

実世界行動マイニングによる
時系列行動ネットワークの構築
～ 震災時の迅速な避難情報の共有に向けて～

NGUYEN MINH THE

電気通信大学大学院情報システム学研究科
博士（工学）の学位申請論文

2012年3月

実世界行動マイニングによる
時系列行動ネットワークの構築
～震災時の迅速な避難情報の共有に向けて～

博士論文審査委員会

主査 大須賀 昭彦 教授

委員 岡本 敏雄 教授

委員 田野 俊一 教授

委員 加藤 聰彦 教授

委員 川村 隆浩 客員准教授

著作權所有者

NGUYEN MINH THE

2012

Building Time-Series Action Network By Human Activity Mining

~ Toward a speedy sharing of information in disaster ~

NGUYEN MINH THE

Abstract

In this paper, we propose an approach that automatically extract users' activities in Japanese sentences retrieved from CGM (consumer generated media). We then design a time-series action network based on Web Ontology Language (OWL). By using the proposed activity extraction approach, we can automatically collect data for the action network. Finally, we propose a novel action-based collaborative filtering, which predicts missing activity data, in order to complement this action network. Moreover, with a combination of collaborative filtering and natural language processing, our method can deal with minority actions such as successful actions. Based on evaluation of 132,244 tweets which related to the massive Tohoku earthquake, we indicated that our approach can provide useful action patterns in real-time. Not only earthquake disaster, our research can also be applied to other disasters and business models, such as typhoon, travel, marketing, etc.

実世界行動マイニングによる時系列行動ネットワークの構築

～ 震災時の迅速な避難情報の共有に向けて～

NGUYEN MINH THE

概要

本論文は、震災時等においてユーザ状況に応じて必要な情報を提供するシステムの構築を目的として、① ソーシャルメディア（ブログ，Twitter）からユーザ行動を自動抽出する手法，② Web オントロジー言語（OWL）を用いた時系列行動ネットワークの設計・構築，および③ 欠損行動（ユーザによって投稿されなかった行動）を行動ベース協調フィルタリングにより推測し，行動ネットワークを補完する手法を提案している．特に，協調フィルタリングと言語処理を統合することで，マイノリティであるが価値のある行動に一定の重み付けを行う点に特徴がある．そして，東日本大震災発生時のつぶやき 132,244 件を対象に評価実験を行った結果，提案手法が避難情報を迅速に提供するのに有効であることが確かめられた．なお，本研究成果は震災時だけでなく，様々な自然災害（台風，大雪），人的災害（ストライキ，テロ）時の行動把握，更に観光案内，マーケティングなど多くの分野に適用可能であると考えられる．

目次

第1章 序論	1
1.1 背景と目的	1
1.2 本研究の貢献	4
1.3 本論文の構成	6
第2章 行動の自動抽出	7
2.1 機械学習と条件付確率場	7
2.1.1 機械学習とは	7
2.1.2 条件付確率場とは	9
2.2 行動の定義	10
2.3 行動抽出の問題点	11
2.4 Webからの行動の自動抽出	12
2.4.1 ベースライン手法(手法A)	12
2.4.2 自己教師あり学習の利用(手法B)	16
2.5 ブログ・Twitterからの行動の自動抽出	19
2.5.1 訓練データの追加	19
2.5.2 行動の時刻と行動の流れ	20
2.6 評価実験と考察	21
2.6.1 実験のデータセット	21
2.6.2 前処理	22
2.6.3 行動属性の抽出	24
2.6.4 実験結果の考察	25
2.7 関連研究	27

2.7.1	マイクロブログからの人間行動の抽出	27
2.7.2	Web からの人間行動の抽出	30
2.7.3	Web からのリレーション抽出	33
2.7.4	オープンリレーション抽出	36
第 3 章	時系列行動ネットワークの構築	41
3.1	時系列行動ネットワークの定義	41
3.2	時系列行動ネットワークの設計	42
3.3	時系列行動ネットワークの構築	45
3.4	評価実験	46
3.5	関連研究	49
第 4 章	時系列行動ネットワークの補完	51
4.1	補完の必要性	51
4.2	欠損行動の推測の課題	53
4.3	欠損行動の推測	54
4.3.1	ユーザ間の類似度に基づく推測	54
4.3.2	行動が起こる可能性に基づく推測	55
4.3.3	実行時間に基づく推測	55
4.3.4	欠損行動の推測式	56
4.4	マイノリティ情報の活用（協調フィルタリングと言語処理の統合）	58
4.5	評価実験	60
4.5.1	ユーザ間の類似度	60
4.5.2	欠損行動の推測	60
4.6	関連研究	62
第 5 章	震災時の応用	65
5.1	震災時の課題	65
5.2	避難・レスキューの応用の可能性	66
5.3	その他の応用の検討	72

5.4	震災時に実用可能なシステムのための分散処理の検討	73
第6章	結論	75
6.1	まとめ	75
6.1.1	行動の自動抽出手法のまとめ	75
6.1.2	時系列行動ネットワークの構築手法のまとめ	79
6.1.3	時系列行動ネットワークの補完手法のまとめ	81
6.1.4	震災および他の応用のまとめ	83
6.2	今後の課題	84
付録 A	時系列行動ネットワーク	93
付録 B	評価実験のデータおよび結果	103
	研究業績 (22 件)	105

目次

1.1	現在地から避難所までの行動を推薦する例	1
1.2	20:00 時“ 何をやったか ”をつぶやかなかった例	2
1.3	本論文のアプローチ	3
1.4	時系列行動ネットワークを参照して, 必需品の情報を取得する例	4
1.5	目的を達成するための行動の系列	5
1.6	MapReduce で大量のつぶやきを処理	6
2.1	機械学習	8
2.2	入力 x と出力 y	9
2.3	動作の頻度	11
2.4	動作と対象の共起頻度	12
2.5	行動抽出のベースライン手法	12
2.6	例文の単語と単語の品詞番号	13
2.7	例文の係り受け解析の結果	13
2.8	自己教師あり学習を用いて行動を抽出	16
2.9	名詞句と動詞句を単純化	17
2.10	訓練データのフォーマット	18
2.11	テンプレートファイル	18
2.12	手法 B を改善する手法 C	19
2.13	助詞を取り除き, 新しい訓練データを作成	20
2.14	新たな行動文を追加する流れ	20
2.15	タイムスタンプによる行動時刻の把握	20
2.16	ユーザ yusukef の行動の流れ	21

2.17	実験データの一部	22
2.18	「A集合」と「B集合」と「C集合」の関係	23
2.19	誤って捨てられた行動を表す文	23
2.20	誤って行動文として扱われてしまった文	24
2.21	お茶の作り方	30
2.22	オントロジーを用いた行動属性の抽出	31
2.23	大規模テキストからの経験マイニング	32
2.24	ブートストラッピング	33
2.25	組織と住所の組を含む文	34
2.26	リレーシヨンの正規表現パターン	34
2.27	ブートストラッピングの問題点	35
2.28	パターンと文をベクトル化	35
2.29	パターンと文のベクトルの類似度	36
2.30	SnowBall の評価方法	36
2.31	KnowItAll の手法	37
2.32	TextRunner のアーキテクチャ	39
3.1	時系列行動ネットワークの一部 (C, I, L はそれぞれクラス, インスタンス, ラベルを表す)	41
3.2	トリプルデータから成る時系列行動ネットワーク	42
3.3	継承する GeoNames, Time line, vCards	42
3.4	時系列行動ネットワークのクラス	43
3.5	時系列行動ネットワークのプロパティ	43
3.6	行動の成功回数	44
3.7	行動の頻度を表すプロパティ	44
3.8	時系列行動ネットワークのデータ (一部)	44
3.9	避難に関する時系列行動ネットワーク構築の流れ	45
3.10	2011-03-11T17:13:00 開始の行動を出力するためのクエリ	46
3.11	2011-03-11T17:13:00 開始の行動	46

3.12 次の行動を探すクエリ	47
3.13 次の行動	47
3.14 SPARQL で適切な避難所を探す例	48
3.15 開放している避難所の結果	48
3.16 ConceptNet の一部	49
3.17 時系列行動ネットワークと ConceptNet の比較	49
3.18 WordNet Search のインタフェース	50
4.1 20:00 時“ 何をやったか ”をつぶやかなかった例	51
4.2 時系列行動ネットワークの欠損の例	52
4.3 時刻 t に対象ユーザの行動	52
4.4 欠損行動を推測するアプローチ	53
4.5 つぶやかなかった時間に複数の行動が行われた場合	55
4.6 言語処理で成功行動を把握	58
4.7 震災時に人々の心理を反映する式	60
4.8 各アイテムに対してユーザの評価を表す行列	63
4.9 ユーザ間でのアイテム購入の先行関係	64
5.1 大震災が発生した場合の課題	65
5.2 時系列行動ネットワークを参照して、必需品の情報を取得する例	66
5.3 時系列行動ネットワークを参照して、運行状況を取得する例	67
5.4 目的を達成するための行動の系列	67
5.5 集合 A , 集合 B , 集合 C , 集合 D , 集合 E の関係	68
5.6 ユーザの状況に応じた避難行動を推薦	71
5.7 秋葉原に関する行動を表す行動ネットワーク (一部)	72
5.8 つぶやきに現れるユーザの購入行動	73
5.9 経験共有サービス	73
5.10 時系列行動ネットワークのデータ	74
5.11 MapReduce による大量のつぶやきの処理	74

6.1 2011年3月11日の大震災直後，新宿駅周辺での避難行動の一部（C，I，L
はそれぞれクラス，インスタンス，ラベルを表す） 102

表 目 次

2.1	Twitter から取得したユーザの行動	21
2.2	実験で利用するコンピュータの性能	22
2.3	機械学習モデル (CRF と SVM) の効果の比較	24
2.4	行動属性抽出の精度	25
2.5	Nilanjan らの手法 [40] との比較	25
2.6	Twitter のメッセージ	29
2.7	行動抽出の関連研究との比較	30
2.8	Web からの関係抽出の手法との比較	33
2.9	組織の住所のシード	34
2.10	Traditional RE と Open RE の比較	37
2.11	TextRunner と O-CRF の比較	38
2.12	O-CRF と本論文の手法の比較	40
4.1	『EC サイトのアイテム』と『震災時の行動』の比較	53
4.2	$t_{before} \sim t_{after}$ の間に, u_a が行った可能性がある行動	59
4.3	ユーザ間の類似度を測る際に考慮すべき要素	61
4.4	削除された行動の再現率	61
4.5	従来の協調フィルタリングとの比較	62
5.1	#hinan タグのつぶやきの質	69
5.2	時系列行動ネットワークを構成する行動の質	69
5.3	周辺行動の提示の処理時間	71

第1章 序論

本章では，本研究の背景と目的を述べた後，目的へのアプローチを説明する．最後に，本研究の主な貢献を説明し，本論文の構成について述べる．

1.1 背景と目的

計算機がユーザの状況（時刻，位置，行動）を把握し，それに応じて行動を推薦することは，コンテキストウェアコンピューティング [33] やユビキタスコンピューティング [43]，特に避難やレスキューといった応用分野において重要な課題とされている [51]．代表的な例として，図 1.1 に示すように，震災時における避難行動の推薦などが挙げられる．東日本大震災が起きた 2011 年 3 月 11 日，東京都内で外出中であった人の数は約 1 千万人と推定され，そのうち 300 万人の帰宅が困難となった [50]．また，今後 30 年以内に大震災が約 87% の確率で起こるとされており，その際の帰宅困難者は 650 万人にのぼると見込まれている [38]．

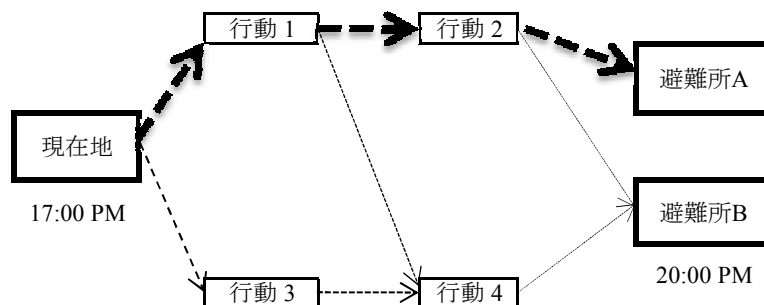


図 1.1: 現在地から避難所までの行動を推薦する例

計算機が人間に行動を推薦するためには，まず推薦すべき行動の獲得が必要である．2011 年 3 月 11 日大震災当日，電話回線が遮断された中，避難行動や避難所の状況など震災に

関して情報交換を行うため、Twitter が頻繁に利用された。日本国内における Twitter のつぶやき数は平時で約 1,800 万件 / 日であるのに対し、震災当日は約 3,300 万件 / 日に達し、1.8 倍に増加した [7]。したがって、震災時には人々の行動を Twitter のつぶやきの集合から獲得することが考えられる。しかし、つぶやきにはノイズ文字列が多く、他のテキストメディアより複雑かつ文法的に正しくない文が多い。また、緊急時であるため、迅速な処理が求められる。

また、行動と行動間の関係を事前に形式化しておくことも必要である。さらに、緊急時であることを考慮すると時間情報が重要となり、各行動の時刻を明確に記述し、時系列的に整理することが必要である。

更に、Twitter などのソーシャルメディアはユーザの自主性に依存しているため、つぶやかれなかった行動も数多く存在すると思われる。例えば、何らかの原因でユーザが 20:00 時に“どこで何をしたか”をつぶやかなかった例を図 1.2 に示す。リアルタイムにユーザの行動に応じた避難情報を推薦するためには、これらの欠損行動を推測する必要がある。欠損行動を推測するにあたっては、人間の行動は必ず前後の行動と依存関係があること、また例えそれがマイノリティであっても成功した行動はきちんと考慮すべきであることなど、いくつかの重要な課題がある。

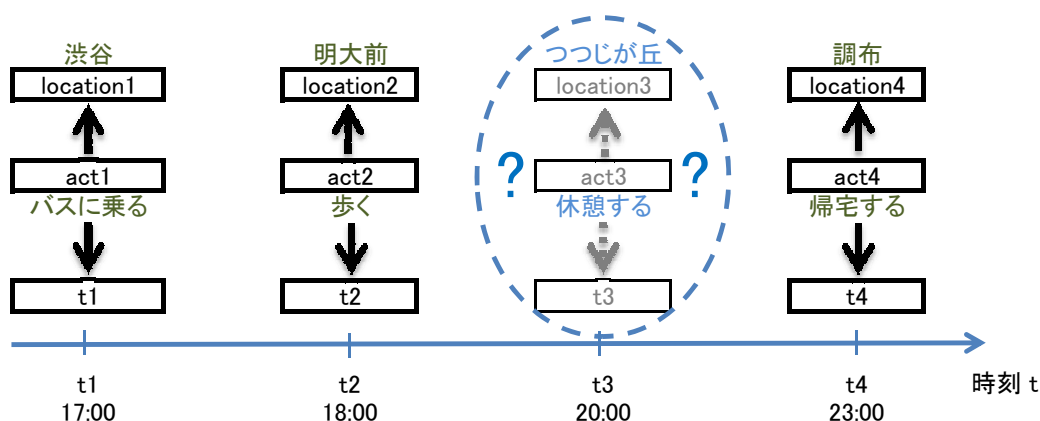


図 1.2: 20:00 時“何をやったか”をつぶやかなかった例

そこで本論文では、まず、ソーシャルメディア（ブログ、Twitter）から取得した日本語の文中に現れる行動を自動的に抽出する手法を提案する（図 1.3 の ①）。提案手法は条件

付確率場と自己教師あり学習を利用する。次に、時系列行動ネットワークを Web オントロジー言語 (OWL) に従って設計し、自動構築手法を提案する (図 1.3 の ②)。この行動ネットワークは各時刻に対する行動と行動間の関係を表す集合知となる。

最後に、行動の性質とユーザのゴール (避難, 帰宅など) を考慮した行動ベース協調フィルタリング手法を提案し、行動ネットワークを補完する (図 1.3 の ③)。つづやかれなかった時間に複数の行動があった場合でも、提案手法は行動の実行時間を考慮して補完することができる。また、協調フィルタリングと言語処理を統合することで、マイノリティでも価値がある行動には一定の重み付けを行う。

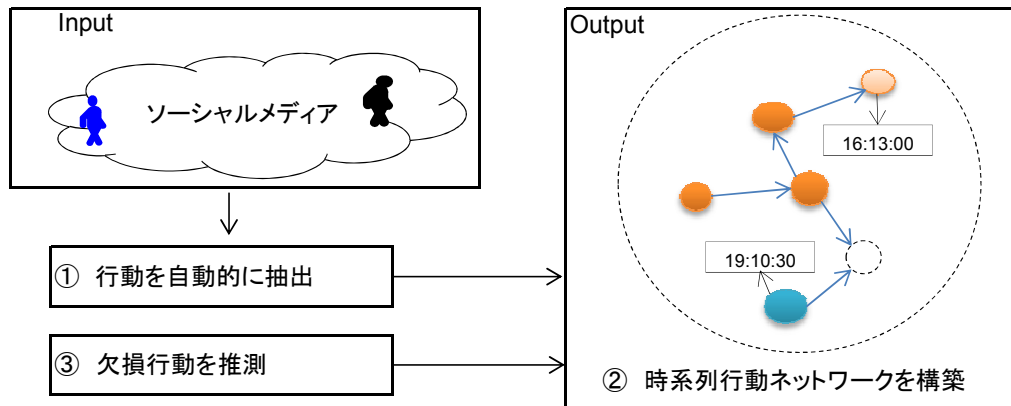


図 1.3: 本論文のアプローチ

1.2 本研究の貢献

本論文はユーザの状況に応じた行動を推薦するシステムの実現を目的として行ったものであり、以下の貢献が挙げられる。

- 日本語の文中に現れる行動属性（行動主、動作、対象、時間）を自動的に抽出する手法を提案した。提案手法では、ラベル編集、初期インスタンスの作成、行動のドメインの定義などを人手で行う必要がなく、準備コストがかからない。また、本手法は先行研究の課題を解決でき、Webの文書だけでなく、Twitterのようなソーシャルメディアの文書にも対応できる。そして、本手法は日本語を対象としたが、英語や他の言語にも適用可能である。更に、本手法は緊急時に大量な文書を迅速に処理可能である。
- 時系列行動ネットワークを提案し、OWLに従って設計した。そして、時系列行動ネットワークのデータを自動的に獲得する手法を提案した。
- 行動の性質とユーザのゴール（避難、帰宅など）を考慮した行動ベース協調フィルタリング手法を提案した。この手法を利用することで、時間毎にユーザの行動を推測でき、時系列行動ネットワークを補完することが可能である。そして協調フィルタリングと言語処理を統合することで、少数であるが価値がある行動（成功行動）にも適用可能にした。

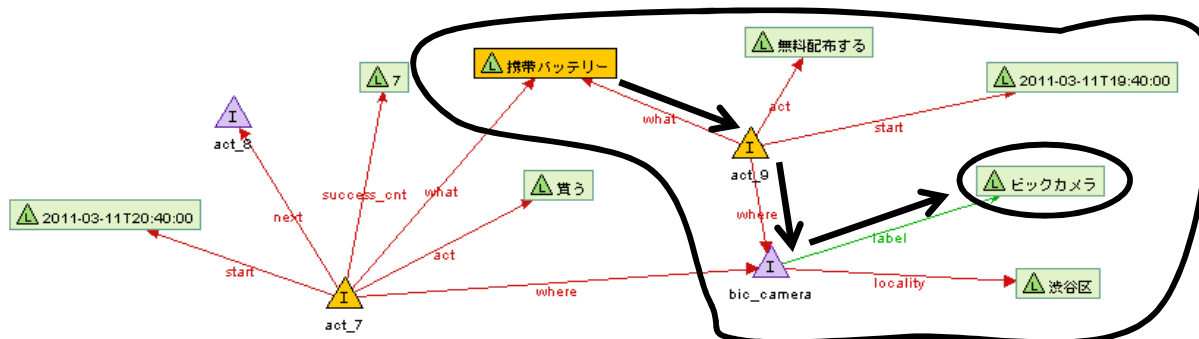


図 1.4: 時系列行動ネットワークを参照して、必需品の情報を取得する例

その上で、東日本大震災の避難に関するつぶやき 132,244 件を題材として評価実験を行った。その結果、時系列行動ネットワークが避難情報を迅速に提供するのに有効であることが確かめられた。具体的では、図 1.4 に示すように、携帯バッテリーのラベルから対象リンク (what) を辿って、行動のインスタンス (act_9) を介し場所 (where) が参照されると、ビックカメラが携帯バッテリーを無料で配布していることが得られる。また、同様に交通機関の運行状況とトイレ・避難所の情報も得られることを確認した。そして、時系列行動ネットワークでは、避難情報を場所ごとに時系列的に整理するので、場所と時刻に応じて把握可能である。更に、図 1.5 に示すように、時系列行動ネットワークを辿ることで、目的を達成するための行動の系列も把握可能である。

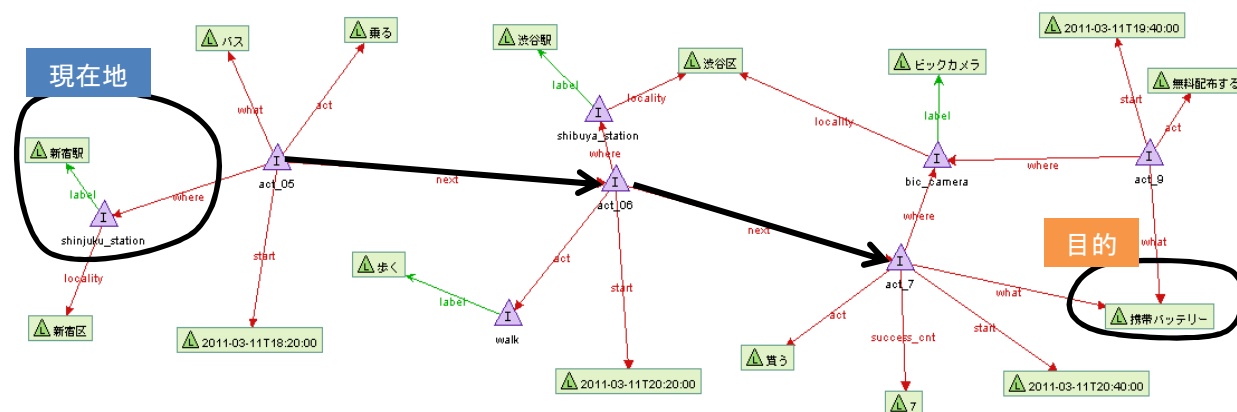


図 1.5: 目的を達成するための行動の系列

図 1.6 に示すように、Map-reduce[22] という分散処理アプローチを用いることで、大量なつぶやきを迅速に処理できるため、本研究の時系列行動ネットワークは震災時に実用可能であると考えている。

尚、本研究成果は震災時だけでなく、様々な自然災害（台風、大雪）、人的災害（ストライキ、テロ）時の行動把握、更に観光案内、マーケティングなど多くの分野に適用可能であると考えられる。

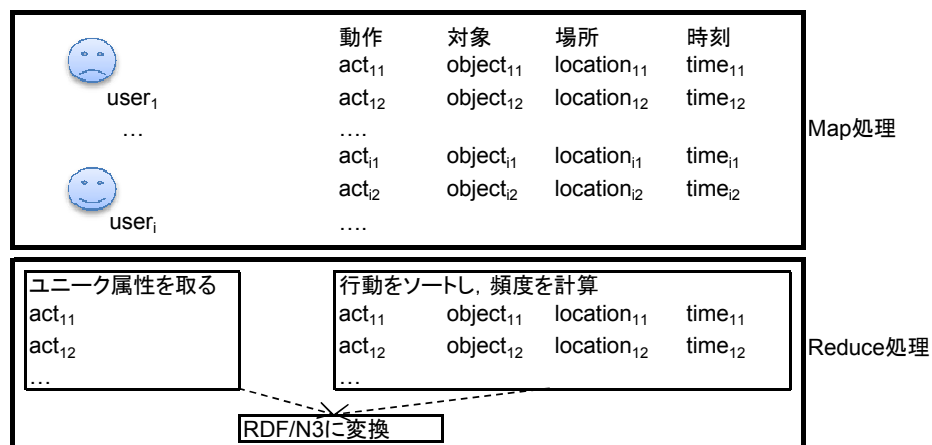


図 1.6: MapReduce で大量のつぶやきを処理

1.3 本論文の構成

本論文の構成は次の通りである．第2章では，まず背景知識として，機械学習と条件付確率場について説明する．次に，Web とソーシャルメディアから行動を自動抽出する手法を提案する．第3章では，時系列行動ネットワークの定義と構築方法を説明する．第4章では，行動ベース協調フィルタリングを提案し，欠損行動を推測する．第5章では，震災や観光案内など本研究の応用について説明する．最後に第6章で，本論文の各提案手法をまとめ，今後の課題を挙げる．

第2章 行動の自動抽出

本章では、まず 2.1 節で背景知識として、機械学習と学習モデルである条件付確率場について説明する。次に、2.2 節で行動の定義について説明する。2.3 節では、文中に現れる行動を抽出する問題点について説明する。2.4 節では、我々が提案する Web から行動の自動抽出手法を説明する。そして、2.5 節で、Weblogs から取得した文中に現れる行動を自動抽出するために、Web から行動の抽出手法を改善する。2.6 節では、評価実験を行い、考察を行う。最後に 2.7 節では、関連研究を説明し、本研究の提案手法との比較を行う。

2.1 機械学習と条件付確率場

2.1.1 機械学習とは

機械学習とは、人工知能における研究課題の一つで、人間が自然に行っている学習能力と同様の機能をコンピュータで実現させるための技術・手法のことである [36]。図 2.1 に示すように、学習モデルは訓練データ (training data) の特徴 (ラベル, ルールなど) を学習し (図 2.1 の ①) 訓練データの特徴モデルを作る (図 2.1 の ②)。そして、この特徴モデルに基づき (図 2.1 の ④)、学習モデルはテストデータ (test data) の特徴を予測する (図 2.1 の ③, ⑤)。

訓練データの作成方法により、機械学習は以下のように分類される。

1. 教師あり学習 (Supervised Learning)
2. 教師なし学習 (Unsupervised Learning)
3. 準教師あり学習 (Semi-Supervised Learning)
4. 自己教師あり学習 (Self-Supervised Learning)

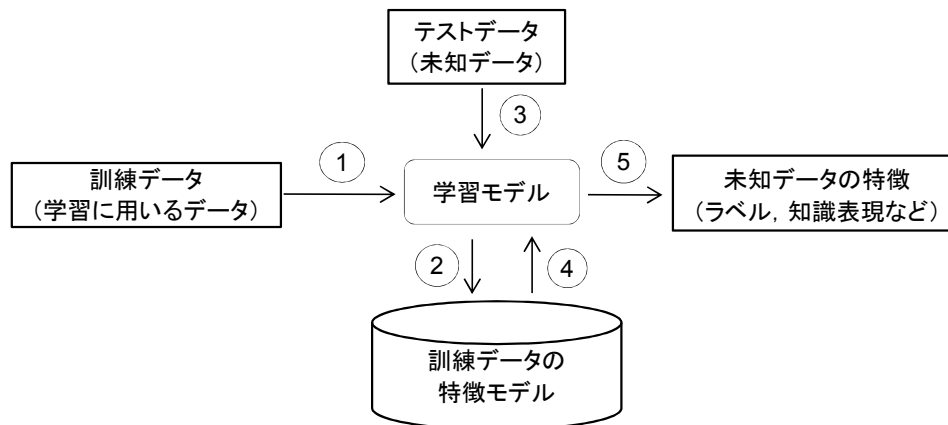


図 2.1: 機械学習

それぞれの手法について以下に解説する。

- 教師あり学習 (Supervised Learning)
教師あり学習では、訓練データを人手で作成する。
- 教師なし学習 (Unsupervised Learning)
教師なし学習では、訓練データがなく、入力データだけが与えられる。例えば、分類問題であれば、入力データ間の距離をその属性から計算し、類似するデータを1つのグループにまとめる。
- 準教師あり学習 (Semi-Supervised Learning)
教師あり学習で使う訓練データは人手で作るため、作成コストが大きい。一方、教師なし学習は、高い精度を得ることが難しい。そこで準教師あり学習では、少数の訓練データを人手で作っておく。そして学習を開始し、学習の過程で訓練データを拡大していく。
- 自己教師あり学習 (Self-Supervised Learning)
自己教師あり学習は教師なし学習の1種で、サンプルデータから訓練データを自動的に作成する。本研究ではこの手法を利用し、また学習モデルとして条件付確率場を使う。

2.1.2 条件付確率場とは

条件付確率場 (Conditional Random Fields) とは, John D. Lafferty ら [30] が提案した系列ラベリング問題に適用した識別モデルである.



図 2.2: 入力 x と出力 y

図 2.2 入力データを x (例えば, 文), 出力データを y (例えば, 固有名詞) とするとき, 学習モデルによって実現したいのは, x が与えられたときに対応する y が正しく出力されるということである. 条件付確率場 (以下 CRF) は一つの指数分布モデルで, この出力系列 $y = y_1, y_2, \dots, y_n$ の入力列 $x = x_1, x_2, \dots, x_n$ に対する条件付確率 $P(y|x)$ を表す.

$$P(y|x) = \frac{\exp(\alpha, \Phi(x, y))}{\sum_{y \in Y} \exp(\Phi(x, y))} \quad (2.1)$$

但し, $\Phi(x, y)$ は系列 $y = y_1, y_2, \dots, y_n$ 上のパスの全ての特徴ベクトルを足し合わせたものであり, α はモデルのパラメータである. そして CRF において, 新たな x が与えられたときの出力予測 \hat{y} は

$$\hat{y} = \operatorname{argmax}_{y \in Y} P(y|x) \quad (2.2)$$

となる. この出力は, Viterbi アルゴリズム [13] を用いることで効率良く解くことができる.

CRF は識別モデル (discriminative model) であり, 重複する特徴をモデルに組み込むことができる. 通常の識別モデルとの違いは, 出力が出力集合の部分集合ではなく, 系列となる点である. CRF は, 品詞付与 [30], テキストチャンキング [46], 固有表現抽出 [34], 形態素解析 [28] などといった系列ラベリング問題に適用され, いずれにおいても高い精度を示している.

2.2 行動の定義

行動の核となる要素は「行動主」、「動作」と「対象」である。そして、ユーザの状況に応じた適切な情報を提供するために、“どこで”、“どんな時に”、“いつ”行動が行われるかは重要である。このため、本論文では、人間の行動は「行動主」、「動作」、「対象」、「場所」、「時間」という5つの基本属性から成ると定義している。そして、これらの属性に以下のようなラベル (Who, Act, What, Where, When) を付ける。

1. 行動主：Who
2. 動作：Act
3. 対象：What
4. 場所：Where
5. 時間：When

「動作」は行動の中核であるため、この属性がない場合、行動として扱っていない。そして、行動を表す文では、「動作」の単独ではなく、「動作」と1つ以上の他の属性（「行動主」、「対象」、「場所」、「時間」）を含む必要がある。そのため、本論文で扱う行動文（行動を表す文）というのは、「動詞句と名詞句を持つ文」（例えば、秋葉原へ行く）又は「名詞句と名詞句が助詞“を”で結ばれた文」（例えば、英語を勉強）となる。例えば、“太郎は今晚渋谷に避難する予定”の文では、以下の行動属性を抽出する。

- 行動主 (Who): 太郎
- 動作 (Act): 避難する
- 場所 (Where): 渋谷
- 時間 (When): 今晚

2.3 行動抽出の問題点

震災等の緊急時には避難情報など行動に関する情報が特に重要であるため、本論文では行動属性の抽出に焦点をあてる。また、緊急時には、抽出処理の迅速さも重要であるだろう。そして、行動の抽出については、以下に示すような問題点がある。

1. 日本国内における Twitter のつぶやき数は平時で約 1,800 万件 / 日であるのに対し、震災当日は約 3,300 万件 / 日に達し、1.8 倍に増加した [7]。したがって、震災時に人手で抽出することは難しく、膨大なコストがかかる。
2. ブログや Twitter といった CGM (Consumer Generated Media) では、複雑かつ文法的に正しくない文が多い。そして CGM から取得する文では、顔文字や“ えーっと”、“。。。”などのようノイズ文字列が含まれる文が多い。
3. 行動属性を抽出する前に、我々は Twitter API を用いて、2010 年 5 月 06 日 ~ 2010 年 5 月 12 日までの「秋葉原」に関するつぶやきを取得した。調査した結果、図 2.3 に示すように、出現頻度が 1 ~ 2 回の動詞の割合は 81.13%であった。また「動作」と「対

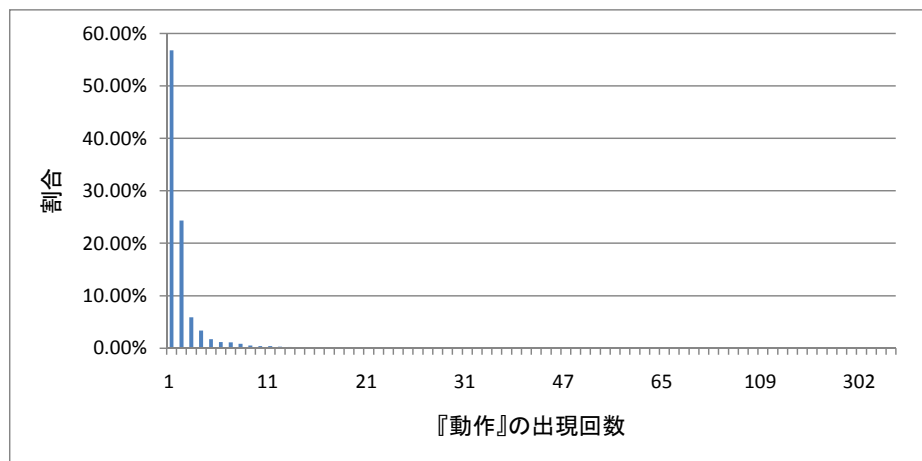


図 2.3: 動作の頻度

象」の共起頻度を図 2.4 に示す。これにより、頻度に基づくと、マイノリティ行動など多くの行動を獲得することはできないことが分かる。

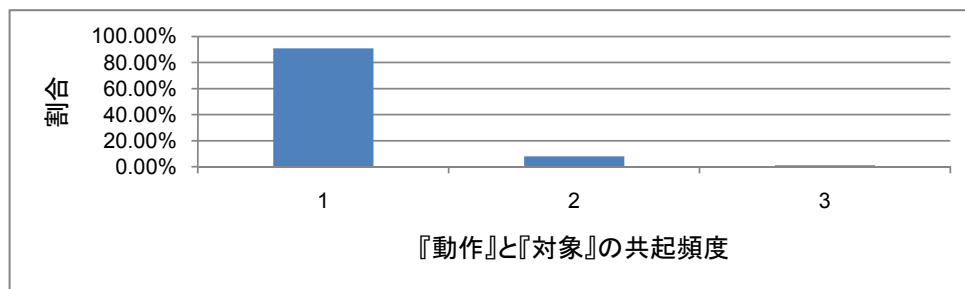


図 2.4: 動作と対象の共起頻度

4. 抽出手法が行動のドメインに依存すると、新たなドメインに適用するために、新たな訓練データの作成は必要になる。

2.4 Web からの行動の自動抽出

2.4.1 ベースライン手法（手法 A）

文中に現れる行動の抽出において、まず考えられる手法は図 2.5 に示すベースライン手法である。この手法では、既存の言語処理ツール（Mecab[28]，Cabocha[27]）と外部のリソース（Google Map API，Wikipedia の人名）を用いて、予め準備した構文パターンに従う文中に現れる行動属性（行動主，動作，対象，時間，場所）を抽出する。

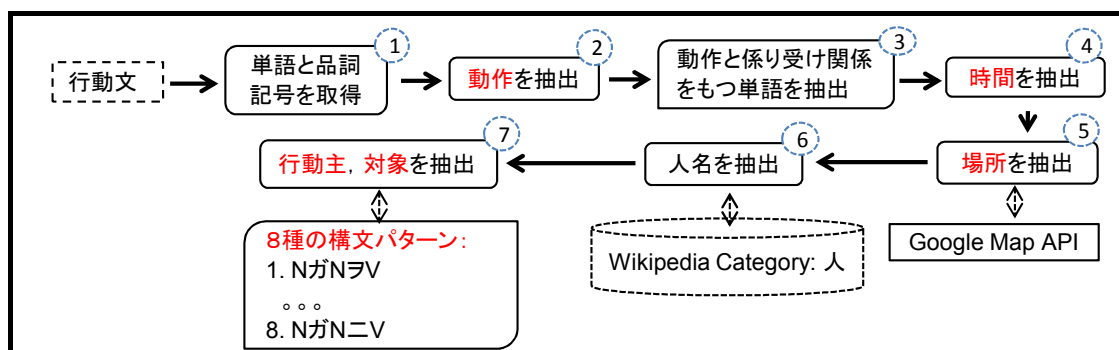


図 2.5: 行動抽出のベースライン手法

以下，“太郎は北海道で白い恋人を買った”という例文を用いて，抽出の流れを説明する。

1. 図 2.6 に示すように Mecab[28] で形態素解析を行い，文の単語と単語の品詞番号を取得する（図 2.5 の ①）。

単語	品詞記号
太郎	44
は	16
北海道	46
で	d
白い	10
恋人	38
を	w
買った	v
た	25

図 2.6: 例文の単語と単語の品詞番号

2. 動作（文の動詞句）を抽出する（図 2.5 の ②）。本例文では，動作は“買った”である。
3. Cabocha[27] で係り受け解析を行い，動作と係り受け関係をもつ単語を抽出する（図 2.5 の ③）。本例文では，図 2.7 に示すように，動作と係り受け関係をもつ単語は“太郎”，“北海道”，“白い恋人”である。

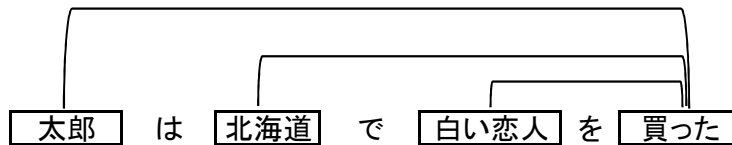


図 2.7: 例文の係り受け解析の結果

4. 動作と係り受け関係をもつ単語の内，時間を表す単語を抽出する（図 2.5 の ④）。
5. 動作と係り受け関係をもつ単語の内，場所を表す単語を抽出する（図 2.5 の ⑤）。抽出精度を向上するために，形態素解析結果に加えて，Google Maps API[17] を用いて場所を判定する。入力名詞句に対して，Google Maps API のレスポンスがアドレスであれば，この入力名詞句は場所であると判定する。

6. 長い人名をカタカナで書くと，Mecab の解析精度が落ちるため，Mecab の解析結果に加えて Wikipedia の人名カテゴリを活用して文の人名を検出する（図 2.5 の ⑥）．この方式以外にも，人名カテゴリ中の人名を Mecab の辞書に組み入れるという方式も考えられる．しかし，Mecab の辞書に組み入れる方式より，Wikipedia を直接に問い合わせる方式の方が柔軟性が高いと考える．
7. 日本語の構文リスト（NP ガ NP ヲ VP，NP ガ NP ニ VP など）に基づき，行動主と対象を抽出する（図 2.5 の ⑦）．本例文では，行動主，対象はそれぞれ“ 太郎 ”，“ 白い恋人 ”である．

次に，構文パターンと時間（時刻，場面）の判定について述べる．

● 構文パターン

我々は日本語の基本的な構文（例えば、『名詞』ハ『名詞』ヲ『動詞』）に基づき，以下の構文パターンを準備した．

1. “ ヲ ”，“ ニ ”又は“ ヘ ”という助詞がある文
 - {O, C}, {ヲ, ニ, ヘ}, V（例：映画を見る）¹
 - S, {O, C}, {ヲ, ニ, ヘ}, V（例：太郎は英語を学んでいる）
 - {O, C}, {ヲ, ニ, ヘ}, V, S（例：マイクロソフトを創業したのはゲイツです）
2. “ ガ ”と“ ハ ”を含む文
 - S ガ V ハ {O, C}（例：太郎が見た映画は面白い）
 - S ガ V {C} ハ {O}（例：ゲイツが創業した会社はマイクロソフトです）
 - S ハ N ガ V（例：太郎はラーメンが大好きなのでよく食べる）
3. その他
 - ヲ N（例：論文を作成）
 - N ガ（ハ）V（例：太郎が読む）

¹以下において，S は Subject（主語），O は Object（述語），C は Complement（補語），V は Verb（動詞），N は Noun（名詞）という意味を表す．

– NヲNニ(例：風景を写真に撮りました)

- 時刻(いつ)

下記の文字列を含む時間表現を時刻として抽出する。

“いま”、“今”、“現在”、“このところ”、“さいきん”、“最近”、“近ごろ”、“近頃”、“このあいだ”、“この間”、“さっき”、“もうすぐ”、“これから”、“しばらく”、“いつも”、“時々”、“たびたび”、“度々”、“たまに”、“年中”、“しばしば”、“しきりに”、“たえず”、“月曜日”、“火曜日”、“水曜日”、“木曜日”、“金曜日”、“土曜日”、“日曜日”、“春”、“夏”、“秋”、“冬”、“おととい”、“一昨日”、“きのう”、“昨日”、“きょう”、“今日”、“あした”、“明日”、“あさって”、“明後日”、“先々週”、“先週”、“今週”、“来週”、“再来週”、“先月”、“今月”、“来月”、“再来月”、“おとし”、“一昨年”、“ことし”、“今年”、“再来年”、“今朝”、“前々日”、“前日”、“当日”、“翌日”、“翌々日”、“午前”、“午前中”、“午後”、“秒”、“分”、“時”、“頃”、“朝”、“昼”、“夜”、“曜日”、“日”、“週”、“月”、“年”、“頃”、“午前”、“午後”、“毎日”、“毎週”、“毎月”、“毎年”、“毎朝”、“週末”、“月末”、“年始”、“年末”、“正午”、“上旬”、“中旬”、“下旬”

- 場面(どんな時に)

下記の日本語の文法パターンを採用して、場면을抽出する。

1. V-た後(に, は)²
2. V-たとたん
3. V-た瞬間
4. V-た直後
5. V - てから
6. V-前に
7. Nのまえ(に)

²VはVerb(動詞), NはNoun(名詞)という意味を表す。

8. V-まで(に)
9. V-るやいなや
10. V - (の)と同時に
11. V-する直前
12. V-している最中に
13. V-している以後,
14. ~時(に,は)
15. ~うち(に,は)
16. ~間(に,は)

2.4.2 自己教師あり学習の利用(手法B)

Web コーパスから取得した文は多様性を持つ．このため，予め準備した構文に従わない文も多い．このような文に対応するために，我々は図 2.8 に示す自己教師あり学習を利用する手法を提案する．

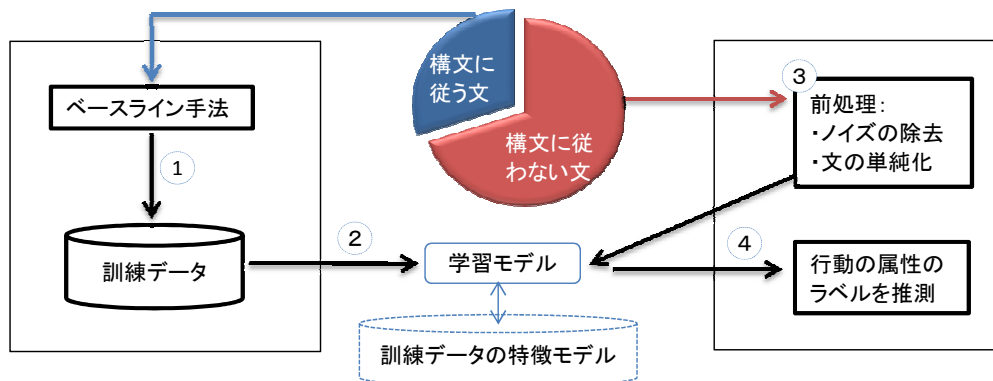


図 2.8: 自己教師あり学習を用いて行動を抽出

提案手法の手順は以下の通りである．

1. 予め準備した構文パターンに従う文を対象に，ベースライン手法を用いて訓練データを作成する(図 2.8 の 1)．

2. 学習モデル (SVM[21], 条件付確率場 [30] など) を用いて, 訓練データの特徴を学習し, 特徴モデルを作成する (図 2.8 の 2) .
3. 予め準備した構文に従わない文の前処理 (ノイズの除去, 文の単純化) を行う (図 2.8 の 3) . 本論文では, 図 2.9 に示すように, 名詞句と動詞句を単純化することで複雑な文を単純化する .

留学	36		
先	51		
の	24		
国	38		
,	9		
地域	38		
,	9		
時期	67	→	
,	9		
留学	36		
の	24		
種類	38		
を	w		
選び	v		
始める	v		
		NP1	38
		を	w
		VP1	v

図 2.9: 名詞句と動詞句を単純化

4. 訓練データの特徴モデルを用いて, 前処理した文中に現れる行動属性のラベルを推測する (図 2.8 の 4) .

本論文では, データ (訓練データ, テストデータ) フォーマットとテンプレートファイルの設計方針は次の 2 つである . 1 つ目は, 訓練データの特徴を全て吸収できることである . 2 つ目は, 学習とテストの計算量を最小限にすることである . これらの設計方針に従い, 本手法は以下のようなデータフォーマットとテンプレートファイルを用いる .

- 訓練データのフォーマット

図 2.10 に示すように, 訓練データは単語列, 品詞記号列, ラベル列から構成される . 単語列は文の単語を表す . 品詞記号列は Mecab の品詞番号, 助詞の記号 (に n, を

単語	品詞記号	ラベル
太郎	44	B-Who
は	16	O
北海道	46	B-Where
で	d	O
白い	10	B-What
恋人	38	I-What
を	w	O
買った	v	B-Act
た	25	I-Act

図 2.10: 訓練データのフォーマット

w, h, d, 動詞の記号 (v) を表す。ラベル列は行動属性のラベルを表す。但し, Who は行動主のラベル, Where は場所のラベル, What は対象のラベル, Act は動作のラベル, B は Begin (属性の先頭), I は Inside (属性の途中), O は Other (属性以外) という意味を表す。

- テストデータのフォーマット

テストデータは単語列と品詞記号列だけがある。ラベルはテストする際にシステムが推測して自動的に付与する。

- テンプレートファイル

# POS column	# Word column
U10:%x[-3,1]	U00:%x[-3,0]
U11:%x[-2,1]	U01:%x[-2,0]
U12:%x[-1,1]	U02:%x[-1,0]
U13:%x[0,1]	U03:%x[0,0]
U14:%x[1,1]	U04:%x[1,0]
U15:%x[2,1]	U05:%x[2,0]
U16:%x[3,1]	U06:%x[3,0]
U17:%x[-3,1]/%x[-2,1]	U07:%x[-1,0]/%x[0,0]
U18:%x[-2,1]/%x[-1,1]	U08:%x[0,0]/%x[1,0]
U19:%x[-1,0]/%x[0,1]	
U110:%x[0,1]/%x[1,1]	#POS column 's junction
U112:%x[1,1]/%x[2,1]	U21:%x[-3,1]/%x[-2,1]/%x[-1,1]
U113:%x[2,1]/%x[3,1]	U22:%x[-2,1]/%x[-1,1]/%x[0,1]
	U23:%x[-1,1]/%x[0,1]/%x[1,1]
U25:%x[1,1]/%x[2,1]/%x[3,1]	U24:%x[0,1]/%x[1,1]/%x[2,1]

図 2.11: テンプレートファイル

図 2.11 はテンプレートファイルを示す。テンプレートファイルとはデータの特徴を表す

ためのファイルである．提案手法が利用する素性（特徴）は単語，品詞，助詞である．テンプレートファイルはこれらの素性と係り受け関係を扱う．そして，長い文に対応させるために，サイズ7のウィンドウ ($\%x[-3,*] \sim \%x[3,*]$) を採用する．図 2.11 にテンプレートファイルの全体を示す．但し， $\%x[i,j]$ は現在の位置からの相対位置で i 行目の j 番目の列の要素を指す．また， U^{**} はテンプレートの記号である．

2.5 ブログ・Twitter からの行動の自動抽出

通常の Web 文書と比べ，ブログ，Twitter 等から取得した文書は複雑かつ文法的に正しくない行動文が多い．これらの文に対応するために，図 2.12 に示すように，我々は訓練データの追加方法を提案し，手法 B を改善する（手法 C）．

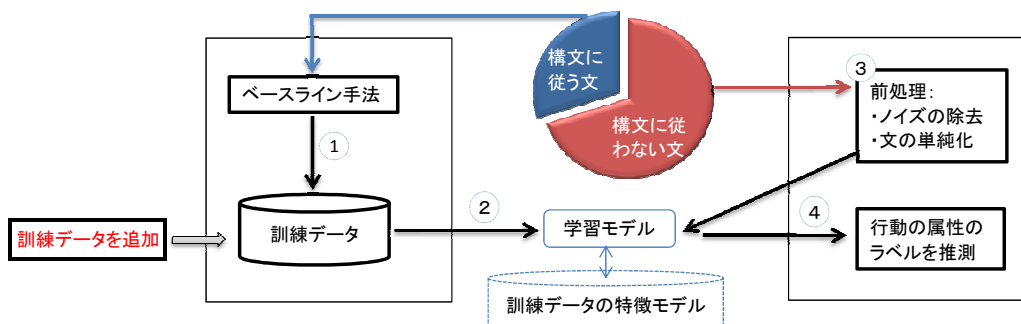


図 2.12: 手法 B を改善する手法 C

本節では，まず訓練データの追加方法を説明する．そして，Twitter のつぶやきのタイムスタンプを利用して，行動の時刻と行動の流れを把握する．

2.5.1 訓練データの追加

ブログ・Twitter から取得した多くの行動文では，助詞（ヲ，ニ，ヘ）が省略されている．これらの行動文に対応するために，本手法では，図 2.13 に示すように元の訓練データの助詞の行を取り除き，新たな訓練データを作成する．

そして，図 2.14 はもう一つの新たな訓練データの追加方法を示す．この方法では，まず，単純な行動文中に現れ行動属性（動作と対象）を抽出する．次に，抽出された行動属性を

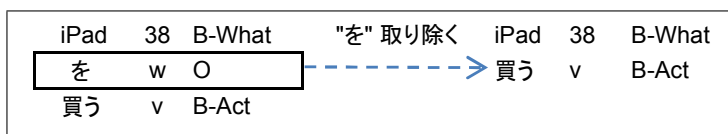


図 2.13: 助詞を取り除き，新しい訓練データを作成

用いて，検索キーワードを作成する．これらのキーワードをブログ検索エンジンに入力して，新たな行動文を取得する．最後に，取得した新たな行動文の訓練データを作成する．

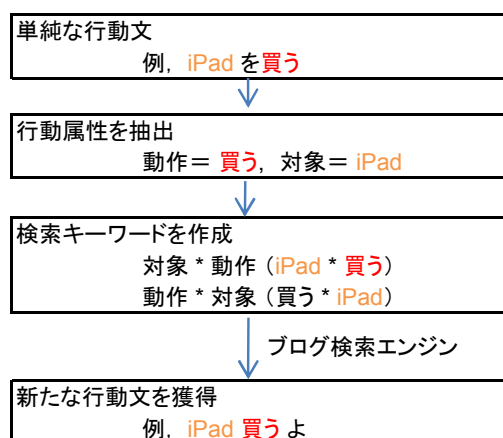


図 2.14: 新たな行動文を追加する流れ

2.5.2 行動の時刻と行動の流れ

図 2.15 に示すように，提案手法では，つぶやきの文書中に現れる時間表現（時，なう，今晚，明日の朝など）とつぶやきの発生時刻を考慮して，行動の時間を把握する．



図 2.15: タイムスタンプによる行動時刻の把握

そして，行動の時刻に基づき，ユーザの行動の流れを判定する．例えば，表 2.1 は yusukef というユーザの行動を表す．これらの行動の時間に基づき，yusukef の行動の流れは図 2.16

の通りである．

表 2.1: Twitter から取得したユーザの行動

ユーザ	動作	対象	場所	時間
yusukef	歩く		豊島区西池袋	2011/3/11 18:01:00
yusukef	買う	自転車	豊島区池袋本町	2011/3/11 18:37:00
yusukef	運転する	自転車	豊島区池袋本町	2011/3/11 18:45:00
yusukef	食べる		埼玉県戸田市下戸田	2011/3/11 19:17:00
yusukef	帰宅する		埼玉県戸田市下戸田	2011/3/11 19:41:00

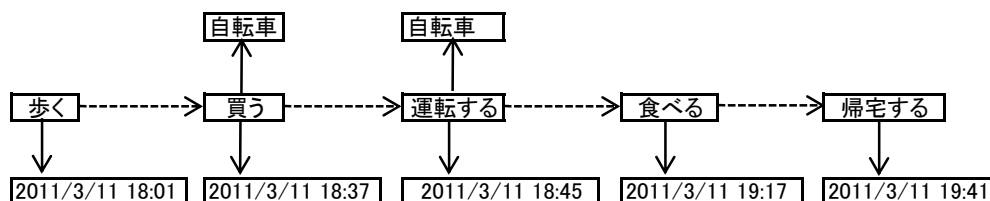


図 2.16: ユーザ yusukef の行動の流れ

2.6 評価実験と考察

本節では、日本語の行動文中に現れる行動属性（行動主，動作，対象，時間，場所）の抽出を実験する．まず，行動文を取得するための前処理を評価する．次に，提案手法を利用した機械学習モデルである条件付確率場（CRF）の有効性を評価する．そして，前節で説明した手法 A，手法 B，手法 C の比較を行う．最後に，実験結果に基づいて考察を行う．

2.6.1 実験のデータセット

我々は Google Blog Search API と Twitter Search API を用いて，2010 年 5 月 06 日～2010 年 5 月 12 日までにブログ・Twitter 上に投稿された日本語の文書を取得した．次に，これらの文書の前処理を行い，行動文を獲得した．そして，行動文の集合から行動文をランダムに選び，3 つの実験データセットを作成し，行動属性を抽出した．それぞれの実験データセットは 1000 件の行動文からなる．図 2.17 は実験データの一部を示す．

日本へ帰国したついでに、秋葉原で一眼レフLUMIX GP1を購入。
 ヨドバシ秋葉原でSDHC4GB買った。
 秋葉原クラブセガ新館5Fでゴーストスカッドやるの。
 PCは秋葉原で買うことにした。
 生協が推奨するPCは無駄に高性能なんで。
 今日は説明会のため秋葉原へ行った。
 途中で婦警コスプレで猫耳着けたお姉さんを見た。
 でも帽子も持ってた。
 これから成田に移動します。
 また感想をアップしますね。
 バスを待っている間、回りの景色を眺めていた。

図 2.17: 実験データの一部

また、実験で利用したコンピュータの性能は表 2.2 のようなものである。

表 2.2: 実験で利用するコンピュータの性能

OS	CPU	RAM
Windows 7 Ultimate	Intel(R) Xeon(R) 2.67 GHz	12.0 GB

2.6.2 前処理

本論文ではブログや Twitter 等から取得した文書そのものではなく、前処理を行って、節 2.2 に定義した「行動を表す文」を対象に行動属性を抽出する。Mecab の解析精度は 100% ではないので、前処理の過程で「行動を表す文」が削除されてしまう可能性がある。また、行動を表さない文が行動を表す文として扱われてしまう可能性もある。一般の文では、Mecab での品詞判定精度は約 98% [28] とされているため、前処理の課程で誤って捨てられた又は含められた行動を表す文は約 $\pm 2\%$ 程度だと考えられる。これを確かめるために、我々は Mecab の精度と本論文での「行動を表す文」の定義を含めて、実際の誤差を把握するため、以下の評価実験を行った。

図 2.18 では、以下のような記号を用いて、前処理について説明する。

A は、ブログや Twitter 等から取得した文の集合である。

B は、A の中にある行動を表す文の集合である（行動を表す文の正解集合）。

C は、A を対象に前処理を行った結果として取得され、行動属性を抽出する対象文の集合である。

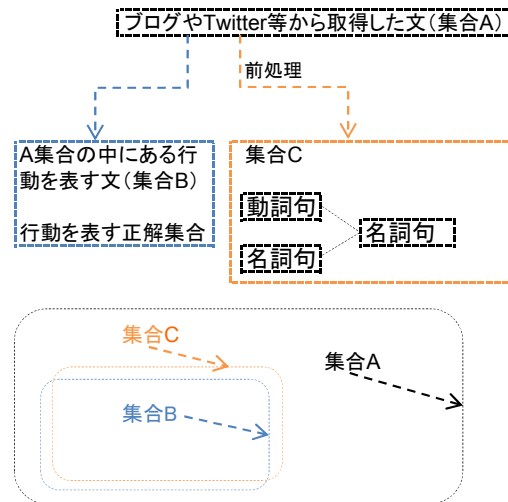


図 2.18: 「A 集合」と「B 集合」と「C 集合」の関係

人手による確認の結果、前処理過程で捨てられた文は 31.43% で、このうち誤って捨てられた行動を表す文は 6.67% であった。行動文によって属性の種類と数は変わるが、約 6% の行動属性を誤って捨ててしまっていることが分かった。尚、前処理過程で、誤って捨てられた行動を表す文は図 2.19 の通りである。これらの文では、動詞が“ お客様訪問終了 ”や“ 妹の電子辞書購入 ”などの名詞となっている。また、人は“ 群馬県太田市へ～。”を行くという行動だと判断できるが、“ 行く ”という動詞が省略されているため、計算機は判断できなかった。

秋葉原にてお客様訪問終了。
 今日はこれから秋葉原で待ち合わせ。
 秋葉原で妹の電子辞書購入。
 秋葉原ノードでリハ。
 これから群馬県太田市へ～。

図 2.19: 誤って捨てられた行動を表す文

前処理過程で、行動を表さないが行動を表すとして扱われてしまった文は図 2.20 の通りである。これらの原因は、Mecab が品詞を誤って判断したためである。“ 田中すだれ店。”の文では、Mecab が“ だれ ”は動詞 (“ だれる ”) であると判断してしまった。“ でも喫茶じゃないでふ。”の文では、Mecab が“ ふ ”は動詞 (“ ふる ”) であると判断してしまった。これらの前処理の誤りについては、今後、改善を検討していく。

田中すだれ店。
でも喫茶じゃないでふ。

図 2.20: 誤って行動文として扱われてしまった文

2.6.3 行動属性の抽出

我々はよく利用されている識別モデル SVM (Support Vector Machine) を対象に比較実験を行った。実験結果では、表 2.3 に示すように、いずれの行動属性においても SVM より CRF の方が高い精度を示している。

表 2.3: 機械学習モデル (CRF と SVM) の効果の比較

	学習モデル	行動	行動主	動作	対象	時間	場所
適合率	Multi-SVM	66.15%	77.22%	90.02%	74.05%	73.51%	75.20%
	CRF	73.21%	82.25%	97.11%	81.23%	80.04%	82.11%
再現率	Multi-SVM	60.03%	72.03%	85.31%	70.02%	71.78%	72.15%
	CRF	66.54%	80.11%	93.18%	76.57%	79.75%	81.02%
F-measure	Multi-SVM	62.94%	74.53%	87.60%	71.98%	72.63%	73.64%
	CRF	69.72%	81.17%	95.10%	78.83%	79.89%	81.56%

表 2.4 は手法 A, 手法 B, 手法 C の比較実験の結果を示す。ただし、手法 A はベースライン手法である。手法 B とは 2.4 節で説明した我々の Web の文書に対応する手法である。手法 C とは 2.5 節で説明した手法 B を改善する手法である。これらの実験結果に基づくと、ベースライン手法では、適合率が高いが、再現率が低い。そして、適合率と再現率の両方において、手法 B より手法 C の方が精度が高いことが確認できた。

Nilanjan らの手法 [40] は、人間の行動に着目して、Twitter ユーザの興味をリアルタイムで把握する手法を提案している。この手法では、まずユーザの興味に関する「カテゴリワード」(例えば、movie, cinema, music, sports など)、「動詞」(例えば、watch, watching, party など)、「時間ワード」(例えば、tonight, today, weekend など)を作っておく。次に、これらのワードに基づいて、Twitter メッセージに対して (カテゴリワード, 動詞) と (カテゴリワード, 時間ワード) の共起頻度を計算する。共起頻度が高いものはユーザの興味があるものと判断する。例えば、“movie”(カテゴリワード)と“go”(動詞)と“tomorrow”(時間ワード)の共起頻度が高ければ、ユーザの興味は“going to a movie tomorrow”(明

表 2.4: 行動属性抽出の精度

	手法	行動	行動主	動作	対象	時間	場所
適合率	手法 A	81.17%	86.32%	98.13%	84.14%	87.96%	88.25%
	手法 B	57.89%	72.79%	82.98%	67.01%	76.40%	80.20%
	手法 C	73.21%	82.25%	97.11%	81.23%	80.04%	82.11%
再現率	手法 A	23.86%	26.38%	28.87%	24.77%	26.20%	26.02%
	手法 B	51.13%	69.13%	90.23%	62.11%	73.51%	77.67%
	手法 C	66.54%	80.11%	93.18%	76.57%	79.75%	81.02%
F-measure	手法 A	36.88%	40.41%	44.61%	38.27%	40.37%	40.19%
	手法 B	54.30%	70.91%	86.45%	64.47%	74.93%	78.91%
	手法 C	69.72%	81.17%	95.10%	78.83%	79.89%	81.56%

日、映画を見に行く)であると判断する。この手法の利点は、Twitter 文書の文法に依存せず、高速に結果を出力できるところにある。しかし、この手法では頻度が低い興味(行動)は、獲得できない。表 2.5 は本論文の提案手法と Nilanjan らの手法の比較実験の結果を示す。この結果に基づくと、Nilanjan らの手法では、適合率が高いが、再現率は非常に低い。

表 2.5: Nilanjan らの手法 [40] との比較

	手法	行動
適合率	Nilanjan et al. (動作と対象の共起頻度 ≥ 3)	96.77%
	提案手法	73.21%
再現率	Nilanjan et al. (動作と対象の共起頻度 ≥ 3)	1.12%
	提案手法	66.54%
F-measure	Nilanjan et al. (動作と対象の共起頻度 ≥ 3)	2.21%
	提案手法	69.72%

2.6.4 実験結果の考察

上記の実験結果から、提案手法は一定の精度で日本語の文中に現れる行動属性を自動的に抽出できる。緊急時にリアルタイムで行動を手で抽出するのは困難であるので、提案手法は重要な役割を果たすだろう。そして、提案手法では、既存の学習モデル(CRF)を採用しているが、以下の工夫点および効果を挙げられる。

- 訓練データを自動的に作成するため、ラベル付けのコストがかからない。

- 自己教師あり学習の適用と訓練データの追加を行うことにより，F-measure はベースライン手法と比べて大幅に向上した．
- 行動の頻度に依存しないため，マイノリティ行動も抽出できる．
- 行動のドメインに依存しないため，震災だけでなく他の分野の行動も抽出できる．

本実験では，1000文を対象とした行動の抽出時間は約0.2秒であった．本実験では一台の計算機による実験であったが，複数台の計算機を用いた分散処理を導入することで，大量なつぶやきの文書を高速に処理することが可能である．具体的には各計算機に，一部のおつぶやきの文章を割り当て，並列処理によって行動を抽出する．したがって，本手法は緊急時に大量なつぶやきの文書を迅速に処理可能である．

2.3節に示した行動属性抽出の難しさに対して，提案手法は以下の対策を行っている．

1. 自己教師あり学習を適用し，行動を自動的に抽出する．
2. 前処理により，文中にあるノイズ文字列を削除している．そして，複雑かつ正しい文法で記述されていない文に対応するために，文の単純化と訓練データの追加を行っている．
3. 行動の頻度ではなく，文の構成と単語の品詞に基づいて行動属性のラベルを推測する．
4. 行動のドメインに依存しないヒューリスティック（構文パターン）を作成する．

2.7 関連研究

関連研究としては、次のようなものが挙げられる。

- マイクロブログに関する研究
- Web から人間行動の抽出
- Web からリレーション抽出
- オープンリレーション抽出

以下に各研究における手法を説明し、本論文が提案する手法（以下、本手法）との比較を行う。

2.7.1 マイクロブログからの人間行動の抽出

我々の調べた結果では、マイクロブログを対象にして人間行動の抽出・解析を行う研究は少なく、Mor ら [37] と Nilanjan ら [40] 挙げられる。

Mor ら [37] は、Twitter ユーザがどんなこと（自分の状況であるか、意見であるかなど）について発言するかを調査した論文である。調査方法は、“Me now”や“Opinions/Complaints”などの9つのカテゴリに基づいて、ユーザが発言したメッセージの内容を分類するというものである。調査結果によると、自分の現状について発言しているメッセージは全体の約41%を占める。

Nilanjan ら [40] は、人間の行動に着目して、Twitter ユーザの興味をリアルタイムで把握する手法を提案している。この手法では、まずユーザの興味に関する「カテゴリワード」（例えば、movie, cinema, music, sports など）、「動詞」（例えば、watch, watching, party など）、「時間ワード」（例えば、tonight, today, weekend など）を作っておく。次に、これらのワードに基づいて、Twitter メッセージに対して {カテゴリワード, 動詞} と {カテゴリワード, 時間ワード} の共起頻度を計算する。共起頻度が高いものはユーザの興味があるものと判断する。例えば、“movie”（カテゴリワード）と“go”（動詞）と“tomorrow”（時間ワード）の共起頻度が高ければ、ユーザの興味は“going to a movie tomorrow”（明日、映画を見に行く）であると判断する。

この手法の利点は、Twitter 文書の文法に依存せず、高速に結果を出力できるところにある。しかし、出力された結果の妥当性を評価していない。また、この手法は以下のような問題点があると考えられる。

1. この手法では頻度が低い興味（行動）は、獲得できない。しかし、頻度が低い興味（行動）でも、重要な経験や意外性を含んでいる可能性は高いと言える。また、頻度が高い興味しか抽出できないので、この手法の再現率は低いと考えられる。
2. “ today ”や“ tomorrow ”などのワードの頻度だけでは、興味（行動）をリアルタイムで把握できるとは言えない。これは、ユーザが発言したメッセージの時刻を考慮しなければ、いつに対しての“ today ”や“ tomorrow ”であるかが分からないためである。更に、メッセージに記述された行動主を把握していないので、誰の興味であるかも分からない。
3. 頻度に基づくため、多くのデータ（メッセージ）から計算しないと正確な結果を得ることができない。

上記のように、我々が調べた限りでは、マイクロブログを対象として、人間の行動を抽出・解析する研究はまだ数が少なく、多くの問題点がある。しかし、タイムスタンプを利用した研究はいずれかにあると思われる。しかし、タイムスタンプに基づいて、行動間の関係を把握するには以下のような問題がある。

1. 行動の発生がメッセージの発言された時刻と逆順である場合、行動間の遷移関係が誤って判断されてしまう。例えば、Twitter から 2.6 のような 2 つのメッセージを取得できるとする。メッセージ 1 では (動作 = 行く, 対象 = 秋葉原, 時刻 = 来週) (行動 A) という行動を表す。メッセージ 2 では (動作 = 食べる, 対象 = ラーメン, 時刻 = 昨日, 場所 = 渋谷) (行動 B) という行動を表す。

Twitter のタイムスタンプ (18:00, 18:30) に基づくと、行動 A が起こった後に、行動 B が起こるという間違っただ判断になってしまう。しかし、我々の提案手法では、文中に表す時刻 (来週と昨日) を解析すると、行動 B が起こった後に行動 A が起こると正しい判断が可能である。つまり、行動間の遷移関係を正確に把握するためには、メッセージの内容を解析する必要があると思われる。

表 2.6: Twitter のメッセージ

No	Twitter に 発言した時刻	行動を表す文
1	18:00	来週, 秋葉原へ行く予定。
2	18:30	昨日, 渋谷でラーメンを食べたよ。

2. また, タイムスタンプを利用することで, ある程度, 行動間の遷移(流れ)を推論できるが, 行動間の因果関係の推論は非常に難しい. 我々の提案手法では, “ ので ”, “ ため ” のような因果関係を表すパターンを訓練データの特徴モデルに加えることにより, 行動間の因果関係を抽出できる. 例えば, “ 俺は iPhone を買うために, 秋葉原へ行ったよ ” の文に対して, “ 動詞 - ために ” というパターンを利用することで, 以下の出力を得ることができる.

行動 A: (行動主 = 俺, 動作 = 買う, 対象 = iPhone)

行動 B: (行動主 = 俺, 動作 = 行った, 対象 = 秋葉原)

行動 B BecauseOf 行動 A (行動 B が起こる原因は行動 A である).

更に, 本論文では, あまり頻度の高くない行動(意外の行動, 珍しい行動, 貴重な経験等)に対しても, 文に明確に記述されていれば抽出可能である. また, “ だから ”, “ したがって ”, “ そして ”, “ それから ”, などの接続詞を利用することで, 文間に表す行動間の関係も把握できる. 例えば, “ 俺は先週秋葉原へ行った. それから, アバターを見に行っちゃったよ. ” の文では, “ それから ” というパターンを用いると, 以下の関係が分かる.

行動 A: (行動主 = 俺, 動作 = 行った, 対象 = 秋葉原)

行動 B: (行動主 = 俺, 動作 = 見に行っちゃった, 対象 = アバター)

行動 A Next 行動 B (行動 A が起こった後に, 行動 B が起こった).

しかし, 現時点では, 暗黙的な行動の関係は抽出できていない. 例えば, “ 俺は先週渋谷にある回転寿司店へ行った. すごく美味しかったよ! ” の文では, “ 寿司を食べた ” という行動が隠れており, 抽出できない. 今後, これらの課題を解決するために, 推論, 相関ルールの適用などを検討している.

2.7.2 Web からの人間行動の抽出

Web からの人間行動抽出の先行研究には, Perkowitz ら [42], 川村ら [24], 倉島ら [29] と Fukazawa ら [15] の研究がある. 本論文の手法とこれらの研究の手法との主な比較は表 2.7 の通りである.

表 2.7: 行動抽出の関連研究との比較

手法等	抽出可能属性	セットアップコスト	対応可能文	因果関係
Perkowitz ら	×		×	不可
Fukazawa ら				不可
倉島ら				N/A ⁴
川村ら				不可
本論文の手法				可能

Perkowitz ら [42] の手法では, 物の作り方のウェブページ (eHow[11]) を対象にして, 利用オブジェクト, 時間, 行動間の順序を抽出する. しかし, 行動属性の抽出は単純なキーワードマッチなので, 作業の手順 (お茶の作り方など) を明示的に書いたウェブページにしか対応できない (図 2.21). また, 文の係り受け関係を解析しない.

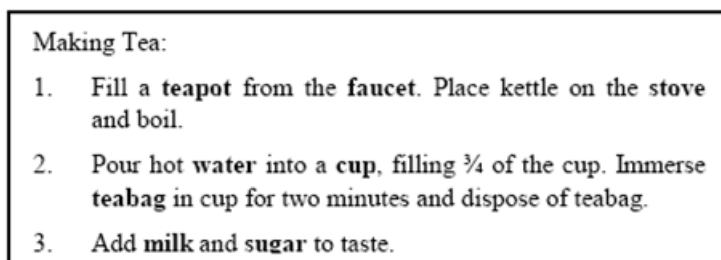


図 2.21: お茶の作り方

川村ら [24] の手法では, まず CGM³ から各トピック (ハリーポッターなど) を対象にして, 行動オントロジーと商品オントロジーを用い動作と対象を抽出する (図 2.22). 続いて, オントロジーを参照して行動属性を結合することで, 行動相関マップを自動的に作成

⁴論文上では, 行動間の因果関係の抽出を述べていない. 自然言語処理ツールを用いて係り受けを上手く解析できると, ある程度行動間の因果関係を抽出できるが, 解析時間がかかる.

³Consumer Generated Media

する．最後に，行動属性の出現頻度と出現順序を考慮して，行動相関マップ内のリンクに重み付けを行う．

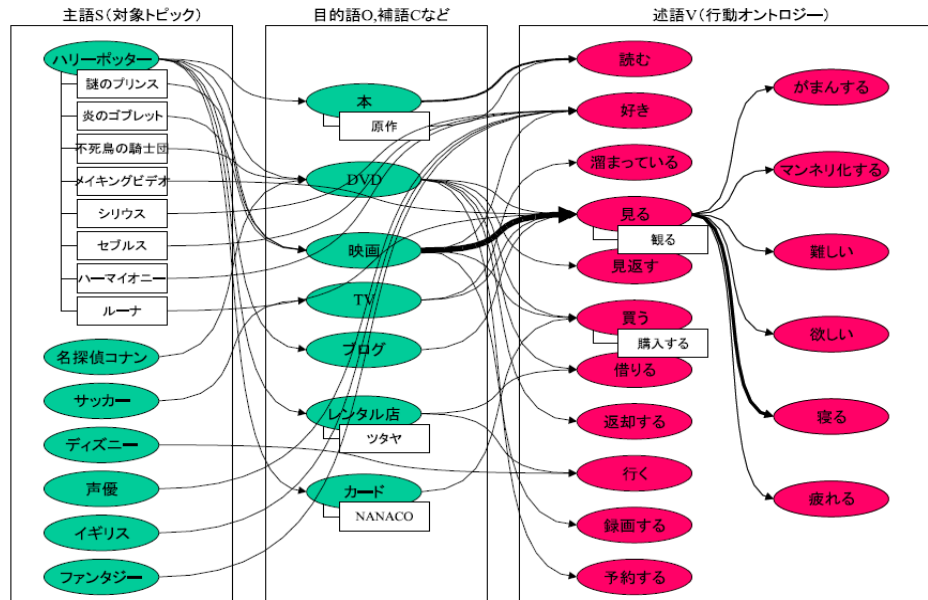


図 2.22: オントロジーを用いた行動属性の抽出

しかし，この手法では行動オントロジーと対象トピックに関する情報（商品名など）のオントロジーを予め準備しておく必要があり，抽出精度（適合率と再現率）はこれらのオントロジーに依存する．

図 2.23 は倉島ら [29] の手法を示す．この手法では，ブログの日付情報から時刻を取得するので，行動文に表す時刻ではない可能性が高い．場所は，固有表現抽出器で“地名”，“組織”と判定される語なので，動作と係り受け関係がない可能性がある．対象と動作の抽出では，係り受けと各分析の自然言語処理ツール（JTAG[14]）を用いる．この方法は JTAG の精度に依存することとなる．また，助詞“を”と“に”が共にある文に対応できない．更に，Banko ら [5] が指摘するように，係り受け解析の自然言語処理ツールを直接用いてエンティティ（行動属性など）の相互関連を判定するのは Web コーパスに適切ではない．Fukazawa ら [15] の手法では，まず「ドメイン + 助詞（を，に） + 動詞」というパターンを用いて，検索エンジンでドメインと動詞を取得する．次に， $Score(\text{ドメイン}, \text{動詞})$ を計算し， 10^{-5} より大きければこのドメインは対象，この動詞は動作として獲得する．

$$Score = \frac{Hits(\text{ドメイン AND 動詞})}{Hits(\text{ドメイン})Hits(\text{動詞})} \quad (2.3)$$

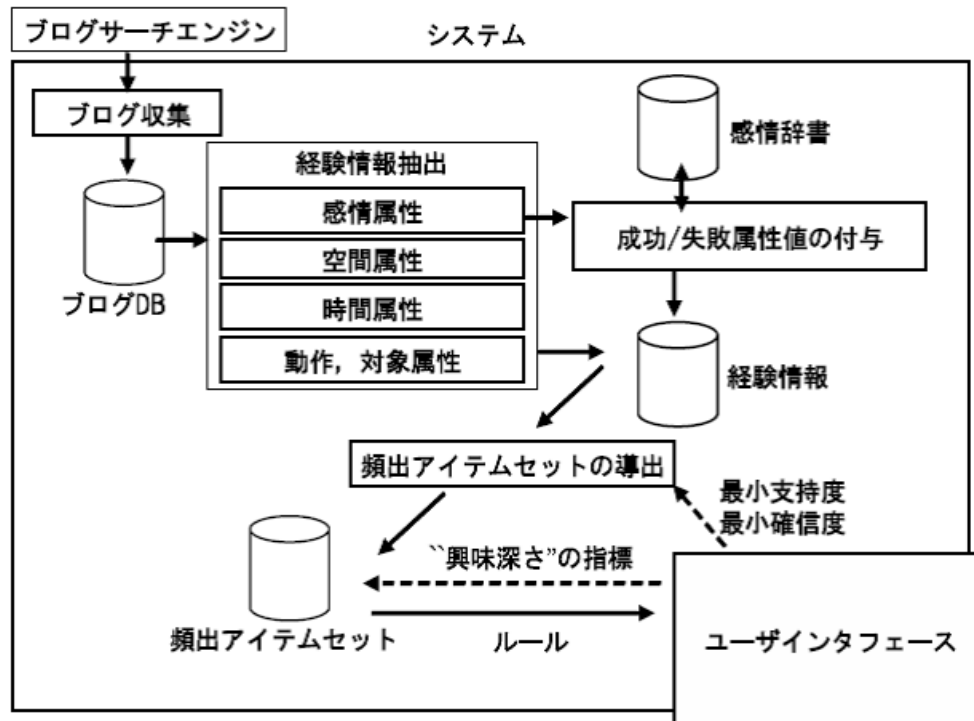


図 2.23: 大規模テキストからの経験マイニング

この手法の利点は、検索エンジンだけ利用することで、対象と動作のペアを獲得できる。しかし、パターンを特定しており、再現率が非常に低いと考えられる。また、価値がある行動パターンでも、共起頻度が低ければ獲得できない。

2.7.3 Web からのリレーション抽出

Web からの関係抽出の代表的な研究として, DIPRE[9], SnowBall[2], KnowItAll[12] が挙げられる. 本論文の手法とこれらの研究の手法との主な比較は表 2.8 の通りである.

表 2.8: Web からの関係抽出の手法との比較

	DIPRE[9]	SnowBall[2]	KnowItAll [12]	本論文の手法
初期シード	必要	必要	必要	必要がない
リレーションの定義	必要	必要	自動発見	自動発見
リレーションタイプ	Binary	Binary	Unary/Binary	5つの行動属性
学習手法	Bootstrapping	Bootstrapping	Bootstrapping	Self-Supervised
学習モデル	パターンマッチ	SVM	Naive Bayes	CRF
シングルパス	NO	NO	NO	YES

DIPRE[9] は Google 創業者である Brin 氏が提案した, Web ページを対象にしたブートストラッピングである. 図 2.24 は手法の手順を示す. 詳細は以下の通りである.

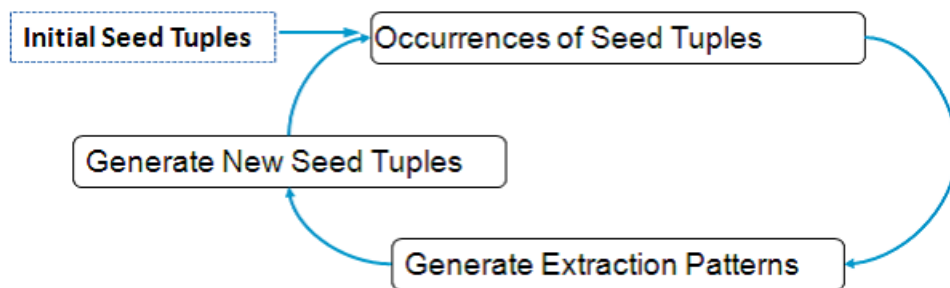


図 2.24: ブートストラッピング

1. Initial Seed Tuples: 表 2.9 に示すようなリレーションのインスタンス (初期シード) をいくつか用意しておく.
2. Occurrences of Seed Tuples: 上位の組を検索エンジンで検索し, 図 2.25 に示すようなリレーションのインスタンスを含む文を取得する.

表 2.9: 組織の住所のシード

仕組み	場所
Microsoft	Redmond
IBM	Armonk
Boeing	Seattle
Intel	Santa Clara

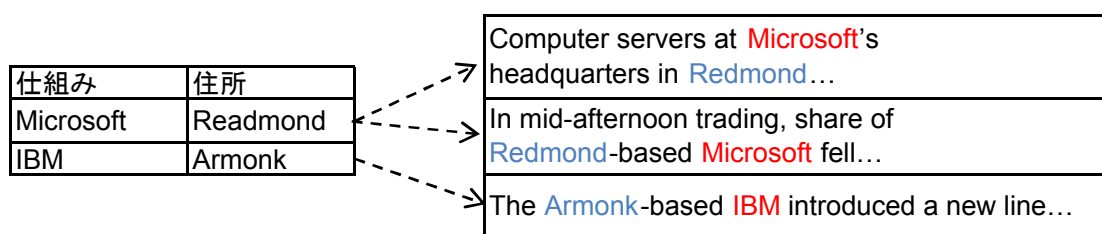


図 2.25: 組織と住所の組を含む文

3. Generate Extraction Patterns: 取得した文から，図 2.26 に示すような正規表現パターンを抽出する．

```
<STRING1> 's headquarters in <STRING2>
<STRING2> -based <STRING1>
```

図 2.26: リレーションの正規表現パターン

4. Generate New Seed Tuples: 抽出した正規表現パターンにマッチするインスタンスをシードに追加する．(2) へ戻る．

ブートストラッピングの利点は，単純なパターンマッチで困難であった，情報を抽出するためのパターンを自動生成することができることにある．一方で，図 2.27 に示すように，ブートストラッピングの欠点としては，「欲しい情報の周辺のパターンを誤って抽出する」

「誤ったパターン周辺の欲しくない情報を抽出する」 「欲しくない情報の周辺の誤ったパターンを抽出する」という問題点がある．また，パターンを自動生成するための方策がヒューリスティクスであり，そのため，意外性や発見性のあるパターンの生成が難しいことも欠点として挙げられる．

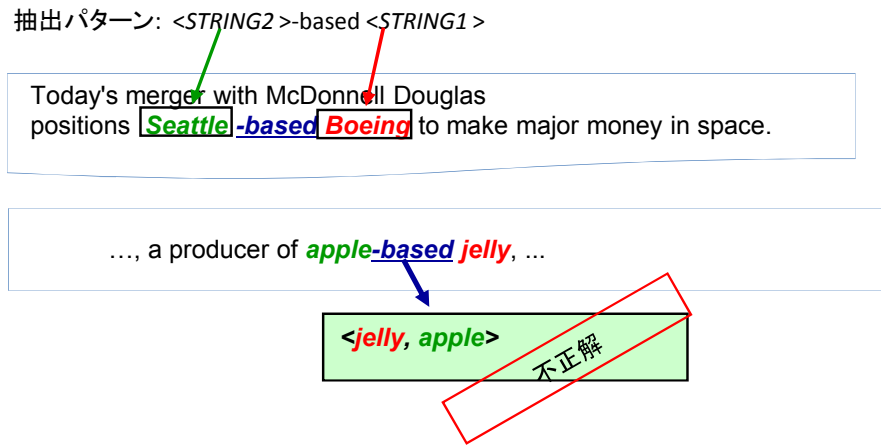


図 2.27: ブートストラッピングの問題点

上記の問題点に加え、DIPRE は正確なパターンマッチ手法を採用するため、限定的なパターンが生成されると、新たなインスタスが見つけれない可能性もある。SnowBall[2] は DIPRE の問題点を解決するために、以下の改良点を工夫した手法である。

- 固有名詞抽出ツールを用いて、パターンと文のタグを抽出する。そして、パターンと文を図 2.28 に示すようにベクトル化する。

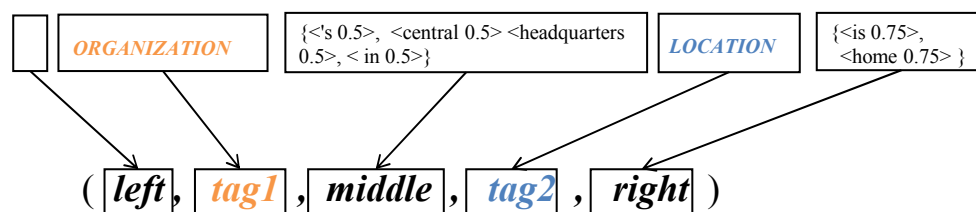


図 2.28: パターンと文をベクトル化

- SVM(Support Vector Machine)[21] を用いて、図 2.29 に示すようにパターンと文のベクトルの類似度を計算する。
- パターンとタプルを選定する前に、図 2.30 に示すように互いに評価する。評価のアイデアとして、良いパターンは多くのポジティブなタプルを生成する。また、正確なタプルは多くの良いパターンから生成される。

$$\begin{aligned}
 \text{パターン}(P) &= ([Lp], \text{tag1}, [Mp], \text{tag2}, [Rp]) \\
 \text{文}(S) &= ([Ls], \text{tag1}, [Ms], \text{tag2}, [Rs]) \\
 \text{Match}(P, S) &= \begin{cases} [Lp] \cdot [Ls] + [Mp] \cdot [Ms] + [Rp] \cdot [Rs] & \text{タグにマッチする場合} \\ 0 & \text{その他} \end{cases}
 \end{aligned}$$

図 2.29: パターンと文のベクトルの類似度

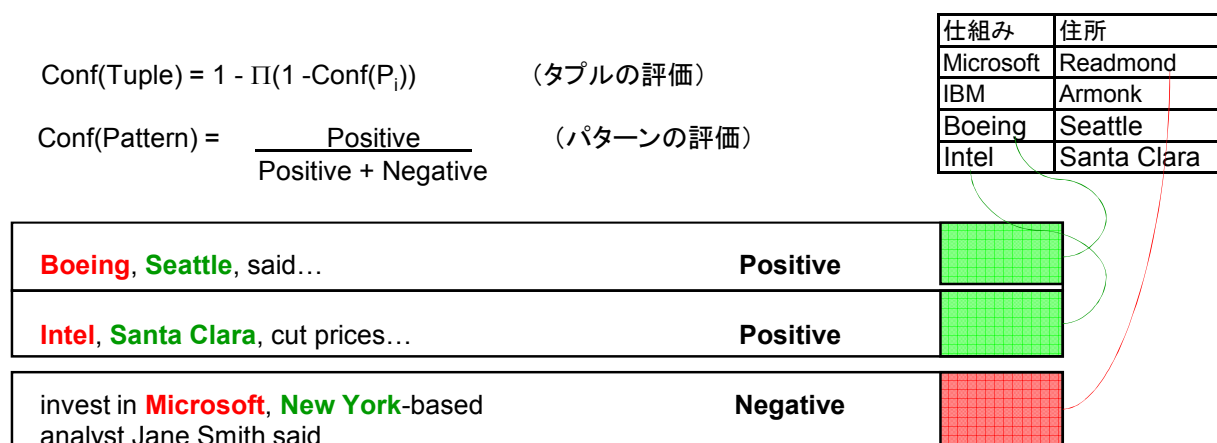


図 2.30: SnowBall の評価方法

図 2.31 に示すように, KnowItAll[12] は DIPRE, SnowBall と同様にブートストラッピングを利用するが, ドメインに依存しないパターンと NP Chunker を用い, Web から全てのリレーションを抽出するという点で異なる. しかし, 大量のクエリと Web ページのダウンロードが必要であるため, リレーション抽出には数週間がかかってしまうという問題点がある.

2.7.4 オープンリレーション抽出

Banko ら [5] によると, Web からリレーション抽出は Traditional Relation Extraction(以下 Traditional RE) と Open Relation Extraction(以下 Open RE) に分ける. サブセクショ

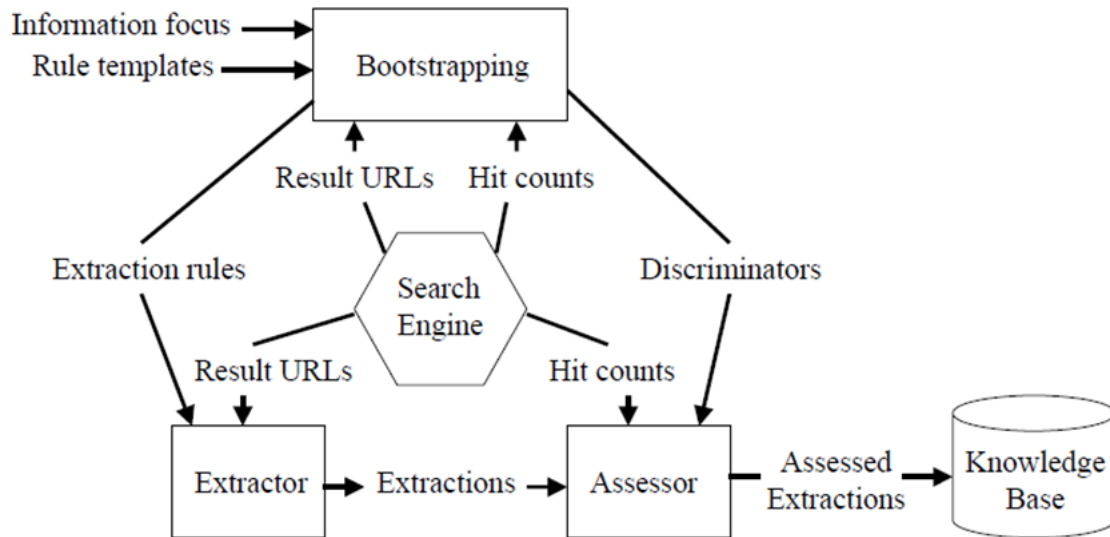


図 2.31: KnowItAll の手法

ン 2.7.3 に説明した手法は Traditional RE である．表 2.10 に示すように，Traditional RE の特徴として，各ドメインごとにリレーションを予め定義しておく必要がある．これに対して，Open RE はリレーションを定義せず，抽出過程でリレーションを発見する．つまり，OpenRE ではすべてのリレーションを検査し抽出する．また，Traditional RE との相違点として，Open RE では抽出しようとするリレーションのインスタンス（初期シード）を作っておく必要がない点が挙げられる．

表 2.10: Traditional RE と Open RE の比較

	Traditional RE	Open RE
入力	コーパスとシード	コーパス
リレーション	定義しておく	自動発見
複雑さ	$O(D \times R)$	$O(D)$ ⁴¹

Open Relation Extraction の概念は Banko ら [5] の研究グループにより定義されたものである．Open RE の実現に向けて，Banko らは 2007 年に TextRunner[6]，2008 年に O-CRF[5] を提案している．O-CRF では，リレーション抽出問題を系列ラベリング問題に置き換える

⁴¹D はドキュメントの数，R はリレーションの数

という点について既存研究と大きく異なる．これにより，表 2.11 に示すように，O-CRF は TextRunner より精度が高い．以下に各手法の詳細を解説する．

表 2.11: TextRunner と O-CRF の比較

	TextRunner	O-CRF
Open RE	YES	YES
リレーションタイプ	Binary	Binary
学習手法	Self-Supervised	Self-Supervised
学習モデル	Naive Bayes (分類)	CRF (系列ラベリング)
外部の 自然言語処理ツール	NP Chunker Dependency Parser (係り受け解析)	NP Chunker Dependency Parser (係り受け解析)
適合率 (Precision)	86.6%	88.3%
再現率 (Recall)	23.2 %	45.2 %
F-measure	36.6%	59.6%

1. TextRunner (University of Washington) [6]

図 2.32 に示すように，TextRunner は Leaner，Extractor，Assessor という 3 つのモジュールに分割される．

- Leaner モジュールでは，小さいコーパスから外部の自然言語処理ツール (NP Chunker，係り受け解析ツール) とヒューリスティクスを用いて，訓練データを自動的に作成する．
- Extractor モジュールでは，Naive Bayes という学習モデルを用いて，訓練データの特徴を学習し特徴モデルを作成する．そして，この特徴モデルを用いて，Web コーパスからタプルを抽出する．
- Assessor モジュールでは，抽出したタプルの正規化 (Normalization) と同定 (Synonym Resolution) を行い，リレーションを出力する．

2. O-CRF (University of Washington) [5]

O-CRF は TextRunner のアーキテクチャ (図 2.32) を採用し，Naive Bayes の代わりに CRF という学習モデルを使っている．現時点では，英語の Web ページからバイ

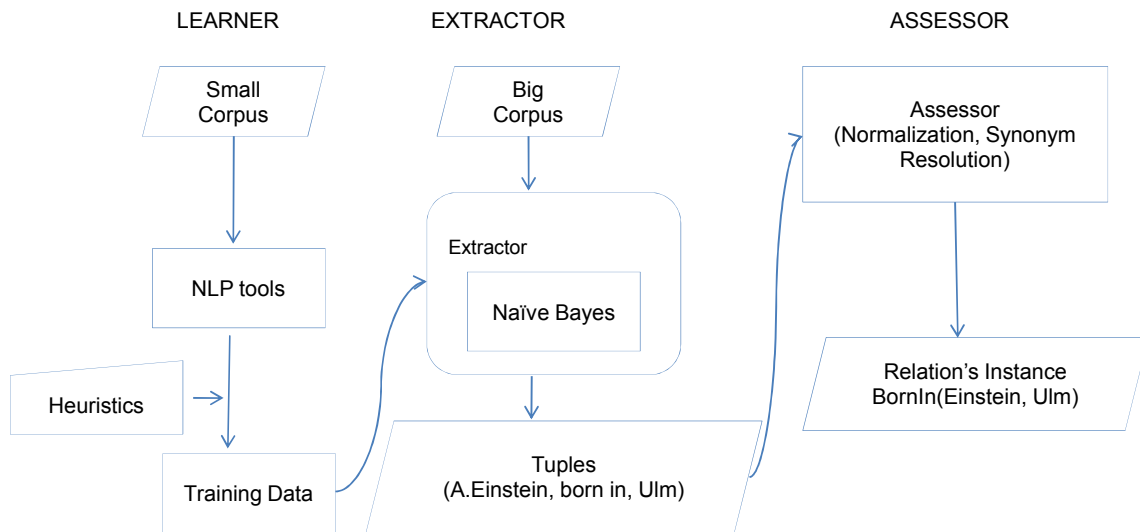


図 2.32: TextRunner のアーキテクチャ

ナリリレーションを抽出する手法として、O-CRF は Open RE の中で抽出の精度が最も高い手法である。しかし、O-CRF は以下のような課題が残っている [4]。

- (a) 英語の文を対象としているので、抽出文の構成は主語 - 動詞-述語 (Subject-Verb-Object) のみとなる。日本語のような自由度が高い言語を対象にする場合、様々な工夫が必要となる。
- (b) リレーションはエンティティと同文に含まれかつ、エンティティの間に出現する必要がある。つまり、文の構成は {Entity1, Relation, Entity2} である（例えば、“ Albert Einstein was born in Ulm . ”のような文）。そして、NP Chunker を用いエンティティを事前に判定しておく必要がある。
- (c) エンティティを意味的に分類しない。
- (d) 既存の研究と同様に、バイナリリレーションを対象にしているため、3 つ以上の要素から構成されるリレーション (Higher-Order Relation) は検討されていない。例えば、“ Bill Gates is the Chairman at Microsoft ”の文では、3 つの要素から構成となるリレーション (Bill Gates, Chairman, Microsoft) を抽出するべきであるが、現時点では、2 つのバイナリリレーション (Bill Gates, Chairman)

と (Bill Gates, Microsoft) を順序に抽出する必要である。

- (e) 暗黙の行動を抽出できない。例えば，“お寿司が大好きなので，和食の店へ行こうよ。”の文では，和食の店へ行って「お寿司を食べる」という暗黙の行動がある。しかし，文にこの行動を記述されていないので，抽出できない。

表 2.12: O-CRF と本論文の手法の比較

	O-CRF	提案手法
Web ページの言語	英語	日本語
抽出対象	リレーション	人間の行動
タプルの要素の数	3	2~6
対応可能な構文	S-V-O	{O, C}, V S, {O, C}, V {O, C}, V, S SガVハ{O,C} SガVCハO SハNガV ヲN Nガ(ハ)V NヲN二
リレーションはエンティティの間に出現する必要がある	YES	NO
エンティティを事前に判定しておく必要がある	YES	NO
エンティティを意味的に分類する	NO	YES

本論文の手法は，一つのラベルだけではなく，文に現れるすべての行動属性と行動間の遷移のラベルを推定する必要がある。また，日本語の文を対象するので，文によって属性の数と位置は変わる。このため，本手法では，エンティティの事前判定，又はリレーションがエンティティの間に現れるといった設定ができない。つまり，O-CRF の解決課題よりも本手法が解決する課題が困難であると考えられる。表 2.12 は O-CRF との比較を示す。

第3章 時系列行動ネットワークの構築

本章では、まず時系列行動ネットワークの定義について説明する。次に、Web オントロジー言語 (OWL) に基づいた時系列行動ネットワークの設計を説明する。そして、設計した時系列行動ネットワークの構築方法を説明する。更に、SPARQL (SPARQL Protocol and RDF Query Language) [49] を用いて、設計した時系列行動ネットワークを確認する。最後に、関連研究について説明し、本研究との比較を行う。

3.1 時系列行動ネットワークの定義

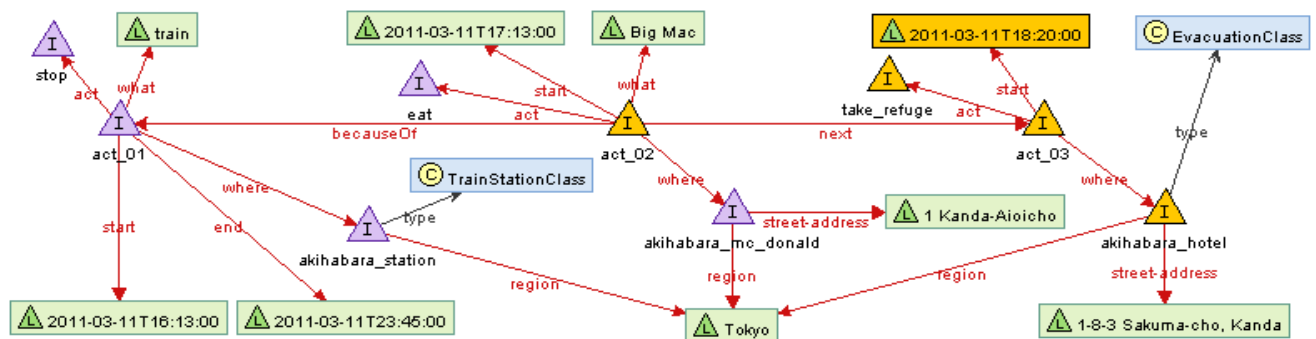


図 3.1: 時系列行動ネットワークの一部 (C, I, L はそれぞれクラス, インスタンス, ラベルを表す)

行動ネットワークとは、行動と行動間の関係を表すネットワーク (有向グラフ) である。前章で抽出した行動属性を本ネットワークを用いて形式化する。特に、緊急時を考慮すると時間情報が重要となるため、各行動の時刻を明確に記述し、時系列的に整理する。そこで、我々は行動ネットワークに時間情報を加えて、時系列行動ネットワークを行動属性の概念 (動作, 対象, 時間, 場所) と行動属性間の関係から成るものと定義する。図 3.1 に

示すように，ノードは行動属性の概念，リンクは行動属性間の関係を表す．行動間のリンクは行動間の遷移関係又は行動間の因果関係を表し，有向リンクである．

3.2 時系列行動ネットワークの設計

時系列行動ネットワークはトリプルデータ（行動属性 - 行動属性間の関係 - 行動属性，行動 - 行動間の関係 - 行動）から成る（図 3.2）．そのため，本研究では RDF（Resource Description Framework）[47] を用いて，行動ネットワークのデータを表現する．W3C が提案した N3[48] はコンパクトで可読性が良いことから，我々は N3 を用いて時系列行動ネットワークのデータ（行動属性，行動属性間の関係）を記述する．

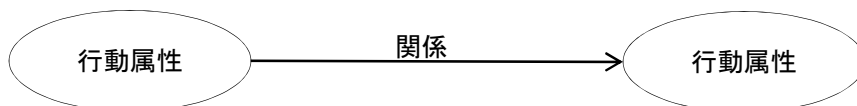


図 3.2: トリプルデータから成る時系列行動ネットワーク

図 3.3 に示すように，拡張性を向上するために，時系列行動ネットワークは GeoNames [16]，Time line [44]，と vCards [18] という外部のオントロジーを継承する．具体的には，場所の経度と緯度を GeoNames で記述する．また，時間情報を Time line で記述する．そして，場所のアドレスを vCards で記述する．

```

@prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#> .
@prefix tl: <http://purl.org/NET/c4dm/timeline.owl#> .
@prefix vcard: <http://www.w3.org/2006/vcard/ns#> .
  
```

図 3.3: 継承する GeoNames，Time line，vCards

図 3.4 は時系列行動ネットワークのクラスを示す．行動，動作，場所，対象のクラスはそれぞれ，ActionClass，ActClass，WhereClass，WhatClass である．場所のサブクラスとして ShopClass，RestaurantClass，TrainStationClass，EvacuationClass を作成する．

図 3.5 は時系列行動ネットワークのプロパティを示す．具体的には，act は動作を表すプロパティである．act のドメイン，レンジはそれぞれ ActionClass と ActClass である．

<pre> ### 行動クラス :ActionClass a owl:Class ; rdfs:subClassOf owl:Thing . ### 動作クラス (ActClass), 場所クラス (WhereClass), 対象クラス (WhatClass) :ActClass a owl:Class ; rdfs:subClassOf owl:Thing . :WhereClass a owl:Class ; rdfs:subClassOf owl:Thing . :WhatClass a owl:Class ; rdfs:subClassOf owl:Thing . </pre>	<pre> ### 場所のサブクラス :ShopClass a owl:Class ; rdfs:subClassOf :WhereClass . :RestaurantClass a owl:Class ; rdfs:subClassOf :WhereClass . :TrainStationClass a owl:Class ; rdfs:subClassOf :WhereClass . :EvacuationClass a owl:Class ; rdfs:subClassOf :WhereClass . </pre>
---	--

図 3.4: 時系列行動ネットワークのクラス

what は対象を表すプロパティである。what のドメイン、レンジはそれぞれ ActClass と WhatClass である。where は場所を表すプロパティである。where のドメイン、レンジはそれぞれ ActClass と WhereClass である。next は行動間の遷移関係を表すプロパティである。next のドメイン、レンジの両方は ActionClass である。becauseOf は行動間の因果関係を表すプロパティである。next と同様に、becauseOf のドメイン、レンジの両方は ActionClass である。

<pre> ### 行動属性 (動作, 対象, 場所) を表すプロパティ :act a owl:ObjectProperty ; rdfs:label "act" ; rdfs:domain :ActionClass ; rdfs:range :ActClass . :what a owl:ObjectProperty ; rdfs:label "what" ; rdfs:domain :ActionClass ; rdfs:range :WhatClass . :where a owl:ObjectProperty ; rdfs:label "where" ; rdfs:domain :ActionClass ; rdfs:range :WhereClass . </pre>	<pre> ### 関係 (遷移, 因果) を表すプロパティ :next a owl:ObjectProperty ; rdfs:label "next" ; rdfs:domain :ActionClass ; rdfs:range :ActionClass . :becauseOf a owl:ObjectProperty ; rdfs:label "becauseOf" ; rdfs:domain :ActionClass ; rdfs:range :ActionClass . </pre>
---	---

図 3.5: 時系列行動ネットワークのプロパティ

できるだけユーザに成功した行動を推薦するため、図 3.6 に示すように時系列行動ネットワークに行動の成功回数を追加する。success_cnt は行動の成功回数を表すプロパティである。success_cnt のドメイン、レンジの両方は ActionClass である。

また、複数のユーザが同じ行動を行った可能性があるので、時系列行動ネットワーク行動に行動の頻度を加える。図 3.7 に示すように、act_freq は行動の頻度を表すプロパティであ

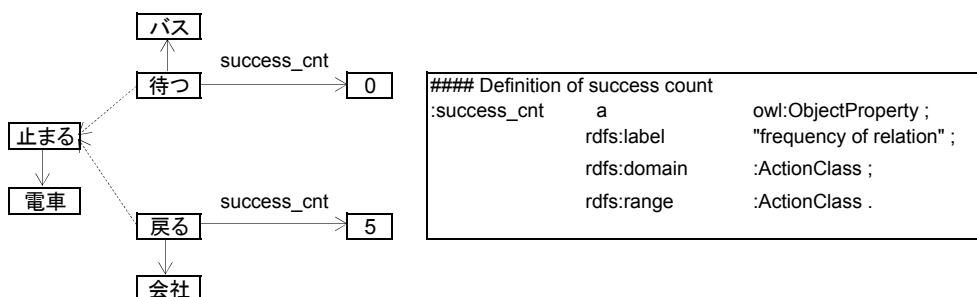


図 3.6: 行動の成功回数

る。success_cnt と同じように、act_freq のドメイン、レンジの両方は ActionClass である。

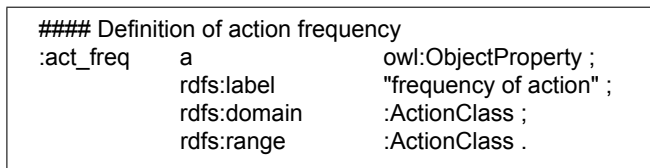


図 3.7: 行動の頻度を表すプロパティ

定義されたクラスとプロパティに基づき、時系列行動ネットワークのデータ（インスタンス）は図 3.8 のように記述できる。

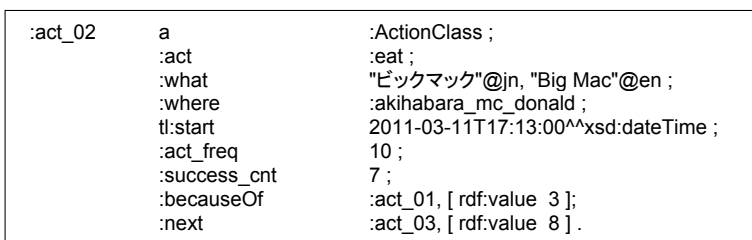


図 3.8: 時系列行動ネットワークのデータ（一部）

3.3 時系列行動ネットワークの構築

図 3.9 は時系列行動ネットワークの構築の流れを表す。以下，避難に関する時系列行動ネットワークの構築を例に説明する。



図 3.9: 避難に関する時系列行動ネットワーク構築の流れ

1. #hinan タグを用いて，避難に関するつぶやきを取得する。
2. 2.5 節で提案した手法を用いて，つぶやきの文中に現れる行動属性（動作，対象，時刻，場所）を自動的に抽出する。
3. つぶやきの文書中に現れる時間表現（なう，今晚，明日の朝など）とつぶやきの発生時刻を考慮して，行動の時刻を決定する。最後に，抽出された行動データ（行動属性と行動属性間の関係）を N3[48] で記述し，行動ネットワークを構成する。そして，我々は Sesame[1] を用いて，これらのデータを格納する。

3.4 評価実験

SPARQL を用いて、以下のようなクエリで時系列行動ネットワークの評価を行った。

- 任意時刻の行動を出力

図 3.10 は 2011-03-11T17:13:00 に開始している行動を出力するためのクエリを示している。

```

PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX tl:<http://purl.org/NET/c4dm/timeline.owl#>
PREFIX geo:<http://www.w3.org/2003/01/geo/wgs84_pos#>
PREFIX owl:<http://www.w3.org/2002/07/owl#>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX daml:<http://www.daml.org/2001/03/daml+oil#>
PREFIX vcard:<http://www.w3.org/2006/vcard/ns#>

SELECT DISTINCT ?act_name ?what ?location
WHERE {
  ?action tl:start "2011-03-11T17:13:00"^^xsd:dateTime .
  ?action :act ?act .
  ?act rdfs:label ?act_name .
  ?action :what ?what .
  ?action :where ?location .
  FILTER(lang(?act_name) = "jn")
}

```

図 3.10: 2011-03-11T17:13:00 開始の行動を出力するためのクエリ

クエリによる出力結果は図 3.11 の通りである。

act_name	what	location
"食べる"@jn	"ビックマック"@jn	:akihabara mc donald
"食べる"@jn	"Big Mac"	:akihabara_mc_donald

図 3.11: 2011-03-11T17:13:00 開始の行動

- 次の行動を出力

図 3.12 は 2011-03-11T17:13:00 に開始している行動の次の行動を出力するためのクエリを示している。

```

PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX tl:<http://purl.org/NET/c4dm/timeline.owl#>
PREFIX geo:<http://www.w3.org/2003/01/geo/wgs84_pos#>
PREFIX owl:<http://www.w3.org/2002/07/owl#>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX daml:<http://www.daml.org/2001/03/daml+oil#>
PREFIX vcard:<http://www.w3.org/2006/vcard/ns#>

SELECT DISTINCT ?name_next_act ?name_next_where
WHERE { ?action tl:start "2011-03-11T17:13:00"^^xsd:dateTime .
       ?action :next ?next_action .
       ?next_action :act ?next_act .
       ?next_act rdfs:label ?name_next_act .
       ?next_action :where ?next_where .
       ?next_where rdfs:label ?name_next_where .
}

```

図 3.12: 次の行動を探すクエリ

クエリによる出力結果は図 3.13 の通りである。

name_next_act	name_next_where
"避難する"@jn	"秋葉原ワシントンホテル"@jn
"避難する"@jn	"Akihabara Washington Hotel"@en
"take refuge"@en	"秋葉原ワシントンホテル"@jn
"take refuge"@en	"Akihabara Washington Hotel"@en

図 3.13: 次の行動

- 開放している避難所を出力

図 3.14 は、2011-03-11T17:00:00 に開放している避難所の場所名、住所、開放時刻、閉鎖時刻を出力するクエリを示している。

```
SELECT DISTINCT ?location_name ?street_address ?end_time
WHERE {?action :act :open .
      ?action tl:start ?start_time .
      ?action tl:end ?end_time .
      ?action :where ?location .
      ?location rdf:type :EvacuationClass .
      ?location rdfs:label ?location_name .
      ?location vcard:locality "Chiyoda-ku"@en .
      ?location vcard:street-address ?street_address .
      FILTER(?start_time <= "2011-03-11T17:00:00"^^xsd:dateTime &&
             ?end_time >= "2011-03-11T17:00:00"^^xsd:dateTime &&
             lang(?street_address) = "en" &&
             lang(?location_name) = "en"
      )
}
```

図 3.14: SPARQL で適切な避難所を探す例

クエリによる出力結果は図 3.15 の通りである。

location_name	"Akihabara Washington Hotel"@en
street_address	"1-8-3 Sakuma-cho, Kanda"@en
start_time	2011-03-11T16:00:00
end_time	2011-03-12T09:00:00

図 3.15: 開放している避難所の結果

以上の結果から、我々が設計した時系列行動ネットワークは OWL に従い、正しく設計・記述されていることが確かめられた。

3.5 関連研究

本研究と最も近い研究としては、MIT Media Lab の ConceptNet[20] とプリンストン大学の認知科学研究所が開発した WordNet[3] が挙げられる。

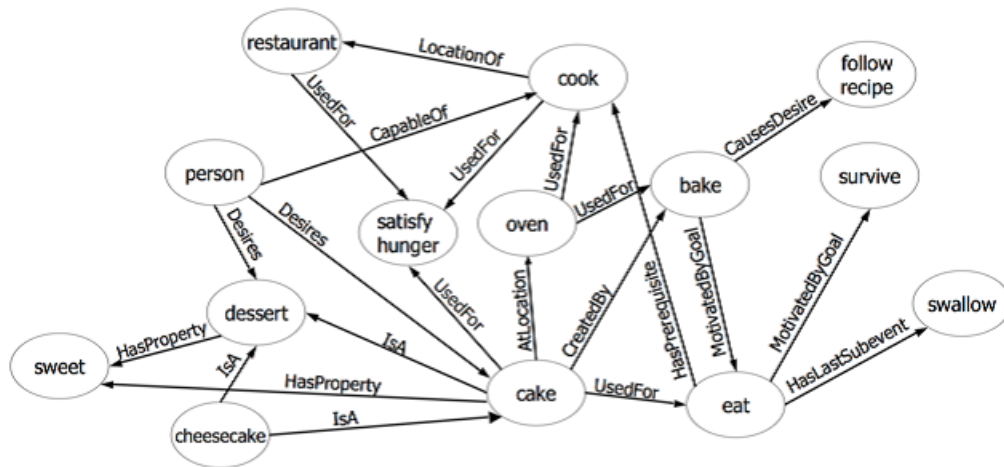


図 3.16: ConceptNet の一部

ConceptNet とは、Web を介して人手で集められる集合知である OpenMind CommonSense[35] のコーパスから、一般的な常識の意味ネットワークを構築する MIT Media Lab のプロジェクトである。図 3.16 は ConceptNet の一部を示す。

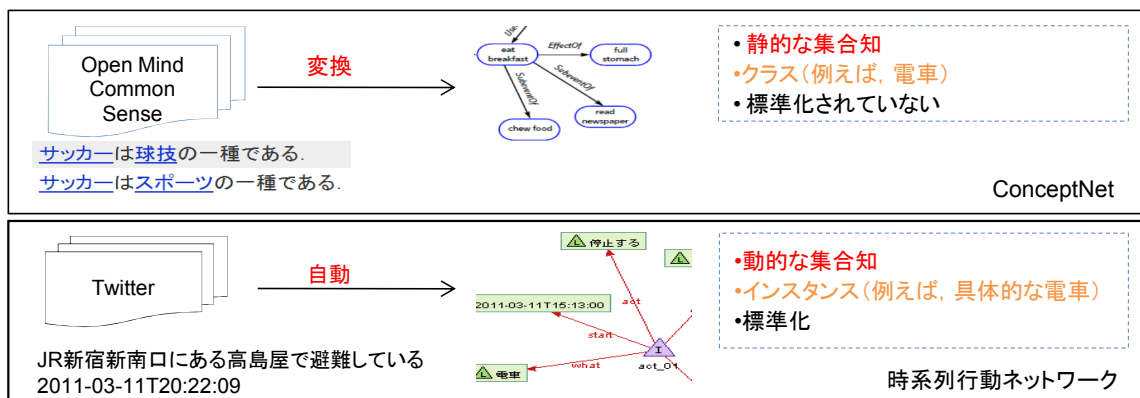


図 3.17: 時系列行動ネットワークと ConceptNet の比較

図 3.17 は ConceptNet と時系列行動ネットワークの比較を示す。ConceptNet が静的な集合知に対して、時系列行動ネットワークは時間の流れで動的に変化する集合知である。また、ConceptNet が行動のクラスまでを表すことに対して、時系列行動ネットワークは行動のインスタンスまで表す。

WordNet は、1985 年から作成されている英語の概念辞書（英単語の意味辞書）である。図 3.18 に示すように、WordNet では、英単語を synset と呼ばれる同義語のグループに分類し、各単語に対して、単語の意味と他グループに属する同義語との関係が定義されている。

WordNet Search - 3.0 - [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for:

Display Options:

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Noun

- [S:](#) (n) **car**, [auto](#), [automobile](#), [machine](#), [motorcar](#) (a motor vehicle with four wheels; *work*)
- [S:](#) (n) **car**, [railcar](#), [railway car](#), [railroad car](#) (a wheeled vehicle adapted to the rails)
- [S:](#) (n) **car**, [gondola](#) (the compartment that is suspended from an airship and that carries passengers)
- [S:](#) (n) **car**, [elevator car](#) (where passengers ride up and down) "*the car was on the*"
- [S:](#) (n) [cable car](#), **car** (a conveyance for passengers or freight on a cable railway) "*t*"

図 3.18: WordNet Search のインタフェース

第4章 時系列行動ネットワークの補完

本章では、まず4.1節で時系列行動ネットワークを補完する必要性について説明する。次に、4.2節で欠損行動を推測する際の課題を述べる。そして、4.3節で行動の性質とユーザのゴール（避難、帰宅など）を考慮した行動ベース協調フィルタリング手法を提案し、欠損行動を推測する。更に4.4節で協調フィルタリングと言語処理の統合を行い、欠損行動の推測手法を改善する。4.5節では、評価実験を行う。最後に、4.6節で関連研究と提案手法との比較を行う。

4.1 補完の必要性

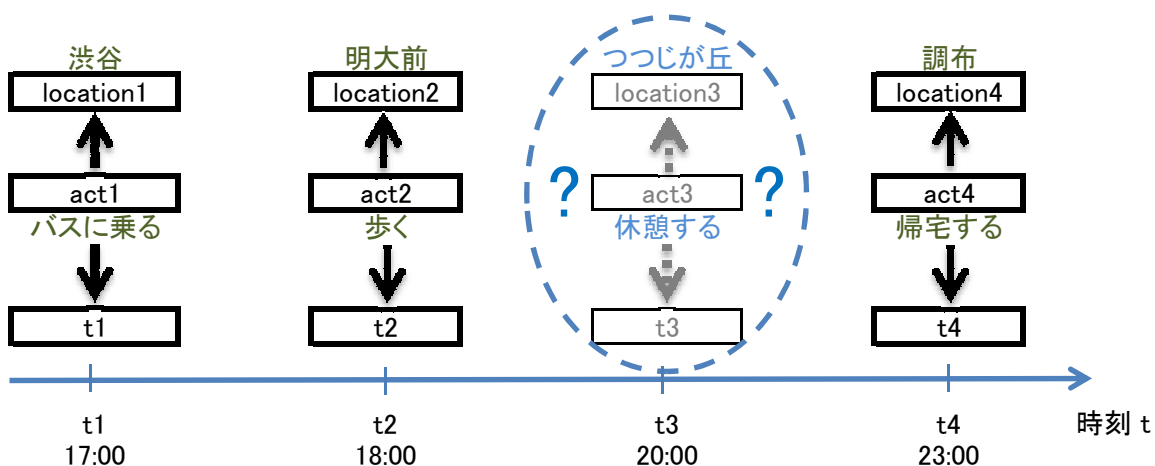


図 4.1: 20:00 時“何をやったか”をつぶやかなかった例

Twitter などのソーシャルメディアはユーザの自主性に依存しているため、つぶやかれなかった行動も数多く存在すると考えられる。つぶやかなかった行動があると、時系列行動ネットワークは欠損してしまう。例えば、図 4.1 に示すように、何らかの原因でユーザ

が20:00時に“どこで何をしたか”をつぶやかなかった場合，図4.2に示すように，時系列行動ネットワークは欠損してしまう．

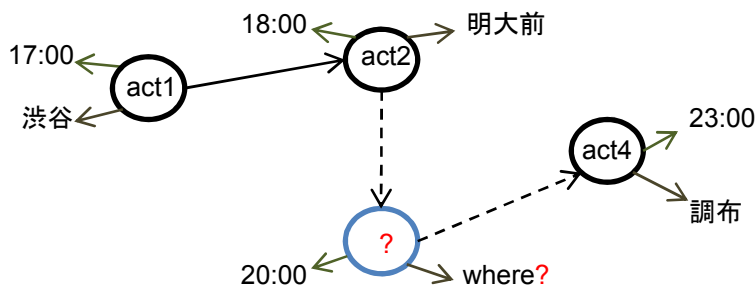


図 4.2: 時系列行動ネットワークの欠損の例

また，リアルタイムにユーザの行動に応じた情報を推薦するためには，これらの欠損行動を推測する必要がある．例えば，図4.3に示すように，時刻 t に対象ユーザが“トイレを探す”という行動を把握できれば，トイレに関する情報を推薦することが考えられる．

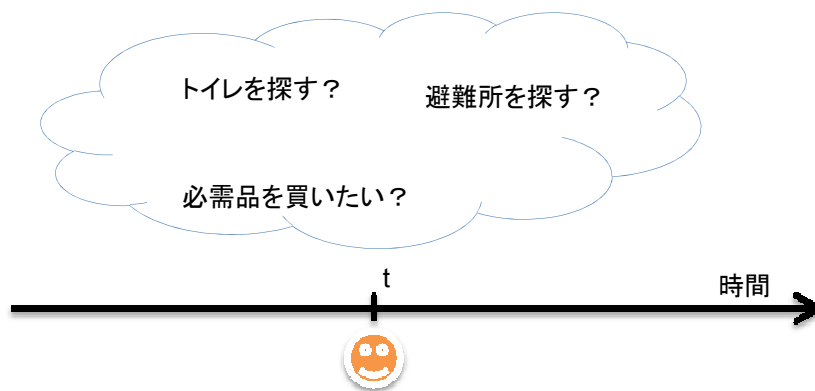


図 4.3: 時刻 t に対象ユーザの行動

4.2 欠損行動の推測の課題

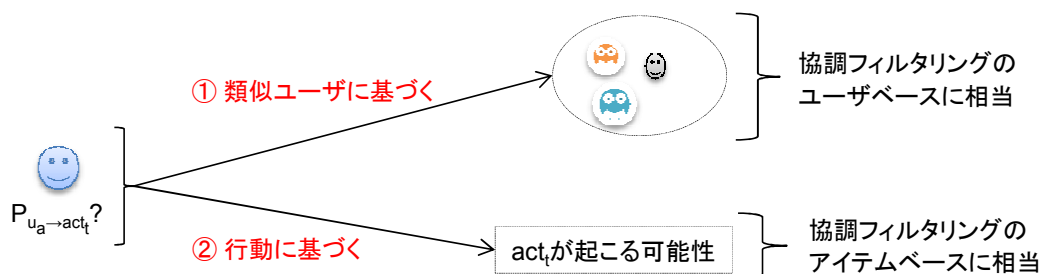


図 4.4: 欠損行動を推測するアプローチ

任意の時刻 t のユーザの行動を推測する問題は、時刻 t に最も行う可能性が高い行動を選ぶ問題に置き換えることができる。そして行動を選ぶ問題とは、時刻 t に対象ユーザ u_a が候補となる行動 act_t を実施する可能性 ($P_{u_a \rightarrow act_t}$) を計算する問題である。

図 4.4 に示すように、協調フィルタリングを利用し $P_{u_a \rightarrow act_t}$ を計算できる。しかし、従来の協調フィルタリングの目的は、ユーザの嗜好に合うアイテム (EC サイト内の商品など) を推薦することである。これに対して、本論文の目的は欠損行動を推測することである。また、表 4.1 に示すように、行動とアイテムは性質が異なる。したがって、震災に欠損行動を推測するために、従来の協調フィルタリングは利用できない。そこで、我々は行動の性質を考慮した協調フィルタリングを以下に提案する。

表 4.1: 『EC サイトのアイテム』と『震災時の行動』の比較

比較観点		EC サイトのアイテム	震災時の行動
複雑さ		1 つの変数	4 つの変数 (動作, 対象, 時刻, 場所)
依存関係	場所	NO	YES
	遷移関係	弱い	強い
実行時間の考慮		不必要	必要

4.3 欠損行動の推測

4.3.1 ユーザ間の類似度に基づく推測

我々は以下の考えに基づき、震災時におけるユーザ間の類似度を計る。

- 類似ユーザは同様の前後行動を行う可能性が高い。
- 同様のゴール（避難したい、帰宅したいなど）を持つユーザは同様の行動を行う可能性が高い。
- 震災時に、一定範囲内にいるユーザは同様の行動を行う可能性が高い。

したがって、震災時にユーザ間の類似度を式 4.1 に示す。ただし、公式内の各要素は以下の通りである。

- $Did(\{a_{before}, l_{before}\}, \{a_{after}, l_{after}\})$ は前後行動を行ったか否かを計る関数である。行った場合は 1、行わなかった場合は 0 になる。
- $SameTarget(a_t, l_t)$ は同様のゴールであるかどうかを判断する関数である。同じ場合は 1 になり、異なる場合は 0 になる。
- $SameLocation(l)$ は同じ場所であるかどうかを判断する関数である。同じ場合は 1 になり、異なる場合は 0 になる。
- β, γ はパラメータであり、 $0 \leq \beta, \gamma, \beta + \gamma \leq 1$ を満たす。

$$\begin{aligned}
 S(u_j, u_a) = & \beta Did(\{a_{before}, l_{before}\}, \{a_{after}, l_{after}\}) \\
 & + \gamma SameTarget(a_t, l_t) \\
 & + (1 - \beta - \gamma) SameLocation(l)
 \end{aligned} \tag{4.1}$$

4.3.2 行動が起こる可能性に基づく推測

実世界の行動は前後の行動，時間，場所，行動間の流れに依存するため，これらの要素を考慮することが必要である．したがって，行動 act_t が起こる可能性 $P(act_t)$ は式 4.2 で表わされる．ただし，公式内の要素は以下の通りである．

- $F(a_{before} \rightarrow act_t)$ は $a_{before} \rightarrow act_t$ の頻度である．
- $F(act_t \rightarrow a_{after})$ は $act_t \rightarrow a_{after}$ の頻度である．
- $F(act_t, t)$ は時刻 t に行動 act_t の頻度である．
- $F(act_t, l)$ は場所 l に行動 act_t の頻度である．
- ρ_a, ρ_t はパラメータであり， $0 \leq \rho_a, \rho_t, \rho_a + \rho_t \leq 1$ を満たす．

$$P(act_t) = \rho_a \{F(a_{before} \rightarrow act_t) + F(act_t \rightarrow a_{after})\} + \rho_t F(act_t, t) + (1 - \rho_a - \rho_t) F(act_t, l) \quad (4.2)$$

4.3.3 実行時間に基づく推測

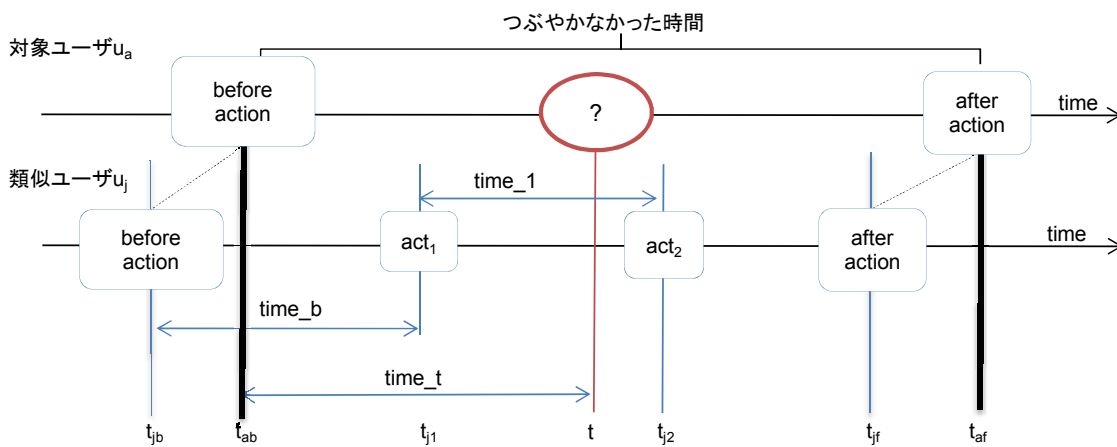


図 4.5: つぶやかなかった時間に複数の行動が行われた場合

実際には、対象ユーザ u_a はつぶやかなかった時間に複数の行動を行った可能性がある。例えば、図 4.5 では、対象ユーザはつぶやかなかった時間に“ before action ”(つぶやいた行動)を除いて、 act_1 と act_2 を行った可能性がある。このため、対象ユーザ u_a はつぶやかなかった時間中の任意の時刻 t に、まだ“ before action ”を行っていたか、又は act_1 と act_2 のどれかを行っていたかを推測する必要がある。この課題を解決するためには、類似ユーザ u_j の行動とその実行時間を考慮する必要がある。

類似ユーザ u_j が行った“ before action ”, act_1 と act_2 の開始時刻はそれぞれ、 t_{jb} , t_{j1} , t_{j2} であるとする。よって、“ before action ”の実行時間 $time_b$ は $[t_{jb}, t_{j1}]$ である。 $t_{ab} \sim t$ までの時間は $time_t$ とする。我々は act_1 の実行時間 $time_1$ と $time_b$ に基づいて、 act_1 を評価する。

$$T(u_j, act_1) = \begin{cases} \frac{time_t}{time_b} & (diff < 0) \\ 1 & (0 \leq diff \leq time_1) \\ \frac{time_1}{diff} & (time_1 < diff) \end{cases} \quad (4.3)$$

where, $diff = time_t - time_b$.

$0 \leq diff \leq time_1$ であると、対象ユーザ u_a が act_1 を行った可能性は最も高い。そして、 $time_t$ と $time_b$ の差が大きいくほど、対象ユーザ u_a が act_1 を行った可能性は減少する。具体的には、任意の時刻 t における act_1 の重みは式 4.3 で評価する。この重みが高いほど、行動が行われた可能性が高い。

4.3.4 欠損行動の推測式

式 4.1, 式 4.2, 式 4.3 を合わせて、 $P_{u_a \rightarrow act_t}$ を式 4.4 で求めることができる。ただし、各要素は以下の通りである。

- L は他のユーザの数である。
- $\omega(u_j, act_t)$ は行動 act_t に対するユーザ u_j の重みである。ユーザ u_j が act_t を行った場合は 1 になり、行わなかった場合は 0 になる。

- α はパラメータであり, $0 \leq \alpha \leq 1$ を満たす. 解決したい課題や行動データによって, これらのパラメータの値を決める.

$$P_{u_a \rightarrow act_t} = \alpha \left(\frac{\sum_{j=1, L} \omega(u_j, act_t) * S(u_j, u_a) * T(u_j, act_t)}{L} \right) + (1 - \alpha) P(act_t) \quad (4.4)$$

4.4 マイノリティ情報の活用（協調フィルタリングと言語処理の統合）

震災時には、成功した行動は低頻度だとしても価値があり重要である。したがって、大量の情報の中に埋もれがちなマイノリティ情報を活用すべきだと考えられる。しかし、協調フィルタリングでは、高頻度のパターンに基づくため、マジョリティ情報を容易に扱えるが、これらのマイノリティ情報に対応するのは困難である。そこで我々は以下のアプローチで成功行動を扱う。

1. 言語処理で成功行動を把握

図 4.6 に示すように、本論文では、“良かった”との係り受け関係を持つ行動は成功行動であると見なす。我々はこのパターンを用いて、つぶやきの文中に現れる成功行動を把握する。

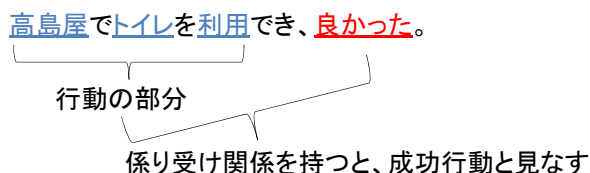


図 4.6: 言語処理で成功行動を把握

2. 対象ユーザが成功行動を行った可能性を推測

我々は以下の考えに基づき、対象ユーザが成功行動を行った可能性を推測する。

- 成功行動の数が多ければ多いほど、行われた可能性が高くなる（式 4.6 の $f(u_a)$ ）
- 人間は成功の回数が多い行動を選ぶ可能性が高い（式 4.7 の $Success(act_t)$ ）

したがって、対象ユーザ u_a が成功行動を行った可能性を式 4.5 で求めることができる。

$$DidSuccess_{u_a \rightarrow act_t} = f(u_a) * Success(act_t) \quad (4.5)$$

$$f(u_a) = \frac{\text{成功した行動の数}}{\text{行動の数}} \quad (4.6)$$

$$Success(act_t) = \frac{act_t \text{の成功回数}}{\text{全ての行動の成功回数の和}} \quad (4.7)$$

例えば， $t_{before} \sim t_{after}$ の間に， u_a が行った可能性がある行動は表 4.2 の通りであるとすると，式 4.8 と式 4.9 の値になる．

$$f(u_a) = \frac{2}{5} \quad (4.8)$$

$$Success(act_2) = \frac{5}{5+1} \quad (4.9)$$

表 4.2: $t_{before} \sim t_{after}$ の間に， u_a が行った可能性がある行動

行動	成功（良かった）	成功の回数	時間
act_{before}	0	0	t_{before}
act_1	0	0	t_1
act_2	1	5	t_2
act_3	1	1	t_3
act_{after}	0	0	t_{after}

3. 協調フィルタリングとの統合

上記により，成功行動を考慮すると，対象ユーザ u_a が行動 act_t を行った可能性を計算する式は最終的に式 4.10 となる．

$$P_{u_a \rightarrow act_t}^* = P_{u_a \rightarrow act_t} + DidSuccess_{u_a \rightarrow act_t} \quad (4.10)$$

図 4.7 に示すように，式 4.10 は震災時の人々の心理を反映する．震災時には，対象ユーザはまず，他の人々の行動を知りたい．そして，調べた行動の中から，成功行動を選んで行いたい．図 4.7 の協調フィルタリング部分は，“他の人々の行動を知りたい”ということを表す．図 4.7 の言語処理の部分は，“成功行動を行いたい”ということを表す．

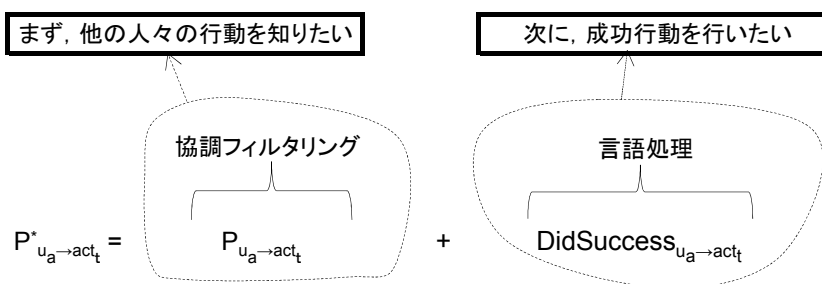


図 4.7: 震災時に人々の心理を反映する式

4.5 評価実験

評価実験を行うために、東日本大震災発生時の避難に関する Twitter ユーザ 59,563 人分のつぶやき 132,244 件を取得した。そして、1 台の PC (CPU : 2.67 Ghz , RAM : 12.0 GB) を用いて評価実験を行った。

4.5.1 ユーザ間の類似度

ユーザ間の類似度を計る際に、考慮すべき要素 (時間, 場所, 行動, ゴール) を知るために、我々は以下の手順で実験を 3 回繰り返した。

1. 行動パターンが違う 39 人のユーザを選ぶ。
2. 各ユーザに対して、行動の時刻をずらして、行動パターンが同様なユーザを生成する。
3. 各要素 (時間, 場所, 行動, ゴール) に基づき、ユーザ間の類似度を測る関数を作成する。この関数を用いて、各ユーザとの同様なユーザを探す。

表 4.3 の実験結果により、ユーザ間の類似度を測る際に、式 4.1 のように時間, 場所, 行動とゴールを考慮するべきであることが分かる。

4.5.2 欠損行動の推測

対象ユーザの欠損行動を推測する際に、以下のようなベースライン手法が考えられる。

表 4.3: ユーザ間の類似度を測る際に考慮すべき要素

考慮の要素	時間	時間・場所	時間・場所・ 行動（動作，対象）	時間・場所・行動 （動作，対象）・ゴール
同様なユーザを 把握する精度	87%	89%	92%	100%

1. 対象ユーザと最も類似しているユーザを探す。
2. 最も類似しているユーザの行動に基づき，対象ユーザの欠損行動を推測する。

ベースライン手法との比較を行うために，我々は以下の手順で実験を 10 回繰り返した。

1. 39 人のユーザをランダムに選ぶ。
2. 選んだユーザの行動をランダムに削除する。
3. 4.1 章で説明した手法を用いて，削除された行動の時刻とユーザ名を入力し，削除された行動を再現できたかを確認する。

表 4.4 の値は 10 回の実験結果の平均値である。この結果により，以下のことが分かった。

- 本手法では，最も類似しているユーザだけでなく，全ての類似ユーザの候補となる行動を考慮する。このため，ベースライン手法と比べ，本手法は欠損行動の推測精度が高い。
- 本手法では，ユーザの action（動作，対象）についての再現率は 69.23%であった。行動の場所の再現率は 76.92%であった。そして（動作，対象）と行動の場所の両方を正確に再現できたのは 43.59%であった。

表 4.4: 削除された行動の再現率

手法	（動作，対象）	場所	（動作，対象，場所）
ベースライン	31.48%	43.09%	27.56%
本手法	69.23%	76.92%	43.59%

4.6 関連研究

協調フィルタリングとは、推薦対象ユーザーと嗜好の類似した他のユーザーの情報を用いて、推薦対象ユーザーの嗜好に合うアイテムを推薦する方法論である。本論文の提案手法は他の類似ユーザの行動に基づいて対象ユーザの欠損行動を推測する手法である。従って、本論文の提案手法は協調フィルタリングの一種であると考えられる。しかし、従来の協調フィルタリングの目的は、ユーザの嗜好に合うアイテム（ECサイト内の商品など）を推薦することである。これに対して、本論文の手法の目的は任意の時刻 t にユーザの行動を把握・推測することである。アイテムと違い、震災時の行動は時間、場所、遷移関係（行動間の遷移関係）に強く依存する。そして、行動の実行時間を考慮することが必要である。表 4.5 は従来の協調フィルタリングとの比較を示す。

表 4.5: 従来の協調フィルタリングとの比較

比較観点	従来の協調フィルタリング	本論文の手法
対象	EC サイト内の商品など	ユーザの行動
対象の複雑さ	1 つの変数（アイテム）	4 つの変数（動作，対象，時刻，場所）
依存関係	場所	NO
	遷移関係	弱い
	ゴールの概念	NO
実行時間の考慮	不必要	必要
自然言語処理との統合	NO	YES

図 4.8 に示す行列は、各アイテムに対してユーザの評価値を表す行列である。実際には、全ての評価値を持つアイテムが少なく、この行列は不完全であり、未評価値は多い。協調フィルタリングには、大きく分けて、アイテムをベースする手法（アイテムベース）[10, 31, 45] とユーザをベースする手法（ユーザベース）[8, 19, 23] がある。アイテムベースとは、他のアイテムとの類似度とそのアイテムの評価値を用いて推薦対象ユーザの未評価値を推測する手法である。ユーザベースとは、他のユーザとの類似度とそのユーザの評価値を用いて、推薦対象のユーザの未評価値を推測する手法である。

協調フィルタリングの最先端手法は Ma ら [32]，Koren[26]，川前ら [52] によって行われている。式 4.11 に示すように、Ma ら [32] の手法はアイテムベースとユーザベースを統合した手法である。しかし、この手法は時間と場所を考慮していない。Koren[26] は時間を考

	i_1	i_2	i_3	i_n	アイテム
u_1					
u_2		$r_{2,2}$			
u_3	$r_{3,1}$		$r_{3,3}$		
u_n					

ユーザ

図 4.8: 各アイテムに対してユーザの評価を表す行列

慮したアイテムベースとユーザベースを統合した手法を提案した．しかし，Koren[26]の手法も場所を考慮していない．

$$P(r_{u,i}) = \alpha \left(\bar{u} + \frac{\sum_{u_a \in S(u)} Sim(u_a, u) * (r_{u_a, i} - \bar{u}_a)}{\sum_{u_a \in S(u)} Sim(u_a, u)} \right) + (1 - \alpha) \left(\bar{i} + \frac{\sum_{i_k \in S(i)} Sim(i_k, i) * (r_{u, i_k} - \bar{i}_k)}{\sum_{i_k \in S(i)} Sim(i_k, i)} \right) \quad (4.11)$$

川前ら [52] は履歴の選別にユーザごとの嗜好及びユーザ間の関係の時系列変化の両方を反映させた協調フィルタリングを提案した．図 4.9 に示すように，川前ら [52] の手法はユーザ間でのアイテム購入の先行関係を用いた協調フィルタリング手法である．そして，この手法は「各ユーザは同様な嗜好をもつ先行ユーザと同様な行動をとり、先行性は時間変化する」という仮説に基づく．先行度の計算に，ユーザの嗜好の先行性を反映するためにユーザ間の購入時刻の差を導入し，ユーザごとの嗜好の時系列性を反映するためにユーザごとの履歴の時間減衰を導入した．ユーザ u_a とユーザ u_b の間の先行度 $PID(u_a, u_b)$ を用いると， u_a がアイテム i を購入する可能性 $PP(i, u_a)$ は式 4.12 で求まる．ただし，ユーザ u_b がアイテム i を購入していれば $\delta(u_b, i) = 1$ ，そうでなければ $\delta(u_b, i) = 0$ となる．

式 4.12 に基づくと，川前ら [52] の手法は時間を考慮したユーザベース協調フィルタリン

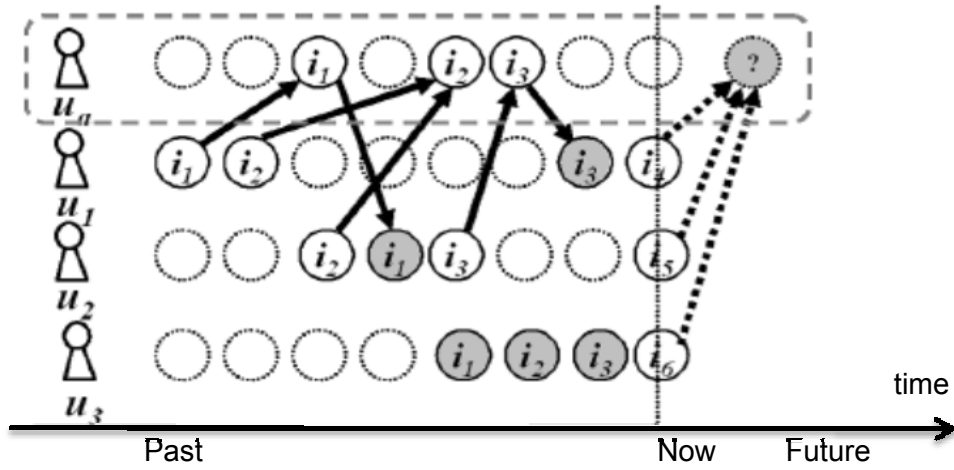


図 4.9: ユーザ間でのアイテム購入の先行関係

手法である．上記の手法と同じように，この手法も場所を考慮していない．

$$PP(i, u_a) = \sum_b \delta(u_b, i) * PID(u_a, u_b) \quad (4.12)$$

第5章 震災時の応用

現在，多くの企業や研究所で，ユーザの行動モデルや行動推測などの研究・開発が行われ，携帯端末向けのサービスなどが試行されている．例えば，NTT DocomoのMy Life Assist Service[41]では，ユーザの行き先を予測し，行き先周辺の店舗情報を提供する．KDDI研究所の「ケータイ de ライフログ」[25]では，ユーザのライフログ（いつ・どこで・誰と・何をしたか，何に興味をもったのか）を収集・管理し，適切な情報を提供することを目指している．本研究は，これらのアプリケーション等に加えて，避難・レスキューや観光案内やマーケティングなど多くの分野に適用できると考えている．以下，まず5.1節で震災時の課題について述べる．次に，5.2節で時系列行動ネットワークを用いて，避難・レスキューの応用の可能性を述べる．そして，5.3節で観光案内やマーケティングなど他の応用について説明する．最後に5.4節で震災時に実用可能なシステムのための分散処理について検討する．

5.1 震災時の課題

内閣府[53]によると，大震災が発生した場合には，以下のような問題が起こる．



①歩道は満員電車状態



②トイレは大渋滞



③避難所は混雑

図 5.1: 大震災が発生した場合の課題

- 皆がむやみに移動すると歩道が満員電車状態になる（図 5.1 の 1）。
- 情報・食料・水・トイレなどの不足が見込まれている。特に、トイレは大幅に不足する可能性がある（図 5.1 の 2）。
- 避難所はスペース不足で大混乱になる。東京都区部の避難所は、地域住民のみを対象としても 60 万人分の不足が推定されている（図 5.1 の 3）。

上記の問題を解決するために、むやみに移動を開始せず、交通機関や必需品やトイレなど避難に関する情報を把握することが必要である。また、緊急時にはこれらの情報が刻一刻と変化するため、場所と時刻に応じてこれらの情報を把握することも必要である。しかし、テレビ・ラジオから震災時に一時的に開放するトイレ・避難所や必需品の獲得は困難である。

5.2 避難・レスキューの応用の可能性

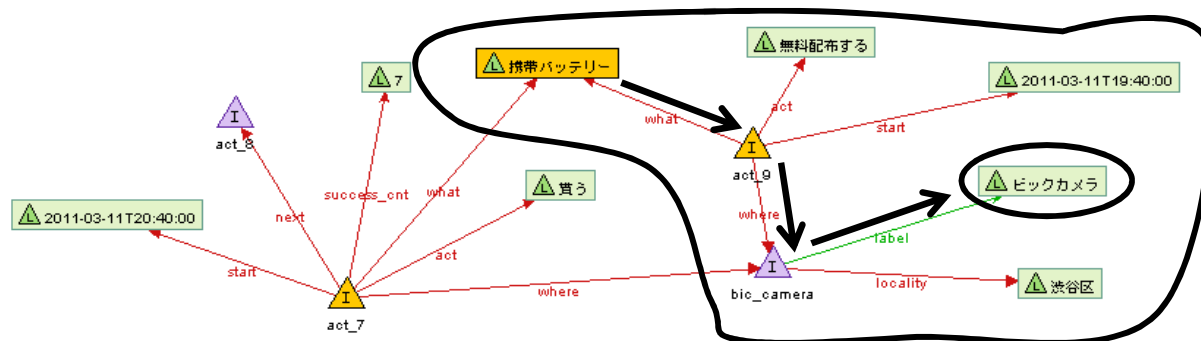


図 5.2: 時系列行動ネットワークを参照して、必需品の情報を取得する例

図 5.2 に示すように、携帯バッテリーのラベルから対象リンク（what）を辿って、行動のインスタンス（act_9）を介し場所（where）が参照されると、ビックカメラが携帯バッテリーを無料で配布していることが得られる。

また、図 5.3 に示すように、バスに乗った後の行動の場所を把握することにより、バスがどこからどこまで運行するかが得られる。

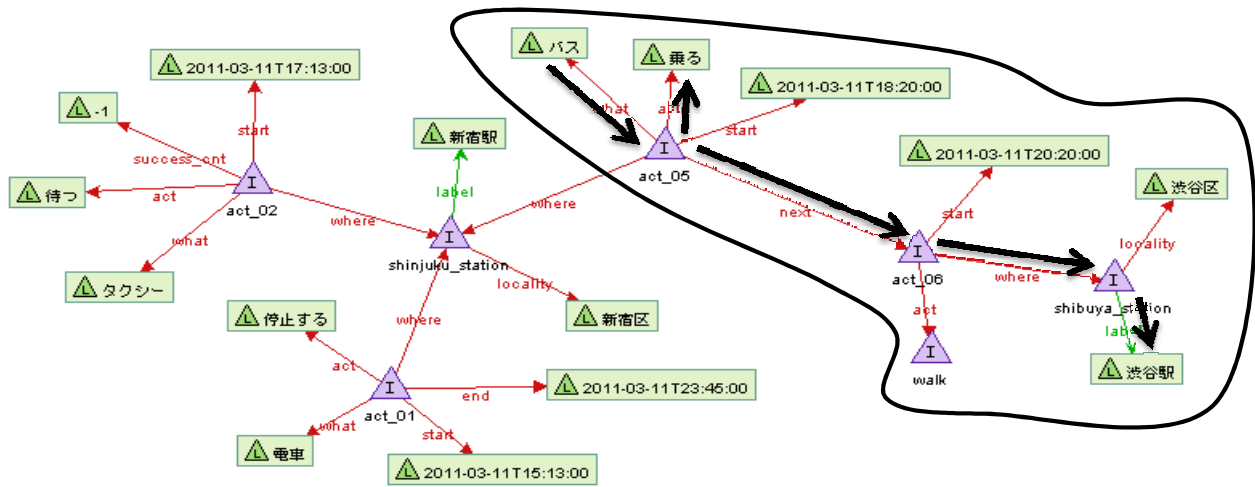


図 5.3: 時系列行動ネットワークを参照して、運行状況を取得する例

同様に時系列行動ネットワークを参照することで、トイレ・避難所の情報も得られることを確認できた。したがって、本研究の時系列行動ネットワークを利用することにより、以下の避難に関する情報を把握でき、節 5.1 で述べた課題を解決することが可能である。

- 飲み物，携帯バッテリーなど必需品
- 電車，バスなど交通機関の運行状況
- トイレや避難所の情報

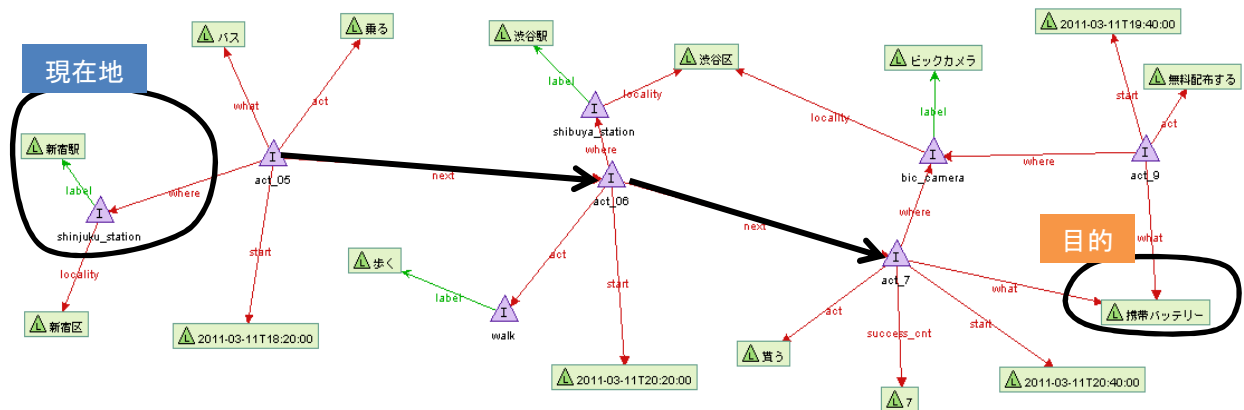


図 5.4: 目的を達成するための行動の系列

時系列行動ネットワークでは、上記の情報を場所ごとに時系列的に整理するので、場所と時刻に応じて把握可能である。更に、図5.4に示すように、時系列行動ネットワークを辿ることで、目的を達成するための行動の系列も把握可能である。

実際では、#hinan タグを含むつぶやきの内、避難情報を含まないつぶやきは数多く存在する。また、#hinan タグを含むつぶやきの文書中に現れる行動の内、避難情報に関係ない行動も存在する。そして、本論文の行動抽出手法の再現率と適合率はそれぞれ66.54%、73.21%である。したがって、時系列行動ネットワークを構成する行動の中には誤った行動又は避難情報に関係ない行動が存在する。時系列行動ネットワークの有用と思われる行動（避難情報を表す行動）の割合を把握するために、我々は以下の評価実験を行った。

図5.5に示すように、我々は以下のような記号を用いて、#hinan タグを含むつぶやきの質を説明する。

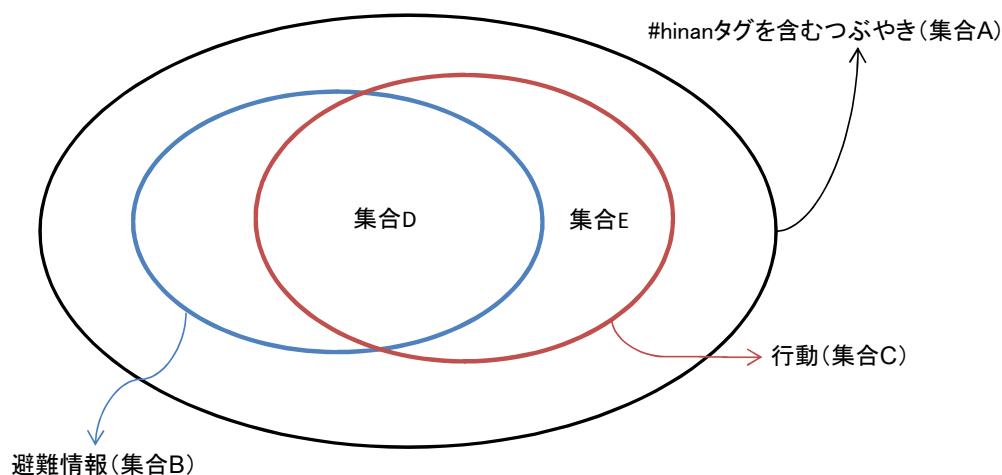


図 5.5: 集合 A, 集合 B, 集合 C, 集合 D, 集合 E の関係

- 集合 A は#hinan タグを含むつぶやきの集合である。
- 集合 B は集合 A のつぶやきの文書中に現れる避難情報の集合である。
- 集合 C は集合 A のつぶやきの文書中に現れる行動の集合である。
- 集合 B と集合 C の共通部分は集合 D である。つまり、集合 D の行動は避難情報を表す行動なので、避難において有用な行動であると考えられる。

- 集合 E は集合 C と集合 D の差集合である (集合 E = 集合 C - 集合 D) . つまり, 集合 E は避難において有効ではない行動の集合である .

表 5.1 に示すように, 我々は #hinan タグのつぶやき 132,244 件の中から 4 つのつぶやきセットをランダムに選び, 評価実験を行った . 評価実験の結果, 集合 C では, 有用な行動の平均割合は 76.73% である .

表 5.1: #hinan タグのつぶやきの質

No	集合 A の要素の個数	集合 B の要素の個数	集合 C の要素の個数	集合 D の要素の個数	集合 E の要素の個数	有用な行動の割合
1	64	23	18	16	2	$16/18 = 88.89\%$
2	68	24	14	11	3	$11/14 = 78.57\%$
3	86	10	8	4	4	$4/8 = 50.00\%$
4	39	18	19	17	2	$17/19 = 89.47\%$
平均	64	19	15	12	3	76.73%

表 5.2: 時系列行動ネットワークを構成する行動の質

No	集合 C の要素の個数	正しく抽出された行動の数	抽出された行動の数	有用な行動数の上限 (ベストケース)	有用な行動数の下限 (ワーストケース)
1	18	$18 \times 66.54\% = 12$	$12/73.21\% = 16$	12	10
2	14	$14 \times 66.54\% = 9$	$9/73.21\% = 12$	9	6
3	8	$8 \times 66.54\% = 5$	$5/73.21\% = 9$	4	1
4	19	$19 \times 66.54\% = 13$	$13/73.21\% = 18$	13	11
平均	15	10	14	10	7

次に, 集合 C の行動を表すつぶやきを対象に, 行動を抽出し, 時系列行動ネットワークを構成することを考える . 表 5.2 に示すように, 本論文の行動を抽出する手法の適合率と再現率に基づく, 集合 C の中から正しく抽出された行動の数と抽出された行動の数を推測することができる . そして, 正しく抽出された行動と有用な行動 (集合 D の行動) に基づく, 時系列行動ネットワークの有用な行動の割合を推測することができる . 正しく抽出された行動は集合 D (有用な行動の集合) 又は集合 E (有用ではない行動の集合) のい

ずれかに含まれる。したがって、時系列行動ネットワークの有用な行動の割合は以下の2つのケースがあると考えられる。

- ベストケース：時系列行動ネットワークを構成する行動が、主として有用な行動の集合（集合D）から抽出された場合。ベストケースにおいて、時系列行動ネットワークの有用な行動の数は、集合Dの要素の数と正しく抽出された行動の数のうち小さい方の値である。表5.2に示すように、ベストケースにおける、時系列行動ネットワークの有用な行動の割合は約71%（10個の有用な行動/14個の抽出された行動）である。
- ワorstケース：時系列行動ネットワークを構成する行動が、主として有用でない行動の集合（集合E）から優先的に抽出された場合。正しく抽出された行動の数が集合Eの要素の数より大きければ、ワorstケースにおいて時系列行動ネットワークの有用な行動の数は、正しく抽出された行動の数から、集合Eの要素の数を引いた値である。逆に、正しく抽出された行動の数が集合Eの要素の数より小さいければ、ワorstケースにおいて時系列行動ネットワークの有用な行動の数は0である。表5.2に示すように、ワorstケースにおける、時系列行動ネットワークの有用な行動の割合は約50%（7個の有用な行動/14個の抽出された行動）である。

したがって、時系列行動ネットワークを構成する行動の内、有用な行動の割合は約50%～71%までであると考えられる。そして、時系列行動ネットワークは#hinan タグを含むつぶやきの文書中に現れる約58%（7個の有用な行動/12個の有用な行動）～83%（10個の有用な行動/12個の有用な行動）までの避難に関する行動を正確に再現している。

上記の時系列行動ネットワークの特性を利用して、伊藤ら [51] は避難行動推薦システムを構築した。これはユーザの持つ携帯端末により現在地を取得することで、ユーザの状況（現在地、時刻）に応じた避難行動を推薦するシステムである。図5.6はAndroid上に表示された推薦画面の例であり、現在地から付近の避難所までの行動を推薦している。緑のマーカはユーザの現在地であり、赤のマーカが避難所である。

通常、時系列行動ネットワークのノード数は時間とともに単調増加し、推薦に必要な処理時間は時間と共に増加する。しかし、必要な情報に絞るなどの工夫をすることで、現実的な処理時間で行動推薦が行えると考えられる。例えば、現在時刻から一定の時間内の行



図 5.6: ユーザの状況に応じた避難行動を推薦

動(ノード), 特定の場所に関する行動(ノード)などに絞ることが挙げられる。実際に伊藤ら [51] では現在地付近の行動(ノード)だけを対象にすることで, 処理時間は時間経過によらず一定であることを示している。表 5.3 に示すように, 周辺行動の提示の処理時間は平均で 144 ミリ秒のため, リアルタイムな推薦に支障はないと考えられる。

表 5.3: 周辺行動の提示の処理時間

時刻	行動ノード数	処理時間 (ms)
15:00	4	125
16:00	26	129
17:00	53	134
18:00	80	145
19:00	117	150
20:00	129	153
21:00	123	149
22:00	113	150
23:00	88	150
24:00	77	155

5.3 その他の応用の検討

本研究は，上記に説明した震災だけでなく，実世界の行動分析，観光案内，マーケティング，経験共有など多くの分野にも適用できる．例えば，Nguyenら [39] は本論文の行動属性の抽出手法を用いて，秋葉原に関する購入行動を分析した．以下に，観光案内，マーケティングと経験共有に応用する方法を説明する．

- 行動ベース観光案内

本論文の手法はドメインに依存しないので，震災だけでなく，実世界の観光に関する行動を表すネットワークも構築できる．そして，この観光行動ネットワークをカーナビゲーションや，電車，Suica，Pasmo，携帯電話などに組み込んで，観光客の行き先周辺の人気がある行動パターンを推薦する．例えば，weblogs や Twitter 等から秋葉原に関する行動を抽出して，図 5.7 のような行動ネットワークを構築できる．この行動ネットワークを参照することで，皆が秋葉原で“ どの店に入る ”，“ どの店で食べる ”，“ 何を買う ”，“ 買い物してからどこに遊ぶ ”など秋葉原に関する行動パターンを把握できる．秋葉原の観光客（初心者，外国人）を街案内するために，これらの行動パターンは重要な情報になる．

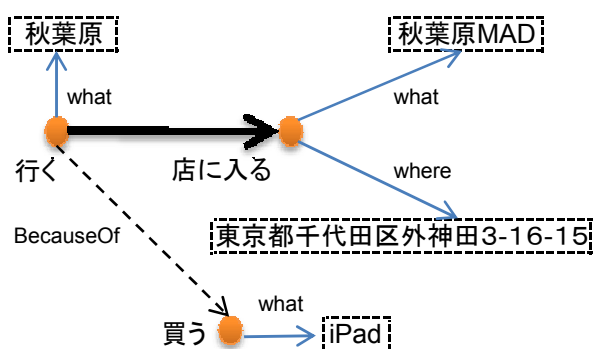


図 5.7: 秋葉原に関する行動を表す行動ネットワーク（一部）

- 行動ベースマーケティング

ブログ・Twitter のつぶやき情報から，図 5.8 に示すように，ユーザが行いたい行動や行おうとする行動を把握できる．行動ベースマーケティングを実現するために，こ

これらの行動は重要な情報になる．例えば，ユーザが“ iPad を買いたい ”ということが分かれば，iPad に関する情報（割引，周辺の販売店）を推薦すると良いマーケティング効果が出ると考えられる．

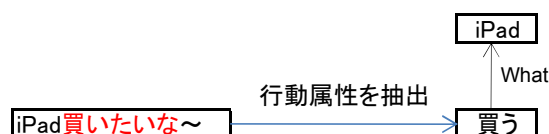


図 5.8: つぶやきに現れるユーザの購入行動

- 経験共有

図 5.9 に示すように，本研究は経験共有に適用できる．具体的には，まずブログの文書等から皆の失敗・成功行動を収集し，これらの行動から有用な行動を初心者にアドバイスすることなどができよう．

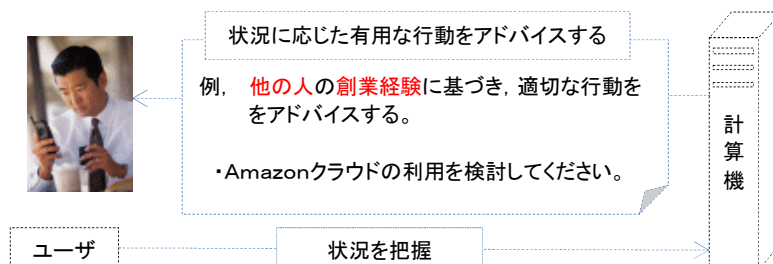


図 5.9: 経験共有サービス

5.4 震災時に実用可能なシステムのための分散処理の検討

震災時の行動推薦システムを考えると，リアルタイム性が重要で，実時間処理可が可能なシステムが必要である．つまり，全てのつぶやきをリアルタイムで処理し，時系列行動ネットワークに新たな行動データを追加するべきである．

東日本大震災当日の 3 月 11 日は Twitter での 1 日の平均総つぶやき数約 1,800 万件に対し，約 3,300 万件とつぶやき件数が 1.8 倍に飛躍的に増加した [7]．これらの大量つぶやきをリアルタイムで処理するために，処理を分散するべきだと考えられる．

図 5.10 に示すように，時系列行動ネットワークの行動データが RDF データであるため，分散しやすい．つまり，行動データを増加したならば，新たな保管デスクを追加すれば問題がない．

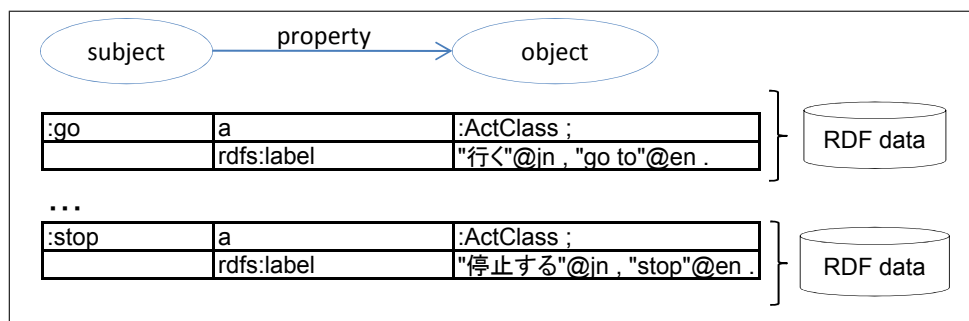


図 5.10: 時系列行動ネットワークのデータ

そして，図 5.11 に示すように，MapReduce の考え方をを用いると，大量のつぶやきをリアルタイムで処理可能である．具体的には，それぞれの計算機は一部のユーザのつぶやきを担当し，つぶやきの文書中に現れる行動属性を抽出する．そして，Reduce の課程で，Map で抽出された行動データをソートしてから，RDF/R3 に変換し時系列行動ネットワークに追加する．

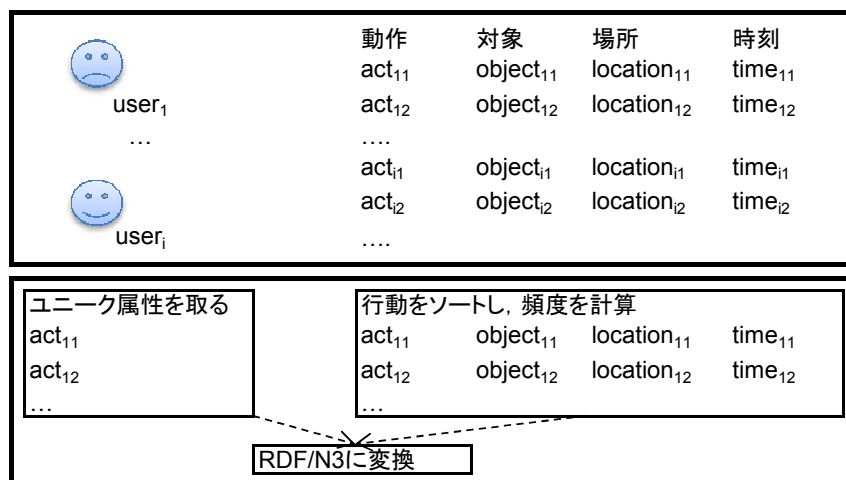


図 5.11: MapReduce による大量のつぶやきの処理

第6章 結論

本章では、まず各提案手法の解決課題、課題の解決方法、オリジナリティ、長所・短所をまとめる。次に、今後の課題について述べる。

6.1 まとめ

本論文は、ユーザの状況に応じた行動を推薦するシステムの実現を目的として、行動の集合知である時系列行動ネットワークを構築する手法を提案した。これは以下の3つの手法から構成される。

1. Twitterのようなソーシャルメディアから人間の行動を自動抽出する手法
2. OWL (Web Ontology Language) で表現される時系列行動ネットワークを自動構築する手法
3. 欠損行動(つぶやかれなかった行動)を推測して時系列行動ネットワークを補完する手法

また、第5章では、これらの提案手法が避難・レスキューや観光案内やマーケティングなど様々な応用分野に適用できることを示した。以下に、各提案手法の解決課題、課題の解決方法、オリジナリティ、長所・短所をまとめる。

6.1.1 行動の自動抽出手法のまとめ

- 課題

Twitterのようなソーシャルメディアからの行動の抽出については、以下のような課題がある。

1. つぶやきの数が大量であるため，人手で抽出すると膨大なコストがかかる．また，緊急時にこの作業をリアルタイムで行うことは困難である．
2. 顔文字や“ えーっと ”，“ 。。。 ”などのようノイズ文字列が多く，他のテキストメディアより複雑かつ文法的に正しくない文が多い．
3. 頻度に基づいた手法では，マイノリティ行動を獲得できない．
4. 抽出手法が行動のドメインに依存すると，新たなドメインに適用するために，新たな訓練データの作成は必要になる．

- 課題の解決方法

上記の課題に対して，提案手法は以下の解決方法を行う．

1. 自己教師あり学習を適用し，行動を自動的に抽出する．
2. 前処理により，文中にあるノイズ文字列を削除している．そして，複雑かつ正しい文法で記述されていない文に対応するために，文の単純化と訓練データの追加を行っている．
3. 行動の頻度ではなく，文の構成と単語の品詞に基づいて行動属性のラベルを推測する．
4. 行動のドメインに依存しないヒューリスティック（構文パターン）を作成する．

- 提案手法のオリジナリティ

提案手法は英語の Web コーパスからの関係抽出の最先端技術である O-CRF[5] をベースとした手法である．しかし，提案手法は O-CRF と違い，一つのラベルだけではなく，文に現れるすべての行動属性を推定する必要がある．また，日本語の文を対象とするので，文によって属性の数と位置は変わる．このため，提案手法では，エンティティの事前判定，又はリレーションがエンティティの間に現れるといった O-CRF の設定ができない．つまり，O-CRF の解決課題よりも本手法が解決する課題の方が困難であると考えられる．

提案手法のオリジナリティと工夫点は以下の通りである．

- 複雑な文を対応するために、訓練データの追加方法を提案した。
- 行動属性抽出の分野において、提案手法は初めて条件付確率場と自己教師あり学習を利用している。
- 時間、場所、人名の抽出や複雑な文の単純化など様々な工夫を実施した。
- データの特徴を表すためのテンプレートファイルを作成した。

- 提案手法の長所

提案手法の主な長所は以下の通りである。

- 訓練データを自動的に作成することで、ラベル付けのコストがかからない。
- ドメイン（行動の種類）に依存せず、汎用性が高い手法である。
- 行動の頻度に依存せずに、マイノリティ行動も抽出できる。
- Web上の文書だけでなく、Twitterのようなソーシャルメディアの文書にも対応できる。
- 緊急時に大量なつぶやきの文書を迅速に処理可能である。
- 本手法は日本語を対象としたが、英語や他の言語にも適用可能である。

- 提案手法の短所

しかし、現時点では、提案手法は以下のような短所がある

- 文の長さや複雑さに依存する。
- 暗黙的な行動を抽出できていない。例えば、“俺は先週渋谷にある回転寿司店へ行った。すごく美味しかったよ!”の文では、“寿司を食べた”という行動が隠れており、抽出できない。今後、これらの課題を解決するために、推論や相関ルールの適用などを検討している。

我々は行動属性の自動抽出手法を以下の学術論文でまとめて発表した。

- The-Minh Nguyen, Takahiro Kawamura, Yasuyuki Tahara, and Akihiko Ohsuga: Self-Supervised Capturing of Users' Activities from Weblogs. International Journal

of Intelligent Information and Database Systems, Vol.6, No.1, pp.61-76, InderScience Publishers, 2012.

- ゲン ミン テイ, 川村隆浩, 中川博之, 田原康之, 大須賀昭彦: 条件付確率場と自己教師あり学習を用いた行動属性の自動抽出と評価. 人工知能学会論文誌, pp.166-178, 26 巻 1 号 SP-R, 2011.
- Nguyen Minh The, Takahiro Kawamura, Hiroyuki Nakagawa, Yasuyuki Tahara, Akihiko Ohsuga: Self-Supervised Mining of Human Activity from CGM. The 11th International Workshop on Knowledge Management and Acquisition for Smart Systems and Services (PKAW 2010), LNAI 6232, pp.61-72, Springer, Heidelberg. Daegu, Korea, 30 August - 3 September, 2010.
- Nguyen Minh The, Takahiro Kawamura, Hiroyuki Nakagawa, Yasuyuki Tahara, Akihiko Ohsuga: Automatic Mining of Human Activity Attributes from Weblogs. In Proc. IEEE/ACIS ICIS2010, Semantic, Social, and Collaborative Intelligence Session, pp.633-644, Yamagata, Japan, 18-20 August, 2010.
- Nguyen Minh The, Takahiro Kawamura, Hiroyuki Nakagawa, Yasuyuki Tahara, Akihiko Ohsuga: Automatic Mining of Human Activity and Its Relationships from CGM. In Proc. ICSOFT2010, volume 1, pp.285-292, Athens, Greece, 22-24 July, 2010.
- Nguyen Minh The, Takahiro Kawamura, Hiroyuki Nakagawa, Ken Nakayama, Yasuyuki Tahara, Akihiko Ohsuga: Human Activity Mining using Conditional Random Fields and Self-Supervised Learning. In Proc. ACIIDS2010, part I, LNAI 5990, pp.140-149. Springer, Heidelberg. Hue City, Vietnam, 24-26 March, 2010.
- ゲン ミン テイ, 川村隆浩, 田原康之, 大須賀昭彦: Twitter から人間行動属性の自動抽出. IEICE・人工知能と知識処理研究会 (Artificial Intelligence and Knowledge-Base Processing), 一般講演 (理論), pp.19-23. 関学東京キャンパス, 2010年6月18日.

- グエン ミン テイ，川村隆浩，中川博之，田原康之，大須賀昭彦：CGMからの自己教師あり学習と条件付確率場を用いた人間行動マイニング．人工知能学会（第24回）全国大会（JSAI2010），2010年6月9～11日．
- グエン ミン テイ，川村隆浩，中川博之，田原康之，大須賀昭彦：Webからの自己教師あり学習を用いた人間行動マイニング．IEICE・人工知能と知識処理研究会，Vol.109，No.386，pp.19-24．「グリーン AI」および一般．関学東京キャンパス，2010年1月15日．
- グエン ミン テイ，川村隆浩，中川博之，中山健，田原康之，大須賀昭彦：条件付確率場と自己教師あり学習を用いた行動属性の自動抽出と評価．合同エージェントワークショップ&シンポジウム2009，pp. 468-475．主催：マルチエージェントと協調計算研究会（日本ソフトウェア科学会），人工知能と知識処理研究会（電子情報通信学会），知能と複雑系研究会（情報処理学会），知識ベースシステム研究会（人工知能学会）．後援：IEEE Computer Society Japan Chapter，2009年10月28～30日．

6.1.2 時系列行動ネットワークの構築手法のまとめ

- 課題
行動を推薦するためには，推薦すべき行動の集合知（行動の構成要素，行動間の関係）が必要である．この集合知を構築する際には以下の課題を解決する必要がある．
 1. 計算機が行動を自動的に推薦するためには，これらの行動の表現，参照の形式化が必要である．さらに，緊急時であることを考慮すると時間情報が重要となり，各行動の時刻を明確に記述し，時系列的に整理することが必要である．
 2. 緊急時における行動の推薦を考えるとリアルタイム性が重要となる．しかし，緊急時にリアルタイムで行動の集合知を人手で構築することは難しく，膨大なコストがかかる．
- 課題の解決方法
我々は上記の課題を以下のように解決した．

1. 行動ネットワークに時間情報に加えて、時系列行動ネットワークを定義した。そして、Web オントロジー言語 (OWL) に基づいて定義された時系列行動ネットワークを設計した。
2. 行動属性の自動抽出手法を用いて、Twitter 上の情報から行動を取得し、時系列行動ネットワークに追加した。

- 提案手法のオリジナリティ

提案手法のオリジナリティは以下の通りである。

- OWL に従う時系列行動ネットワークである。この時系列行動ネットワークは行動のインスタンスまで表し、時間の流れで動的に変化する集合知である。
- Twitter 上の情報から行動を自動的に取得し、時系列行動ネットワークを構成することである。

- 提案手法の長所

提案手法の長所は以下の通りである。

- 各時刻における行動と行動間の関係を表すことができる。
- 緊急時にリアルタイムで行動の集合知を自動的に構築できる。

- 提案手法の短所

しかし、提案手法には以下のような短所がある。

- 構築の精度は行動属性の自動抽出手法に依存する。

我々は時系列行動ネットワークの構築手法を以下の学術論文でまとめて発表した。

- **The-Minh Nguyen**, Kenji Koshikawa, Takahiro Kawamura, Yasuyuki Tahara, and Akihiko Ohsuga: Building Earthquake Semantic Network by Mining Human Activity from Twitter. 2011 IEEE International Conference on Granular Computing, pp.496-501. Kaohsiung, Taiwan, 8-10 November, 2011.

- ゲン ミン テイ，越川 兼地，川村隆浩，田原康之，大須賀昭彦：震災時の効率的な避難のための行動推薦エージェント Ready for 87%の提案（2）～時系列避難行動ネットワークの構築～．合同エージェントワークショップ&シンポジウム 2011．主催：マルチエージェントと協調計算研究会（日本ソフトウェア科学会），人工知能と知識処理研究会（電子情報通信学会），知能システム研究会（情報処理学会），知識ベースシステム研究会（人工知能学会），IEEE Computer Society Japan Chapter, 2011年10月26～28日．
- ゲン ミン テイ，伊藤雅博，川村隆浩，中川博之，田原康之，大須賀昭彦：Webからの実世界行動抽出による意味ネットワーク構築手法の提案．IEICE・人工知能と知識処理研究会，vol.111, no.70, AI2011-8, pp.39-44, 2011年5月26日．

6.1.3 時系列行動ネットワークの補完手法のまとめ

- 課題

Twitterなどのソーシャルメディアはユーザの自主性に依存しているため，つぶやかれなかった行動も数多く存在すると考えられる．また，リアルタイムにユーザの行動に応じた情報を推薦するためには，これらの欠損行動を推測する必要がある．欠損行動を推測する際には以下の課題がある．

1. 以下の性質を考慮する必要がある．
 - － 震災時におけるユーザの行動は場所と時刻に強く依存する．
 - － 行動は動作，対象，時刻，場所という4つの属性を持つ．
 - － 現在の行動は前後の行動と依存関係がある．
2. ユーザ間の類似度を測る際には，上記の行動の性質に加えて，ユーザのゴール（避難したい，帰宅したいなど）を考慮するべきである．
3. つぶやかなかった時間に対象ユーザは複数の行動を行った可能性がある．
4. マイノリティであっても成功した行動はきちんと考慮すべきである．

- 課題の解決方法

我々は上記の課題を以下のように解決した。

1. 行動の性質を考慮した行動ベース協調フィルタリング手法を提案した。
2. ゴール概念をユーザ間の類似度の式に導入した。
3. 行動の実行時間を考慮し、各時刻における行動の重みを付けた。
4. 協調フィルタリングと言語処理を統合し、成功行動を考慮した。

- 提案手法のオリジナリティ

提案手法のオリジナリティは以下の通りである。

- 行動の性質を考慮した行動ベース協調フィルタリング手法。
- 協調フィルタリングと言語処理の統合。

- 提案手法の長所

提案手法の長所は以下の通りである。

- 実世界行動の性質を考慮した。
- 成功行動を考慮した。
- 任意の時刻に対して、行動を把握・推測できる。

しかし、本手法では、過去の欠損行動の推測を行っているが、今後の課題としては将来行う可能性がある行動の推測も検討すべきだと考えられる。

我々は時系列行動ネットワークの補完手法を以下の学術論文でまとめて発表した。

- The-Minh Nguyen, Takahiro Kawamura, Yasuyuki Tahara, and Akihiko Ohsuga: Building a Time Series Action Network for Earthquake Disaster. 4th International Conference on Agents and Artificial Intelligence (ICAART2012), volume 1, pp.100-108, Vilamoura, Algarve, Portugal, 6-8 February, 2012.

- グエン ミン テイ , 越川 兼地 , 川村隆浩 , 田原康之 , 大須賀昭彦 : 震災時の効率的な避難のための行動推薦エージェント Ready for 87%の提案 (2) ~時系列避難行動ネットワークの構築~ . 合同エージェントワークショップ&シンポジウム 2011 . 主催 : マルチエージェントと協調計算研究会 (日本ソフトウェア科学会) , 人工知能と知識処理研究会 (電子情報通信学会) , 知能システム研究会 (情報処理学会) , 知識ベースシステム研究会 (人工知能学会) , IEEE Computer Society Japan Chapter, 2011 年 10 月 26 ~ 28 日 .

6.1.4 震災および他の応用のまとめ

東日本大震災発生時のつぶやき 132,244 件を対象に評価実験を行った結果, 提案手法が避難情報を迅速に提供するのに有効であることが確かめられた. なお, 本研究成果は震災時だけでなく, 様々な自然災害 (台風, 大雪), 人的災害 (ストライキ, テロ) 時の行動把握, 更に観光案内, マーケティングなど多くの分野に適用可能であると考えられる.

我々は震災および他の応用を以下の学術論文でまとめて発表した.

- Isabel Shizu Miyamae Iwanaga, The-Minh Nguyen, Takahiro Kawamura, Hiroyuki Nakagawa, Yasuyuki Tahara, and Akihiko Ohsuga: Building an Earthquake Evacuation Ontology from Twitter. 2011 IEEE International Conference on Granular Computing, pp.306-311. Kaohsiung, Taiwan, 8-10 November, 2011.
- グエン ミン テイ , 越川 兼地 , 川村隆浩 , 田原康之 , 大須賀昭彦 : 震災時の効率的な避難のための行動推薦エージェント Ready for 87%の提案 (2) ~時系列避難行動ネットワークの構築~ . 合同エージェントワークショップ&シンポジウム 2011 . 主催 : マルチエージェントと協調計算研究会 (日本ソフトウェア科学会) , 人工知能と知識処理研究会 (電子情報通信学会) , 知能システム研究会 (情報処理学会) , 知識ベースシステム研究会 (人工知能学会) , IEEE Computer Society Japan Chapter, 2011 年 10 月 26 ~ 28 日 .
- ミヤマエ イワナガ イサベル シズ , グエン ミン テイ , 川村隆浩 , 中川博之 , 田原康之 , 大須賀昭彦 : 震災時の効率的な避難のための行動推薦エージェント Ready for

87%の提案(1)～避難計画オントロジーの設計～. 合同エージェントワークショップ & シンポジウム 2011. 主催: マルチエージェントと協調計算研究会 (日本ソフトウェア科学会), 人工知能と知識処理研究会 (電子情報通信学会), 知能システム研究会 (情報処理学会), 知識ベースシステム研究会 (人工知能学会), IEEE Computer Society Japan Chapter, 2011年10月26～28日.

- 伊藤雅博, グエン ミン テイ, 川村隆浩, 中川博之, 田原康之, 大須賀昭彦: 震災時の効率的な避難のための行動推薦エージェント Ready for 87%の提案(3)～行動ネットワークを用いた避難行動の推薦～. 合同エージェントワークショップ & シンポジウム 2011. 主催: マルチエージェントと協調計算研究会 (日本ソフトウェア科学会), 人工知能と知識処理研究会 (電子情報通信学会), 知能システム研究会 (情報処理学会), 知識ベースシステム研究会 (人工知能学会), IEEE Computer Society Japan Chapter, 2011年10月26～28日.
- Nguyen Minh The, Takahiro Kawamura, Yasuyuki Tahara, Akihiko Ohsuga: Capturing Users' Buying Activity at Akihabara Electric Town from Twitter. In Proc. ICCCI2010, part II, LNAI 6422, pp.163-171. Springer, Heidelberg, Kaohsiung, Taiwan, November 10-12, 2010.

6.2 今後の課題

今後の課題として以下のようなタスクが挙げられる.

- つぶやきの前処理を強化し, 行動を抽出する手法の精度を改善する.
- 欠損行動の推測手法と時系列行動ネットワークを用いてユーザの将来行動を予測する.
- つぶやきの信頼性を検討する.

そして, 伊藤ら [51] のように, 本研究の時系列行動ネットワークを用いて, 震災や観光案内などにおける, ユーザの状況に応じた行動を推薦することも重要な課題である.

謝辞

本研究にあたり，ご多忙の中適切なご指導を下さった大須賀 昭彦 教授，川村 隆浩 客員准教授，田原 康之准教授に感謝いたします．また，日頃から実験や日本語のチェックなど様々な助言・協力をして下さった大須賀・田原・川村研究室の皆様感謝の意を表します．そして，投稿した論文等に対して，国内外の多くの査読者から様々なコメントをいただきました．

参考文献

- [1] Aduna. A de-facto standard framework for processing rdf data. <http://www.openrdf.org/about.jsp>, 2011.
- [2] E. Agichtein and L. Gravano. Snowball: Extracting relations from large plain-text collections. In *Proc. ACM DL*, 2000.
- [3] G. A.Miller and C. Fellbaum. Wordnet. <http://wordnet.princeton.edu/>.
- [4] M. Banko. Open information extraction for the web. PhD thesis, University of Washington, 2009.
- [5] M. Banko and O. Etzioni. The tradeoffs between open and traditional relation extraction. In *Proc. ACL*, 2008.
- [6] M. Banko, M. J.Cafarella, S. Soderland, M. Broadhead, and O. Etzioni. Open information extraction from the web. In *Proc. IJCAI*, pp. 2670–2676, 2007.
- [7] Biglobe, 2011. <http://tr.twipple.jp/info/bunseki/20110427.html>.
- [8] J.S. Breese, D. Heckerman, and C. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In *Proc. UAI*, 1998.
- [9] S. Brin. Extracting patterns and relations from the world wide web. In *Proc. EDBT*, pp. 172–183, 1998.
- [10] M. Deshpande and G. Karypis. Item-based top-n recommendation. *ACM Trans. Inf. Syst.*, Vol. 22, No. 1, pp. 143–177, 2004.

- [11] eHow. <http://www.ehow.com/>.
- [12] O. Etzioni, M. Cafarella, D. Downey, A. Popescu, T. Shaked, S. Soderland, D. S.Weld, and A. Yates. Methods for domain-independent information extraction from the web: An experimental comparison. In *Proc. AAAI*, 2004.
- [13] G.D. Forney. The viterbi algorithm. In *Proc. IEEE*, 第 61 卷, 1973.
- [14] T. Fuchi and S. Takagi. Japanese morphological analyzer using word co-occurrence-jtag. In *Proc. ACL*, pp. 409–413, 1998.
- [15] Y. Fukazawa and J. Ota. Learning user’s real world activity model from the web. In *IEICE SIG Notes*, 2009.
- [16] Geo, 2003. http://www.w3.org/2003/01/geo/wgs84_pos.
- [17] Google. Google Maps API Services.
- [18] H. Halpin, R. Iannella, B. Suda, and N. Walsh, 2010. <http://www.w3.org/2006/vcard/ns>.
- [19] J.L. Herlocker, J.A. Konstan, A. Borchers, and J. Riedl. An algorithmic framework for performing collaborative filtering. In *Proc. SIGIR*, 1999.
- [20] L. Hugo and S. Push. a practical commonsense reasoning toolkit. *BT Technology Journal*, Vol. 22, No. 4, 2004.
- [21] C. J.C.Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 1998.
- [22] D. Jeffrey and G. Sanjay. Mapreduce: Simplified data processing on large clusters. In *Proc. OSDI*, 2004.
- [23] R. Jin, J.Y. Chai, and L. Si. An automatic weighting scheme for collaborative filtering. In *Proc. SIGIR*, 2004.

- [24] T. Kawamura, T-M. Nguyen, and A. Ohsuga. Building of human activity correlation map from weblogs. In *Proc. ICSOFT*, 2009.
- [25] KDDI, 2009. Mobile phone based Lifelog.
- [26] Y. Koren. Collaborative filtering with temporal dynamics. In *Proc. KDD*, 2009.
- [27] T. Kudo and Y. Matsumoto. Japanese dependency analysis using cascaded chunking. In *Proc. CoNLL*, pp. 63–69, 2002.
- [28] T. Kudo, K. Yamamoto, and Y. Matsumoto. Applying conditional random fields to japanese morphologicaical analysis. In *Proc. EMNLP*, pp. 230–237, 2004.
- [29] T. Kurashima, K. Fujimura, and H. Okuda. Discovering association rules on experiences from large-scale weblogs entries. In *Proc. ECIR. LNCS vol 5478. Springer*, 2009.
- [30] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proc. ICML*, 2001.
- [31] G. Linden, B. Smith, and J. York. Amazon.com recommendations: Item-to-item collaborative filtering. In *IEEE Internet Computing*, pp. 76–80, 2003.
- [32] H. Ma, I. King, and M. R-Lyu. Effective missing data prediction for collaborative filtering. In *Proc. SIGIR*, 2007.
- [33] Y. Matsuo, N. Okazaki, K. Izumi, Y. Nakamura, T. Nishimura, and K. Hasida. Inferring long-term user properties based on users’ location history. In *Proc. IJCAI*, pp. 2159–2165, 2007.
- [34] A. McCallum and W. Li. Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons. 2003.
- [35] MIT, 2011. <http://openmind.media.mit.edu/>.

- [36] Thomas Mitchell. *Machine Learning*. McGraw-Hill ISBN 978-0071154673, 1997.
- [37] N. Mor, B. Jeffrey, and L. Chih-Hui. Is it really about me. message content in social awareness streams. In *Proc. CSC*, 2010.
- [38] I. Nakabayashi. Development of urban disaster prevention systems in japan - from the mid-1980s. *Journal of Disaster Research*, Vol. 1, , 2006.
- [39] T-M. Nguyen, T. Kawamura, Y. Tahara, and A. Ohsuga. Capturing users' buying activity at akihabara electric town from twitter. In *Proc. ICCCI, part II, LNAI 6422*, pp.163-171. Springer, Heidelberg, 2011.
- [40] B. Nilanjan, C. Dipanjan, D. Koustuv, J. Anupam, M. Sumit, N. Seema, R. Angshu, and M. Sameer. User interests in social media sites: An exploration with micro-blogs. In *Proc. CIKM*, 2009.
- [41] NTTDocomo, 2009. My Life Assist Service.
- [42] M. Perkowski, M. Philipose, and K. Fishkin J. Donald. Mining models of human activities from the web. In *Proc. WWW*, 2004.
- [43] S. Poslad. *Ubiquitous Computing Smart Devices, Environments and Interactions*. Wiley, ISBN: 978-0-470-03560-3, 2009.
- [44] Y. Raimond and S. Abdallah, 2007. <http://purl.org/NET/c4dm/timeline.owl>.
- [45] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *Proc. WWW*, 2001.
- [46] F. Sha and F. Pereira. Shallow parsing with conditional random fields. In *Proc. HLTNAACL*, pp. 213–220, 2003.
- [47] W3C, 2004. <http://www.w3.org/TR/rdf-concepts/>.
- [48] W3C, 2006. <http://www.w3.org/DesignIssues/Notation3>.

- [49] W3C, 2008. <http://www.w3.org/TR/rdf-sparql-query/>.
- [50] ミヤマエ イワナガ イサベルシズ, ゲン ミンテイ, 川村隆浩, 中川博之, 田原康之, 大須賀昭彦. 震災時の効率的な避難のための行動推薦エージェント ready for 87%の提案 (1) ~ 避難計画オントロジーの設計 ~. In *JAWS2011*, 2011.
- [51] 伊藤雅博, ゲン ミンテイ, 川村隆浩, 中川博之, 田原康之, 大須賀昭彦. 震災時の効率的な避難のための行動推薦エージェント ready for 87%の提案 (3) ~ 行動ネットワークを用いた避難行動の推薦 ~. In *JAWS2011*, 2011.
- [52] 川前徳章, 坂野鋭, 山田武士, 上田修功. ユーザの嗜好の時系列性と先行性に着目した協調フィルタリング. 電子情報通信学会論文誌 D, Vol. 92, No. 6, 2009.
- [53] 内閣府 (防災担当). 震災時の帰宅行動, そのときあなたはどのようにする?, 2010.

付録 A 時系列行動ネットワーク

```

@prefix xsd:      <http://www.w3.org/2001/XMLSchema#> .
@prefix rdfs:     <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf:      <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix daml:     <http://www.daml.org/2001/03/daml+oil#> .
@prefix owl:    <http://www.w3.org/2002/07/owl#> .
@prefix geo:      <http://www.w3.org/2003/01/geo/wgs84_pos#> .
@prefix vcard:    <http://www.w3.org/2006/vcard/ns#> .
@prefix tl:       <http://purl.org/NET/c4dm/timeline.owl#> .

```

```

@prefix :         <http://localhost:8080/OWL/TiSAN.owl#> .

```

```

<http://localhost:8080/OWL/TiSAN.owl>

```

```

a          owl:Ontology ;
owl:versionInfo "Time-Series Action Network 1.4
(c) Minh N.T. Update: 2011-10-07"^^xsd:string .

```

```

#### Definition of main Class

```

```

:ActionClass a          owl:Class ;
rdfs:subClassOf owl:Thing .
:ActClass a owl:Class ;
rdfs:subClassOf owl:Thing .
:WhereClass a owl:Class ;
rdfs:subClassOf owl:Thing .

```



```
:WhatClass a owl:Class ;  
rdfs:subClassOf owl:Thing .
```

```
#### Definition sub-class of WhereClass
```

```
:ShopClass a owl:Class ;  
rdfs:subClassOf :WhereClass .
```

```
:RestaurantClass a owl:Class ;  
rdfs:subClassOf :WhereClass .
```

```
:TrainStationClass a owl:Class ;  
rdf:subClassOf :WhereClass .
```

```
:EvacuationClass a owl:Class ;  
rdf:subClassOf :WhereClass .
```

```
### Definition of properties
```

```
:act a owl:ObjectProperty ;  
    rdfs:label "act" ;  
    rdfs:domain :ActionClass ;  
    rdfs:range :ActClass .
```

```
:what a owl:ObjectProperty ;  
rdfs:label "what" ;  
    rdfs:domain :ActionClass ;  
    rdfs:range :WhatClass .
```

```
:where a owl:ObjectProperty ;
```

```
rdfs:label "from" ;  
rdfs:domain :ActionClass ;  
    rdfs:range :WhereClass .
```

```
#### Definition of relation
```

```
:next a owl:ObjectProperty ;  
    rdfs:label "next" ;  
    rdfs:domain :ActionClass ;  
    rdfs:range :ActionClass .
```

```
:becauseOf a owl:ObjectProperty ;  
    rdfs:label "becauseOf" ;  
    rdfs:domain :ActionClass ;  
    rdfs:range :ActionClass .
```

```
#### Definition of action frequency
```

```
:act_freq a owl:ObjectProperty ;  
rdfs:label "frequency of action" ;  
rdfs:domain :ActionClass ;  
rdfs:range :ActionClass .
```

```
#### Definition of success count
```

```
:success_cnt a owl:ObjectProperty ;  
rdfs:label "frequency of relation" ;  
rdfs:domain :ActionClass ;  
rdfs:range :ActionClass .
```

```
##### Data of the action network (sample) #####
```

```
#### Definition of action
```

```
:stop a :ActClass ;
```

```
rdfs:label "停止する"@jn , "stop"@en .
```

```
:eat a :ActClass ;
```

```
rdfs:label "食べる"@jn , "eat"@en .
```

```
:take_refuge a :ActClass ;
```

```
rdfs:label "避難する"@jn , "take refuge"@en .
```

```
:open a :ActClass ;
```

```
rdfs:label "開放する"@jn , "open"@en .
```

```
:walk a :ActClass ;
```

```
rdfs:label "歩く"@jn .
```

```
#### Definition of location
```

```
:shinjuku_station a :TrainStationClass ;
```

```
rdfs:label "新宿駅"@jn ;
```

```
vcard:locality "新宿区"@jn .
```

```
:takashimaya a :EvacuationClass ;
```

```
rdfs:label "高島屋"@jn ;
```

```
vcard:locality "新宿区"@jn ;
```

```
vcard:street-address "JR 新宿新南口" .
```

```
:bic_camera a :ShopClass ;
```

rdfs:label "ビックカメラ"@jn ;

vcard:locality "渋谷区"@jn .

:shibuya_station a :TrainStationClass ;

rdfs:label "渋谷駅"@jn ;

vcard:locality "渋谷区"@jn .

:eurospace a :EvacuationClass ;

rdfs:label "ユーロスペース"@jn ;

vcard:locality "渋谷区"@jn ;

vcard:street-address "山町1 - 5 Q-A Xビル3 F"@jn .

:akihabara_station a :TrainStationClass ;

rdfs:label "秋葉原駅"@jn, "Akihabara station"@en ;

vcard:region "東京都"@jn, "Tokyo"@en ;

vcard:locality "千代田区"@jn, "Chiyoda-ku"@en ;

vcard:street-address "外神田一丁目17-6"@jn , "1-17-6 Sotokanda"@en ;

geo:lat 35.69858 ;

geo:lat 139.773108 .

:akihabara_mc_donald a :RestaurantClass ;

rdfs:label "マクドナルド"@jn, "McDonald"@en ;

vcard:region "東京都"@jn, "Tokyo"@en ;

vcard:locality "千代田区"@jn, "Chiyoda-ku"@en ;

vcard:street-address "神田相生町1"@jn , "1 Kanda-Aioicho"@en ;

geo:lat 35.699336 ;

geo:long 139.77388 .

```
:akihabara_hotel a :EvacuationClass ;
rdfs:label "秋葉原ワシントンホテル"@jn, "Akihabara Washington Hotel"@en ;
vcard:region "東京都"@jn, "Tokyo"@en ;
vcard:locality "千代田区"@jn, "Chiyoda-ku"@en ;
vcard:street-address "神田佐久間町1丁目8?3"@jn , "1-8-3 Sakuma-cho, Kanda"@en ;
geo:lat 35.697297 ;
geo:long 139.773477 .
```

```
##### Activities #####
```

```
:act_01 a :ActionClass ;
:act :stop ;
:what "電車"@jn ;
:where :shinjuku_station ;
tl:start "2011-03-11T15:13:00"^^xsd:dateTime ;
tl:end "2011-03-11T23:45:00"^^xsd:dateTime .
```

```
:act_02 a :ActionClass ;
:act "待つ"@jn ;
:what "タクシー"@jn ;
:where :shinjuku_station ;
tl:start "2011-03-11T17:13:00"^^xsd:dateTime ;
:success_cnt -1 .
```

```
:act_03 a :ActionClass ;
:act :walk ;
:where :shinjuku_station ;
tl:start "2011-03-11T19:52:09"^^xsd:dateTime ;
:next :act_04 .
```

```
:act_04 a :ActionClass ;  
:act :take_refuge ;  
:where :takashimaya ;  
tl:start "2011-03-11T20:22:09"^^xsd:dateTime ;  
:act_freq 10 ;  
:success_cnt 8 .
```

```
:act_05 a :ActionClass ;  
:act "乗る"@jn ;  
:what "バス"@jn ;  
:where :shinjuku_station ;  
tl:start "2011-03-11T18:20:00"^^xsd:dateTime ;  
:next :act_06 .
```

```
:act_06 a :ActionClass ;  
:act :walk ;  
:where :shibuya_station ;  
tl:start "2011-03-11T20:20:00"^^xsd:dateTime ;  
:next :act_7 .
```

```
:act_7 a :ActionClass ;  
:act "貰う"@jn ;  
:what "携帯バッテリー"@jn ;  
:where :bic_camera ;  
tl:start "2011-03-11T20:40:00"^^xsd:dateTime ;  
:next :act_8 .
```

100

```
:act_08 a :ActionClass ;  
:act :take_refuge ;  
:where :eurospace ;  
tl:start "2011-03-11T21:02:00"^^xsd:dateTime .
```

```
:act_9 a :ActionClass ;  
:act "無料配布する"@jn ;  
:what "携帯バッテリー"@jn ;  
:where :bic_camera ;  
tl:start "2011-03-11T19:40:00"^^xsd:dateTime .
```

```
:act_10 a :ActionClass ;  
:act :open ;  
:where :eurospace ;  
tl:start "2011-03-11T20:10:40"^^xsd:dateTime .
```

```
:act_11 a :ActionClass ;  
:act :stop ;  
:what "電車"@jn, "train"@en ;  
:where :akihabara_station ;  
tl:start "2011-03-11T16:13:00"^^xsd:dateTime ;  
tl:end "2011-03-11T23:45:00"^^xsd:dateTime .
```

```
:act_12 a :ActionClass ;  
:act :eat ;  
:what "ビックマック"@jn, "Big Mac" ;  
:where :akihabara_mc_donald ;  
tl:start "2011-03-11T17:13:00"^^xsd:dateTime ;
```

```
:act_freq 10 ;
:success_cnt 7 ;
:becauseOf :act_11, [ rdf:value 3 ];
:next :act_13, [ rdf:value 8 ] .

:act_13 a :ActionClass ;
:act :take_refuge ;
:where :akihabara_hotel ;
tl:start "2011-03-11T18:20:00"^^xsd:dateTime ;
:act_freq 70 ;
:success_cnt 50 .

:act_14 a :ActionClass ;
:act :open ;
:where :akihabara_hotel ;
tl:start "2011-03-11T16:00:00"^^xsd:dateTime ;
tl:end "2011-03-12T09:00:00"^^xsd:dateTime .
```

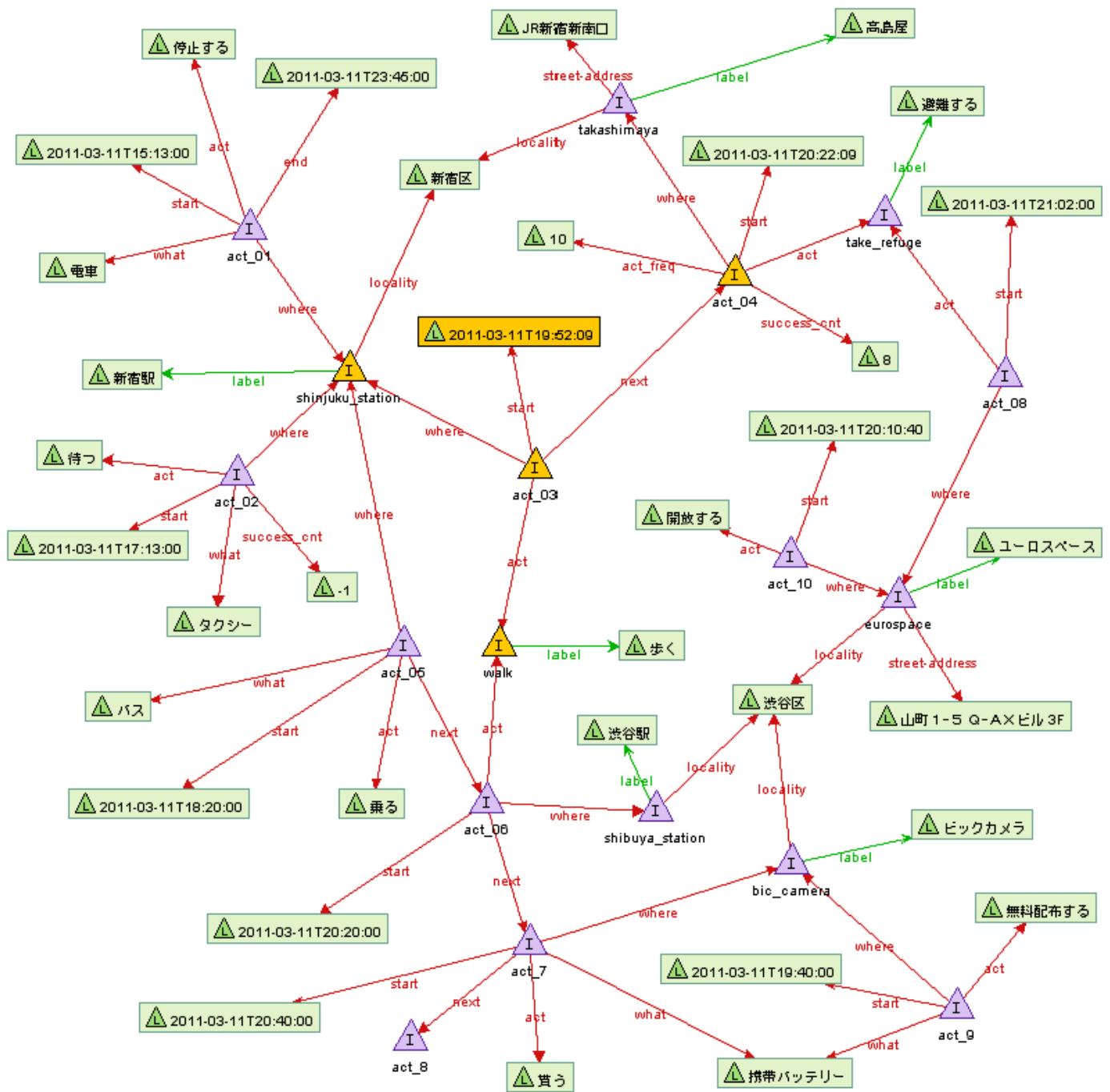



図 6.1: 2011 年 3 月 11 日の大震災直後，新宿駅周辺での避難行動の一部（C, I, L はそれぞれクラス，インスタンス，ラベルを表す）

付録B 評価実験のデータおよび結果

- 避難に関するつぶやき: <http://p.tl/Jbaw>
- ユーザの避難行動: <http://p.tl/P09B>
- 行動を補完した結果: <http://p.tl/VwKh>

研究業績（22件）

学術雑誌（3件）

1. The-Minh Nguyen, Takahiro Kawamura, Yasuyuki Tahara, and Akihiko Ohsuga: Self-Supervised Capturing of Users' Activities from Weblogs. International Journal of Intelligent Information and Database Systems, Vol.6, No.1, pp.61-76, InderScience Publishers, 2012.
2. 田中俊行, グエン ミン テイ, 中川博之, 田原康之, 大須賀昭彦: 評判分析システムのための教師あり学習を用いた意見抽出. 電子情報通信学会論文誌 D, ソフトウェアエージェントとその応用特集号, Vol.J94-D, No.11, pp.1751-1761, 2011.
3. グエン ミン テイ, 川村隆浩, 中川博之, 田原康之, 大須賀昭彦: 条件付確率場と自己教師あり学習を用いた行動属性の自動抽出と評価. 人工知能学会論文誌, pp.166-178, 26 巻 1 号 SP-R, 2011.

国際会議（8件）

4. The-Minh Nguyen, Takahiro Kawamura, Yasuyuki Tahara, and Akihiko Ohsuga: Building a Time Series Action Network for Earthquake Disaster. 4th International Conference on Agents and Artificial Intelligence (ICAART2012), volume 1, pp.100-108, Vilamoura, Algarve, Portugal, 6-8 February, 2012. In Cooperation with: Portuguese Association for Artificial Intelligence (APPIA), Spanish Association of Artificial Intelligence (AEPIA), Association for the Advancement of Artificial Intelligence (AAAI).

5. **The-Minh Nguyen**, Kenji Koshikawa, Takahiro Kawamura, Yasuyuki Tahara, and Akihiko Ohsuga: Building Earthquake Semantic Network by Mining Human Activity from Twitter. 2011 IEEE International Conference on Granular Computing, pp.496-501. Kaohsiung, Taiwan, 8-10 November, 2011.
6. Isabel Shizu Miyamae Iwanaga, **The-Minh Nguyen**, Takahiro Kawamura, Hiroyuki Nakagawa, Yasuyuki Tahara, and Akihiko Ohsuga: Building an Earthquake Evacuation Ontology from Twitter. 2011 IEEE International Conference on Granular Computing, pp.306-311. Kaohsiung, Taiwan, 8-10 November, 2011.
7. **Nguyen Minh The**, Takahiro Kawamura, Yasuyuki Tahara, Akihiko Ohsuga: Capturing Users' Buying Activity at Akihabara Electric Town from Twitter. In Proc. ICCCI2010, part II, LNAI 6422, pp.163-171. Springer, Heidelberg, Kaohsiung, Taiwan, November 10-12, 2010.
8. **Nguyen Minh The**, Takahiro Kawamura, Hiroyuki Nakagawa, Yasuyuki Tahara, Akihiko Ohsuga: Automatic Mining of Human Activity Attributes from Weblogs. In Proc. IEEE/ACIS ICIS2010, Semantic, Social, and Collaborative Intelligence Session, pp.633-644, Yamagata, Japan, 18-20 August, 2010.
9. **Nguyen Minh The**, Takahiro Kawamura, Hiroyuki Nakagawa, Yasuyuki Tahara, Akihiko Ohsuga: Automatic Mining of Human Activity and Its Relationships from CGM. In Proc. ICSOFT2010, volume 1, pp.285-292, Athens, Greece, 22-24 July, 2010.
10. **Nguyen Minh The**, Takahiro Kawamura, Hiroyuki Nakagawa, Ken Nakayama, Yasuyuki Tahara, Akihiko Ohsuga: Human Activity Mining using Conditional Random Fields and Self-Supervised Learning. In Proc. ACIIDS2010, part I, LNAI 5990, pp.140-149. Springer, Heidelberg. Hue City, Vietnam, 24-26 March, 2010.
11. Takahiro Kawamura, **Nguyen Minh The**, Akihiko Ohsuga: Building of Human

Activity Correlation Map from Weblogs. In Proc. ICSOFT2009, volume 2, pp. 346-352. INSTICC Press, ISBN 978-989-674-010-8. Sofia, Bulgaria, 26-29 July, 2009.

査読付国際ワークショップ（1件）

12. Nguyen Minh The, Takahiro Kawamura, Hiroyuki Nakagawa, Yasuyuki Tahara, Akihiko Ohsuga: Self-Supervised Mining of Human Activity from CGM. The 11th International Workshop on Knowledge Management and Acquisition for Smart Systems and Services (PKAW 2010), LNAI 6232, pp.61-72, Springer, Heidelberg. Daegu, Korea, 30 August - 3 September, 2010.

査読付国内シンポジウム（5件）

13. グエン ミン テイ, 越川 兼地, 川村隆浩, 田原康之, 大須賀昭彦: 震災時の効率的な避難のための行動推薦エージェント Ready for 87%の提案(2) ~ 時系列避難行動ネットワークの構築 ~ . 合同エージェントワークショップ&シンポジウム 2011 . 主催: マルチエージェントと協調計算研究会 (日本ソフトウェア科学会), 人工知能と知識処理研究会 (電子情報通信学会), 知能システム研究会 (情報処理学会), 知識ベースシステム研究会 (人工知能学会), IEEE Computer Society Japan Chapter, 2011年10月26~28日 .
14. ミヤマエ イワナガ イサベル シズ, グエン ミン テイ, 川村隆浩, 中川博之, 田原康之, 大須賀昭彦: 震災時の効率的な避難のための行動推薦エージェント Ready for 87%の提案(1) ~ 避難計画オントロジーの設計 ~ . 合同エージェントワークショップ & シンポジウム 2011 . 主催: マルチエージェントと協調計算研究会 (日本ソフトウェア科学会), 人工知能と知識処理研究会 (電子情報通信学会), 知能システム研究会 (情報処理学会), 知識ベースシステム研究会 (人工知能学会), IEEE Computer Society Japan Chapter, 2011年10月26~28日 .
15. 伊藤雅博, グエン ミン テイ, 川村隆浩, 中川博之, 田原康之, 大須賀昭彦: 震災時の効率的な避難のための行動推薦エージェント Ready for 87%の提案(3) ~ 行動ネッ

トワークを用いた避難行動の推薦～．合同エージェントワークショップ&シンポジウム 2011．主催：マルチエージェントと協調計算研究会（日本ソフトウェア科学会），人工知能と知識処理研究会（電子情報通信学会），知能システム研究会（情報処理学会），知識ベースシステム研究会（人工知能学会），IEEE Computer Society Japan Chapter, 2011年10月26～28日．

16. 田中俊行，ゲン ミン テイ，中川博之，田原康之，大須賀昭彦：教師あり学習を用いたレビューサイトからの意見抽出．合同エージェントワークショップ&シンポジウム 2010．主催：マルチエージェントと協調計算研究会（日本ソフトウェア科学会），人工知能と知識処理研究会（電子情報通信学会），知能システム研究会（情報処理学会），知識ベースシステム研究会（人工知能学会）．後援：IEEE Computer Society Japan Chapter, 2010年10月27～29日（学生奨励賞）
17. ゲン ミン テイ，川村隆浩，中川博之，中山健，田原康之，大須賀昭彦：条件付確率場と自己教師あり学習を用いた行動属性の自動抽出と評価．合同エージェントワークショップ&シンポジウム 2009，pp. 468-475．主催：マルチエージェントと協調計算研究会（日本ソフトウェア科学会），人工知能と知識処理研究会（電子情報通信学会），知能と複雑系研究会（情報処理学会），知識ベースシステム研究会（人工知能学会）．後援：IEEE Computer Society Japan Chapter, 2009年10月28～30日．

国内大会・研究会（5件）

18. ゲン ミン テイ，伊藤雅博，川村隆浩，中川博之，田原康之，大須賀昭彦：Webからの実世界行動抽出による意味ネットワーク構築手法の提案．IEICE・人工知能と知識処理研究会，vol.111, no.70, AI2011-8, pp.39-44, 2011年5月26日．
19. 田中俊行，ゲン ミン テイ，中川博之，田原康之，大須賀昭彦：シーソー型インタフェースを用いた評判比較システムの開発．情報処理学会，インタラクシオン 2011，2011年3月10～12日．
20. ゲン ミン テイ，川村隆浩，田原康之，大須賀昭彦：Twitterからの人間行動属性の

- 自動抽出 . IEICE・人工知能と知識処理研究会 (Artificial Intelligence and Knowledge-Base Processing) , 一般講演 (理論) , pp.19-23 . 関学東京キャンパス , 2010 年 6 月 18 日 .
21. ゲン ミン テイ , 川村隆浩 , 中川博之 , 田原康之 , 大須賀昭彦 : CGM からの自己教師あり学習と条件付確率場を用いた人間行動マイニング . 人工知能学会 (第 24 回) 全国大会 (JSAI2010) , 2010 年 6 月 9 ~ 11 日 .
22. ゲン ミン テイ , 川村隆浩 , 中川博之 , 田原康之 , 大須賀昭彦 : Web からの自己教師あり学習を用いた人間行動マイニング . IEICE・人工知能と知識処理研究会 , Vol.109, No.386, pp.19-24. 「グリーン AI」および一般 . 関学東京キャンパス , 2010 年 1 月 15 日 .

関連論文の印刷公表の方法及び時期

学術雑誌（2件）

1. 全著者名：The-Minh Nguyen, Takahiro Kawamura, Yasuyuki Tahara, and Akihiko Ohsuga
論文題目：Self-Supervised Capturing of Users' Activities from Weblogs
印刷公表の方法及び時期：International Journal of Intelligent Information and Database Systems, Vol.6, No.1, pp.61-76, InderScience Publishers, 2012
（第2章の内容）
2. 全著者名：ゲン ミン テイ，川村隆浩，中川博之，田原康之，大須賀昭彦
論文題目：条件付確率場と自己教師あり学習を用いた行動属性の自動抽出と評価
印刷公表の方法及び時期：人工知能学会論文誌，pp.166-178，26 巻1号 SP-R，2011
（第2章及び第5章の内容）

国際会議（6件）

3. 全著者名：The-Minh Nguyen, Takahiro Kawamura, Yasuyuki Tahara, and Akihiko Ohsuga
論文題目：Building a Time Series Action Network for Earthquake Disaster
印刷公表の方法及び時期：4th International Conference on Agents and Artificial Intelligence (ICAART2012), volume 1, pp.100-108, Vilamoura, Algarve, Portugal, 6-8 February, 2012. In Cooperation with: Portuguese Association for Artificial Intelligence (APPIA), Spanish Association of Artificial Intelligence (AEPIA), Association for the Advancement of Artificial Intelligence (AAAI)

(第 4 章の内容)

4. 全著者名 : The-Minh Nguyen, Kenji Koshikawa, Takahiro Kawamura, Yasuyuki Tahara, and Akihiko Ohsuga

論文題目 : Building Earthquake Semantic Network by Mining Human Activity from Twitter

印刷公表の方法及び時期 : 2011 IEEE International Conference on Granular Computing, pp.496-501. Kaohsiung, Taiwan, 8-10 November, 2011

(第 3 章の内容)

5. 全著者名 : Nguyen Minh The, Takahiro Kawamura, Yasuyuki Tahara, Akihiko Ohsuga

論文題目 : Capturing Users' Buying Activity at Akihabara Electric Town from Twitter

印刷公表の方法及び時期 : In Proc. ICCCI2010, part II, LNAI 6422, pp.163-171. Springer, Heidelberg, Kaohsiung, Taiwan, 10-12 November, 2010

(第 5 章の内容)

6. 全著者名 : Nguyen Minh The, Takahiro Kawamura, Hiroyuki Nakagawa, Yasuyuki Tahara, Akihiko Ohsuga

論文題目 : Automatic Mining of Human Activity Attributes from Weblogs

印刷公表の方法及び時期 : In Proc. IEEE/ACIS ICIS2010, Semantic, Social, and Collaborative Intelligence Session, pp.633-644, Yamagata, Japan, 18-20 August, 2010

(第 2 章の内容)

7. 全著者名 : Nguyen Minh The, Takahiro Kawamura, Hiroyuki Nakagawa, Yasuyuki Tahara, Akihiko Ohsuga

論文題目 : Automatic Mining of Human Activity and Its Relationships from CGM

印刷公表の方法及び時期 : In Proc. ICSOFT2010, volume 1, pp.285-292, Athens, Greece, 22-24 July, 2010

(第 2 章の内容)

8. 全著者名:Nguyen Minh The, Takahiro Kawamura, Hiroyuki Nakagawa, Ken Nakayama, Yasuyuki Tahara, Akihiko Ohsuga
論文題目:Human Activity Mining using Conditional Random Fields and Self-Supervised Learning
印刷公表の方法及び時期 : In Proc. ACIIDS2010, part I, LNAI 5990, pp.140-149. Springer, Heidelberg. Hue City, Vietnam, 24-26 March, 2010
(第 2 章の内容)

査読付国際ワークショップ (1 件)

9. 全著者名 : Nguyen Minh The, Takahiro Kawamura, Hiroyuki Nakagawa, Yasuyuki Tahara, Akihiko Ohsuga
論文題目 : Self-Supervised Mining of Human Activity from CGM
印刷公表の方法及び時期 : The 11th International Workshop on Knowledge Management and Acquisition for Smart Systems and Services (PKAW 2010), LNAI 6232, pp.61-72, Springer, Heidelberg. Daegu, Korea, 30 August - 3 September, 2010
(第 2 章の内容)

参考論文の印刷公表の方法及び時期

1. 全著者名：田中俊行，グエン ミン テイ，中川博之，田原康之，大須賀昭彦
論文題目：評判分析システムのための教師あり学習を用いた意見抽出
印刷公表の方法及び時期：電子情報通信学会論文誌 D，ソフトウェアエージェントとその応用特集号，Vol.J94-D ，No.11 ，pp.1751-1761 ，2011
2. 全著者名：Isabel Shizu Miyamae Iwanaga, The-Minh Nguyen, Takahiro Kawamura, Hiroyuki Nakagawa, Yasuyuki Tahara, and Akihiko Ohsuga
論文題目：Building an Earthquake Evacuation Ontology from Twitter
印刷公表の方法及び時期：2011 IEEE International Conference on Granular Computing, pp.306-311. Kaohsiung, Taiwan, 8-10 November, 2011

著者略歴

NGUYEN MINH THE (グエン ミン テイ)

- 1981年10月17日 ベトナム・ハノイに生まれる
- 1999年9月 ハノイ工科大学入学
- 2001年4月 国費留学生として日本に留学
- 2002年3月 東京日本語教育センター JASSO 卒業
- 2005年3月 長野高専電子情報工学科卒業
- 2007年3月 室蘭工業大学情報工学科卒業
- 2007年4月 電気通信大学大学院情報システム学研究科社会知能情報学専攻
博士前期課程入学
- 2009年9月 電気通信大学大学院情報システム学研究科社会知能情報学専攻
博士前期課程卒業
- 2009年10月 電気通信大学大学院情報システム学研究科社会知能情報学専攻
博士後期課程入学
- 2012年3月 電気通信大学大学院情報システム学研究科社会知能情報学専攻
博士後期課程修了予定