

# アクション検索のためのブートストラップ的クエリ生成

北口 善紀<sup>1,a)</sup> 大島 裕明<sup>1,b)</sup> 田中 克己<sup>1,c)</sup>

**概要:** 本研究では、抽象的な行動を表す動詞が与えられたときに、より具体的・実地的・詳細な行動について Web で検索するために利用可能なクエリを生成する手法を提案する。たとえば「ゴルフ」というトピックと「上達する」という動詞が与えられたときに、(シャンク, 直す), (距離感, 掴む) という語の組を生成する。取得されるアクションは多様である事が望ましい。クエリの生成は、ブートストラップ法による動詞から名詞の検索, 名詞から動詞の検索の繰り返しにより行う。適切なクエリを取得するために、得られた語が検索に有用かどうかを確認する手法を採用する。また、検索にかかるコストを抑制するために、本研究ではブートストラップ法に検索結果多様性の技術を適用する。

**キーワード:** ブートストラップ法, アクション検索, クエリ生成, 検索結果多様性, semantic drift

## Query Formulation for Action Search by Bootstrapping

KITAGUCHI YOSHINORI<sup>1,a)</sup> OHSHIMA HIROAKI<sup>1,b)</sup> TANAKA KATSUMI<sup>1,c)</sup>

**Abstract:** In this paper, we propose a method to formulate queries to search for 'actions' from the Web. The goal of this paper is to search for several actions related to a given query. For example, if a user's initial query 'Golf improve' is input to know what concrete actions he should take for improving his golf game, then the proposing method can generate possible queries such as 'shanking stop' and 'distance adjust'. To formulate queries for the 'action' search, our algorithm repeatedly searches for nouns from verbs and verbs from nouns in a bootstrap manner. To obtain relevant query terms, we verify the usefulness of terms. To reduce search costs, we apply search-result diversification technique to the bootstrap term finding.

**Keywords:** bootstrap, action search, query formulation, search-result diversification, semantic drift

### 1. はじめに

自分がとるべき具体的な行動について知りたいという意図で検索を行うニーズがある。ゴルフを始めたばかりのユーザが上手になるためにどのような行動をとればよいか検索する場合について考える。このユーザがとるべき適切な行動として、シャンクを直すためにグリップを調整する、距離感を掴むために練習場で自分のクラブの飛距離を確かめる、といった行動が挙げられる。これらの行動について

調べるためには「シャンク 直す」、「距離感 掴む」などのクエリによる検索が必要となる。しかし、ゴルフの知識が乏しいユーザはそれらのクエリを自分で生成することが困難であるため、抽象的な「ゴルフ 上達」というクエリで検索することしかできない。この問題を解決するために、本研究では抽象的な行動を表す動詞を用いた検索により、具体的・実践的・詳細な行動について Web で検索する際に利用できるクエリの生成手法を提案する。

現状でアクションを Web で検索したい場合には、Google<sup>\*1</sup>, Yahoo<sup>\*2</sup>, Bing<sup>\*3</sup> などの検索エンジンのクエリボックスに適切なクエリを入力することとなる。ゴルフを

<sup>1</sup> 京都大学大学院情報学研究科社会情報学専攻  
Department of Social Informatics, Graduate School of Informatics, Kyoto University Yoshida Honmachi, Sakyo, Kyoto 606-8501, Japan

a) kitaguchi@dl.kuis.kyoto-u.ac.jp  
b) ohshima@dl.kuis.kyoto-u.ac.jp  
c) tanaka@dl.kuis.kyoto-u.ac.jp

\*1 <https://www.google.co.jp/>  
\*2 <https://www.yahoo.com/>  
\*3 <http://www.bing.com/>

上達させたいといった抽象的な意図を達成するのに有効な行動は数多くあるため、行動を取得するのに有用なクエリの数も多くなる。そのようなクエリを多く獲得しようと思うと、クエリを変えて何度も検索を行う必要がある。この際に以下の二種類の無駄な検索が行われうる。

(1) 適切な結果を得られない検索

(2) 新しい行動についての知識が得られない検索

一番目の問題の原因としてはユーザの知識が少ないことが挙げられる。ゴルフについての知識が乏しいユーザはどの語が検索に有効なのかを理解していないことにより、誤ったクエリでの検索を行ってしまうのである。二番目の問題は多くの適合した結果を得ようとする検索では避けられない。何度も検索を繰り返し、アクションの検索に適切なクエリがある程度得られた状態を想定する。この際に新しく有効なクエリを得ようと思い検索を行っても、今までに得られたクエリしか検索結果に含まれないということが多い。また、現時点で十分なクエリを取得できていて、これ以上検索を続けても新たなクエリを得る事ができないという状況であったとしても、ユーザがそれを知ることは難しい。本研究の提案する手法では有効な語のみを検索に利用し、同じ結果しか得られない検索をできる限り減らす工夫を取り入れることで無駄な検索の抑制を目指す。

本研究ではクエリの生成にブートストラップ法を用いる。ブートストラップ法では入力として与えた少数の語から抽出を開始し、得られた結果を次の抽出に利用することを繰り返すため、多くの語を抽出することができる。そのため、多くのクエリを生成したいという本研究の目標に合致していると言える。本研究ではトピックを表す語と抽象的な行動を表す動詞を入力として用いる。ゴルフを上達したいという意図での検索では「ゴルフ」がトピック語に、「上達する」が抽象的な行動を表す動詞に当たる。それらの入力を利用し、本研究では名詞と動詞の語の組をクエリとして生成する。クエリ生成のために、動詞をクエリとした検索結果から名詞を抽出、名詞をクエリとした検索結果から動詞を抽出という過程を新たな動詞が得られなくなるまで繰り返す。そうして得られた名詞の集合と動詞の集合から名詞と動詞の組を取り出し、クエリとして生成する。ブートストラップ法による検索においても、無駄な検索による適合率の低下、検索コストの上昇という問題は発生する。本研究の手法ではそれらの問題を解決する手法を取り入れる。

ブートストラップ法における適合率の問題は semantic drift と呼ばれる [1]。semantic drift とは出現頻度が高い表現のスコアが高くなるために元々の検索意図を満たす結果が得られなくなることを言う。ゴルフの上達のための検索だと「球を打つ」という表現の出現頻度が高く、「球」、「打つ」という語のスコアが高くなり結果の適合率が下がることがありうる。この問題を解決するために、本研究では得られた語が本当に検索に有効かどうかを抽出元の語を

得る事ができるか確認することで判断し、有効でない語を結果から取り除くという手段をとる。

検索コストに関する問題は検索結果多様性の概念を用いる事で解決する。通常の検索であれば新たに生成したクエリが今までには得られていない新しい結果を取得できるかどうかを事前に判断することは難しい。しかし、本研究でのブートストラップ法を用いた検索では抽出された語による検索結果に抽出元の語が含まれる可能性が高いため、クエリが新しい結果を得られるかどうかを事前に調べることができる。ゴルフというトピックにおいて、名詞「シャンク」を利用した検索結果から動詞「直す」を抽出した場合を考える。この際、「直す」をクエリとして得られた検索結果からも「シャンク」という抽出元の名詞が得られることが想定される。これをふまえて、本研究では得られた語の集合ができるだけ多様となるように類似度の高い語を結果から除外することによる検索コストの抑制を試みる。

## 2. 関連研究

本研究の関連研究としては、アクション検索、ブートストラップ抽出、パターンマッチングが挙げられる。

### 2.1 アクション検索

先行研究ではクエリログからアクションを発見する手法が提案されている。Lucchese ら [2] は 2 段階の手法でユーザが検索エンジン上で行おうとしたアクションを取得する方法を提案している。Lin ら [3] は、実世界のエンティティに対するアクションを取得する手法を提案している。

### 2.2 ブートストラップ

ブートストラップ法は語の抽出に用いられる。河合ら [4] は Web から同じクラスに属するキーワードを抽出するコストの低い検索手法を提案している。先行研究では semantic drift の抑制を対象としたものもある [5], [6]。このことから、semantic drift はブートストラップ法を用いる上で重要な問題であると言える。

### 2.3 パターンマッチング

パターンマッチングに関する手法は多くの研究者によって提案されている [7], [8], [9]。本研究の手法では名詞と動詞の関係を利用した手法を提案しているが、このように名詞や動詞を利用しているパターンマッチングに関する研究として Pekar ら [10], Torisawa ら [11] の研究がある。Torisawa らは名詞と動詞の関係を利用して同義となる動詞の組を取得する手法を提案しており、名詞と動詞の関係性からアクション取得のためのクエリを生成する本研究との関連が強いと言える。

### 3. アクション取得のためのクエリ生成

本研究ではトピック語と動詞を入力として、名詞と動詞の語の組で表されるアクション取得のためのクエリを生成する。クエリの生成のために、ブートストラップ法による名詞・動詞の抽出を行う。その際に、semantic drift を抑制するために、名詞のフィルタリングを行う。また検索コストを減らすために、多様性を用いたフィルタリングも行う。ブートストラップ法により名詞・動詞を抽出したのち、得られた名詞と動詞を対応付けることでクエリを生成する。以下で、名詞・動詞の抽出、semantic drift 抑制のための名詞フィルタリング、多様性を用いたフィルタリング、クエリ生成という四手法の詳細を説明する。

#### 3.1 名詞・動詞の抽出

クエリ生成のために必要な名詞や動詞を抽出する手法について説明する。ここでは、動詞集合  $V$  から名詞集合  $N$  を抽出する方法を説明する。名詞から動詞の抽出も同様である。動詞集合  $V$  中のすべての動詞  $v$  について、トピック語と動詞  $v$  をクエリとして検索 API により検索結果の上位  $k$  件を取得する。取得した検索結果からクエリとして与えた動詞  $v$  を、その動詞の目的語となる名詞  $n$  とともに抽出する。入力されたクエリが「ゴルフ 上達する」であれば、(アプローチ, 上達する), (パター, 上達する) といった語の組が抽出される。ここでストップワードに含まれる名詞は抽出対象から外すこととする。名詞集合  $N$  は動詞  $v$  との共起数が閾値を超える名詞  $n$  から構成される。動詞  $v$  と名詞  $n$  の共起数を  $Col(v, n)$  とすると、 $Col(v, n)$  が閾値を超えない名詞  $n$  は名詞集合  $N$  に含まれない。

ここで  $Col(v, n)$  は  $Col(n, v)$  と必ずしも同じ値とはならない。 $Col(v, n)$  が動詞  $v$  をクエリに用いて得られた検索結果内での名詞  $n$  と動詞  $v$  の共起数を表すのに対し、 $Col(n, v)$  は名詞  $n$  をクエリに用いて得られた検索結果内での名詞  $n$  と動詞  $v$  の共起数を表す。

$Col(v, n)$  が算出された後、名詞  $n$  のスコア  $Score(n)$  を以下の式で算出する。

$$Score(n) = \sum_{v \in V} link\_exists(Col(n, v)) \cdot Score(v) \quad (1)$$

$link\_exists(s)$  は  $s$  が閾値  $\theta_n$  を超えていれば 1 で、そうでない場合 0 となる。動詞のスコアを求める際には閾値に  $\theta_v$  が用いられる。名詞集合中の名詞は  $Score(n)$  によりランク付けされる。

上記のようにして動詞集合  $V$  から名詞集合  $N$  を取得する。動詞集合  $V$  の取得も同様である。ブートストラップ抽出は動詞集合  $V$  に新しい動詞が追加されなくなるまで続けられる。

ここまでで述べたのは単純なブートストラップ法によ

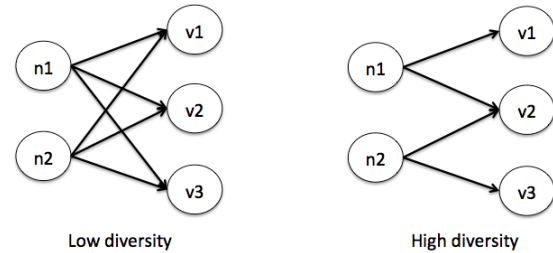


図 1 動詞集合  $V$  の多様性

る抽出手法である。何の工夫もしなければ、検索コスト、semantic drift に関する問題を解決できない。そこで多様性によるフィルタリング、semantic drift 抑制のための名詞フィルタリングによりこれらの問題を解決する。

#### 3.2 多様性によるフィルタリング

検索コストを減らすために動詞と名詞の共起関係から算出される多様性を利用する。ここでは名詞を利用した動詞の抽出を例として説明する。

図 1 の左側のグラフでは全ての動詞が同じ名詞  $n1, n2$  から抽出されている。これらの動詞は次の抽出に利用しても同じような名詞しか抽出できない可能性が高い。そこで次の抽出には得られた動詞の一部を利用すれば十分だと考えられる。それに対して、図 1 の右側のグラフでは動詞  $v1$  と共起する名詞が動詞  $v3$  とは異なる。このような場合、得られた動詞がそれぞれ異なる名詞を抽出するのに有用となる可能性が高い。そのため次の抽出に得られた動詞をすべて利用すべきだと考えられる。

以上をふまえて、抽出された動詞  $v$  が次の抽出に有用かどうかを  $MMR$  (*Maximal Marginal Relevance*) [12] アルゴリズムを使って判断する。動詞  $MMR$  は以下の式で選出される。

$$MMR = Arg \max_{v \in V \setminus S} [\lambda(Score(v)) - (1-\lambda) \max_{v' \in S} \frac{Sim(v, v')}{|V|}] \quad (2)$$

ここで  $S$  はすでに選択済みの動詞集合であり、 $v$  と  $v'$  の類似度  $Sim(v, v')$  は以下の式で求められる。

$$Sim(v, v') = \sum_{n \in N} link\_exists(Col(n, v)) \cdot link\_exists(Col(n, v')) \quad (3)$$

$MMR$  アルゴリズムにより動詞集合  $V$  から  $S$  に入る動詞を 1 つずつ選出するという作業を、 $MMR$  の値が 0 という閾値を下回るまで繰り返す。そうして最終的に得られた集合  $S$  を動詞集合  $V$  として取得する。 $MMR$  アルゴリズムを用いることにより、動詞集合中の動詞が同じ名詞を抽出するというケースが少なくなる。そのため、次の名詞抽出の際に coverage を保ちつつ検索コストを抑制できる。

#### 3.3 semantic drift 抑制のための名詞フィルタリング

単純なブートストラップ法ではトピックに頻出の表現による影響を受け、抽出される語が検索意図から離れた

ものになるという semantic drift を避けることが難しい。そこで本研究では抽出された名詞  $n$  が抽出元の動詞集合  $V_{pref}$  中の動詞を正しく抽出できているかを表すスコア  $Eligibility(n)$  を用いて semantic drift を抑制する手法を導入した。  $Eligibility(n)$  は以下の式で算出できる。

$$Eligibility(n) = \left( \prod_{v \in V_{elite}} \log(Col(n, v) + 1) + 1 \right)^{\frac{1}{|V_{elite}|}} \quad (4)$$

$V_{elite}$  は  $V_{pref}$  のサブセットであり、以下の式で求められる。

$$V_{elite} = \{v | v \in V_{prev} \text{ and } Score(v) > \frac{1}{|V_{prev}|}\} \quad (5)$$

$Eligibility(n)$  が閾値  $\theta_{sd}$  より小さかった場合には、名詞  $n$  は次の抽出の際のクエリとはならない。

### 3.4 クエリの生成

ブートストラップ法により抽出を繰り返し、名詞集合  $N$  と動詞集合  $V$  を得た後、名詞と動詞の対応付けを行うことでクエリを生成する。名詞集合  $N$  中の各名詞  $n$  について、  $Col(n, v) * Score(v)$  が最大となる動詞集合  $V$  中の動詞  $v$  を対応付ける。すべての名詞への対応付けが終わり次第、名詞と動詞の組の集合をクエリ集合として出力する。

## 4. 実験・評価

### 4.1 実験環境

実験では日本語の Web ページを Bing 検索 API<sup>\*4</sup> を利用して検索し、アクション取得のためのクエリを生成した。ストップワードリストの作成には、SlothLib のストップワードリスト<sup>\*5</sup> を利用した。

実験ではトピック語 12 個に以下に示す各目的を表す動詞を加えたクエリで検索を開始し、アクション取得のためのクエリを生成することとした。

- (1) カテゴリ 1: 悪い状態への遷移を防ぐ
- (2) カテゴリ 2: 何かを習得する

以前挙げた「ゴルフ 上達する」というクエリはカテゴリ 2 に含まれる。カテゴリ 1 に含まれるクエリとしては「パソコン 防ぐ」が挙げられる。この場合クエリ自体が抽象的な行動を表している訳ではないが、「防ぐ」という動詞により悪条件を回避するアクションを取得するためのクエリを生成できることが予備実験でわかっているため、このようなクエリも実験に利用することとする。

多様性によるフィルタリングは検索コスト、coverage への影響が大きいと考えられる。この手法は検索コストを抑制するために採用していて、新しい名詞・動詞を抽出できないクエリを排除する手法となっている。そのため、誤って有効なクエリを排除してしまっていた場合には、coverage

<sup>\*4</sup> <https://datamarket.azure.com/dataset/bing/search>

<sup>\*5</sup> <http://svn.sourceforge.jp/svnroot/slothlib/CSharp/Version1/SlothLib/NLP/Filter/StopWord/word/Japanese.txt>

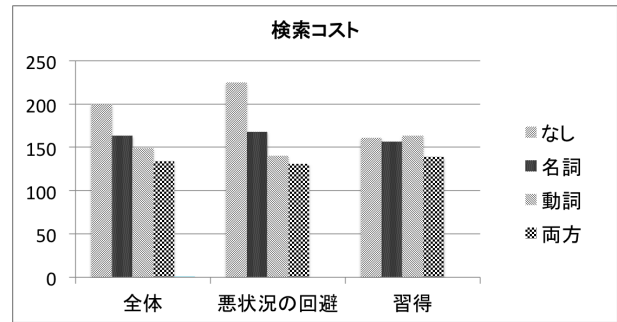


図 2 検索コスト

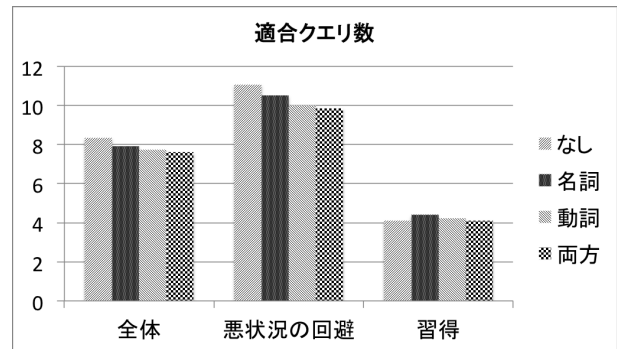


図 3 適合クエリ数

が下がることになる。coverage を保ちながら検索コストを抑制することが本研究での理想となる。

多様性フィルタリングの効果を調べるために、以下の四条件で実験を行った。

- (1) 手法 1: フィルタリングなし
- (2) 手法 2: 名詞をフィルタリング
- (3) 手法 3: 動詞をフィルタリング
- (4) 手法 4: 両方ともフィルタリング

すべての条件で、semantic drift を抑制する手法は同様のものを用いた。それぞれの手法で得られた結果から検索コスト、適合クエリ数、適合率、coverage を算出した。実験では次のようにパラメータを設定した。 $(k, \lambda, \theta_n, \theta_v, \theta_{sd}) = (100, 0.5, 2, 4, 1.3)$

### 4.2 結果

手法別の検索コストの算出結果は図 2 のようになった。検索コストは API の使用数で算出し、クエリごとの検索コストの平均値をグラフでは示している。これは他のグラフについても同様である。

手法 4(両方ともフィルタリング) がクエリ全体での検索における検索コストを減少させた。手法 4 では検索コストを手法 1(フィルタリングなし) の約 67% に抑えることに成功している。手法 3(動詞をフィルタリング) の検索コストは手法 2(名詞をフィルタリング) の検索コストを 14 下回った。このことから、今回の実験では検索コストを減らすのに動詞のフィルタリングが名詞のフィルタリングより有効だったと言える。

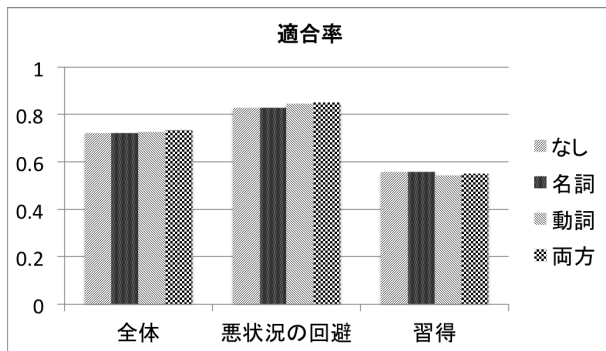


図 4 適合率

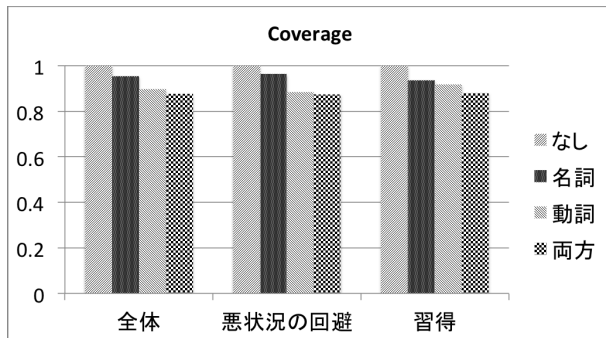


図 5 Coverage

カテゴリ 1(状況の悪化を防ぐ) に対して手法 4 がカテゴリ 2(何かを習得する) よりも検索コストを多く削減することができていた。カテゴリ 1 では手法 4 における検索コストは手法 1 の約 58% に抑えられている。それに対して、カテゴリ 2 では手法 4 における検索コストが手法 1 の約 86% であり、カテゴリ 1 と比較して検索コストの削減がうまくいっていなかった。

図 3 は手法ごとの獲得できた適合クエリ数を示している。手法ごとの差はほとんど見られなかった。しかし、動詞のカテゴリごとの比較では差が見られた。カテゴリ 1 については、すべての手法で約 10 個の適合クエリが獲得できていた。それに対して、カテゴリ 2 では、すべての手法について約 4 個しか適合クエリを獲得できていなかった。

図 4 は手法ごとの適合率を示している。適合率についても、手法ごとの差はほとんど見られなかった。しかし、アクション獲得数と同様に、カテゴリごとの比較では差が見られた。カテゴリ 1 については、すべての手法で適合率が 0.8 を超えていた。それに対して、カテゴリ 2 ではすべての手法の適合率が 0.6 を下回る結果となっていた。

図 5 は手法ごとの coverage を示している。coverage は各手法で獲得された適合クエリ数を手法 1 で獲得された適合クエリ数で割った値で表される。coverage については、カテゴリごとの差がほとんど見られなかった。すべてのカテゴリについて、手法 4 の coverage は 0.8 より大きい値を記録した。

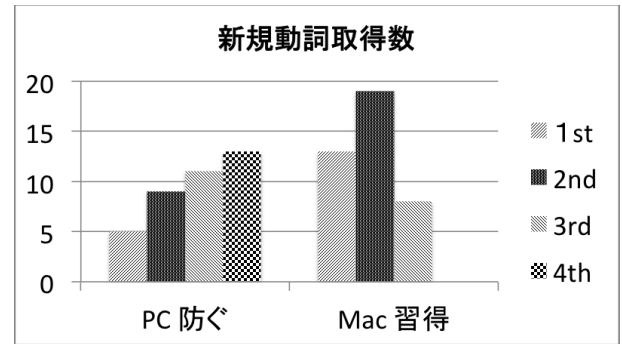


図 6 ステップごとの新規動詞取得数

### 4.3 評価

多様性によるフィルタリングにより検索コストを減らすことには成功したが、フィルタリング対象を増やすにつれて coverage が下がっていることから不適切なクエリだけを排除することができているとは言えない。また、semantic drift を防ぐ手法の効果が薄く、drift を防いでいないクエリがあった。そのようなクエリに関しては多様性によるフィルタリングを行うと逆に検索コストが多くなってしまいうという結果となっていた。

動詞をフィルタリングした際の検索コストが名詞をフィルタリングした際の検索コストを下回った。この要因としては semantic drift の抑制の際に名詞をフィルタリングしていたことが考えられる。semantic drift 抑制のための名詞フィルタリングが強く効いていたために、名詞のみのフィルタリングでは効果が薄かったものと思われる。

獲得できた適合クエリ数、適合率が多様性フィルタリングの手法を変えても変わらないことは予想通りであった。この手法は探索範囲を絞り込みながらも、同じ正解を得られることを期待した手法だからである。適合率、適合クエリ数を改善するためには、semantic drift 抑制手法の改善が必要となる。

coverage については概ね高い精度であったと言えるが、完全に同じ正解を取得できていないわけではないのでまだまだ改良の余地がある。

### 4.4 改善案

semantic drift が起こった理由を調べるために、クエリについてブートストラップ抽出の各段階における動詞の増加数を算出した。クエリごとの動詞の増加数は図 6 のようになった。「パソコン 防ぐ」というクエリについては、新規獲得動詞数は段階的に増加している。また初回で獲得された動詞数が 5 であり、あまり多くない。このケースでは初回に獲得された動詞の意味が広がりすぎなかったために、semantic drift が起きなかったと考えられる。それに対して、「Mac 覚える」というクエリでは最初のステップで大量の動詞が取得されてしまっている。これが原因で動詞の意味が発散してしまい、semantic drift が起こったものと推

測される。

semantic drift の抑制法として考えられるのは WordNet<sup>\*6</sup> など得られる同意語情報を利用して意味が離れすぎないようにフィルターをかける方法である。この方法だと、ある動詞が元の動詞の同意語でなければ排除されるため、意味の変化が起こりにくい。しかし、この方法では同意語でない語については抽出されないという問題がある。

適合率や正解獲得数を上げる方法としてクエリの生成法を変えろという方法が考えられる。他のクエリの生成法として、最終的に得られた動詞集合からもう一度抽出を行い、得られたすべての名詞をその抽出元の動詞と対応付け、得られた名詞と動詞の組をすべて取得するという方法が考えられる。この方法だと多様性によるフィルタリングで排除された名詞についても取得することができる。ここで動詞による名詞の抽出を提案しているのはアクションの目的が動詞に大きく依存していると考えたためである。また、正しいアクションを取得できる動詞から不適切なアクションを取得できるケースは少ないと予想されることも動詞による名詞の抽出を提案する理由の一つである。ただし、この方法だと取得された名詞の多様性については保証されず、同じようなクエリが多く生成されてしまう可能性がある。

## 5. まとめ

本研究ではブートストラップ法を用いて与えられたクエリに適合するアクションを取得するためのクエリを生成する手法を提案した。ブートストラップ法による抽出には二つの問題が考えられる。一つ目の問題は検索コストが大きいことである。本研究ではこの問題を、新しいアクションを得られないようなクエリによる検索を多様性を考慮したフィルタリングで減らすことにより解決した。もう一つの問題は semantic drift と呼ばれる適合性が低下する問題である。本研究ではこの問題を名詞が抽出元の動詞を正しく抽出することができるかどうか確認し、不適切な名詞を次の抽出に利用しないことで解決した。提案手法は検索コストをフィルタリングを行わない場合と比較して7割以下に減らす事ができた。しかし、適合率が0.6以下となっていたカテゴリが存在したため、semantic drift を抑制できていたとは言えない。そこで、今後は semantic drift 抑制手法を向上させることが重要であると考えている。

## 参考文献

[1] Curran, J. R., Murphy, T. and Scholz, B.: Minimising semantic drift with Mutual Exclusion Bootstrapping, *In Proc. of PAFLING* (2007).  
 [2] Lucchese, C., Orlando, S., Perego, R., Silvestri, F. and Tolomei, G.: Discovering Tasks from Search Engine Query Logs, *ACM Trans. Inf. Syst.*, Vol. 31, No. 3, pp. 14:1-14:43 (online), DOI: 10.1145/2493175.2493179

(2013).  
 [3] Lin, T., Pantel, P., Gamon, M., Kannan, A. and Fuxman, A.: Active Objects: Actions for Entity-centric Search, *Proceedings of the 21st International Conference on World Wide Web, WWW '12*, New York, NY, USA, ACM, pp. 589-598 (online), DOI: 10.1145/2187836.2187916 (2012).  
 [4] Kawai, H., Mizuguchi, H. and Tsuchida, M.: Cost-effective web search in bootstrapping for named entity recognition, *Database Systems for Advanced Applications*, Springer, pp. 393-407 (2008).  
 [5] McIntosh, T. and Curran, J. R.: Reducing Semantic Drift with Bagging and Distributional Similarity, *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1 - Volume 1*, ACL '09, Stroudsburg, PA, USA, Association for Computational Linguistics, pp. 396-404 (online), available from <http://dl.acm.org/citation.cfm?id=1687878.1687935> (2009).  
 [6] McIntosh, T.: Unsupervised Discovery of Negative Categories in Lexicon Bootstrapping, *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, EMNLP '10*, Stroudsburg, PA, USA, Association for Computational Linguistics, pp. 356-365 (online), available from <http://dl.acm.org/citation.cfm?id=1870658.1870693> (2010).  
 [7] Bhagat, R., Hovy, E. and Patwardhan, S.: Acquiring Paraphrases from Text Corpora, *Proceedings of the Fifth International Conference on Knowledge Capture, K-CAP '09*, New York, NY, USA, ACM, pp. 161-168 (online), DOI: 10.1145/1597735.1597764 (2009).  
 [8] Sang, E. T. K. and Hofmann, K.: Lexical Patterns or Dependency Patterns: Which is Better for Hypernym Extraction?, *Proceedings of the Thirteenth Conference on Computational Natural Language Learning, CoNLL '09*, Stroudsburg, PA, USA, Association for Computational Linguistics, pp. 174-182 (online), available from <http://dl.acm.org/citation.cfm?id=1596374.1596402> (2009).  
 [9] Chang, C.-H. and Lui, S.-C.: IEPAD: Information Extraction Based on Pattern Discovery, *Proceedings of the 10th International Conference on World Wide Web, WWW '01*, New York, NY, USA, ACM, pp. 681-688 (online), DOI: 10.1145/371920.372182 (2001).  
 [10] Pekar, V.: Acquisition of Verb Entailment from Text, *Proceedings of the Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, HLT-NAACL '06*, Stroudsburg, PA, USA, Association for Computational Linguistics, pp. 49-56 (online), DOI: 10.3115/1220835.1220842 (2006).  
 [11] Torisawa, K.: Acquiring Inference Rules with Temporal Constraints by Using Japanese Coordinated Sentences and Noun-verb Co-occurrences, *Proceedings of the Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, HLT-NAACL '06*, Stroudsburg, PA, USA, Association for Computational Linguistics, pp. 57-64 (online), DOI: 10.3115/1220835.1220843 (2006).  
 [12] Carbonell, J. and Goldstein, J.: The use of MMR, diversity-based reranking for reordering documents and producing summaries, *In SIGIR*, pp. 335-336 (1998).

\*6 <http://wordnet.princeton.edu/>