 回人工知能学会全国大会論文集, 4K1-5. 2015.
- 大塚淳史, 関洋平, 佐藤哲司. 話題の周期性に着目した情報要求言語化のためのクエリ拡張手法の提案. 第 5 回データ工学と情報マネジメントに関するフォーラム, C9-1.

2013.

- 大塚淳史, 関洋平, 佐藤哲司. 時系列トピックモデルを用いたコミュニティ QA からの話題変動の抽出. 第 5 回 Web とデータベースに関するフォーラム, A3-3.2012.
- 大塚淳史, 関洋平, 神門典子, 佐藤哲司. コンテキスト切替による多様な情報要求に対する Web 検索手法の提案. 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, F8-4.2012.
- 大塚淳史, 関洋平, 神門典子, 佐藤哲司. 情報要求の言語化支援のためのコンテキスト提示型クエリ拡張法の提案と評価. 情報処理学会, マルチメディア, 分散, 協調とモバイルシンポジウム論文集, 1B-2. 2011.