

RDF を用いたユーザのコンテキスト推定方法

2012SE190 尾形 純輝 2012SE229 塩見 進 2012SE239 鈴木 健人

指導教員 青山 幹雄

1. 研究背景

近年、コンテキストウェアサービスが注目を集めている。そのため、ユーザを取り巻く膨大なコンテキストデータを構造化する方法やセンサなどで収集できないようなユーザのコンテキストを推定する方法が求められている。

2. 研究課題

本研究ではセンサなどで収集できないユーザのコンテキストを推定するため、以下の2点を研究課題とする。

- (1)ユーザのコンテキストデータ構造化
- (2)ユーザのコンテキスト推定

3. 関連研究

(1)コンテキストウェアサービス

コンテキストウェアサービスとは、ユーザのコンテキストを収集しコンピュータがそのコンテキストに応じたサービスを提供する概念である[4]。コンテキストウェアサービスは計測と収集、分析、実行の3つの機能からなる。

(2)RDF

RDF(Resource Description Framework)とは、リソースの意味を定義する仕組みである。RDF は主語、述語、目的語の3要素で記述する[7]。

(3)SPARQL

SPARQL(SPARQL Protocol And RDF Query Language)とは、RDFDB に問い合わせを行うためのクエリ言語である[8]。

(4)コサイン類似度

コサイン類似度とは、2つのベクトル間の類似度をベクトルの内積として定義したものである。

4. アプローチ

本研究では、収集したコンテキストデータを RDF を用いて構造化することで、知識領域を前提とせずに運用できる形でリソースを記述できると考えた。また、推定にコサイン類似度を用いることで、既存のコンテキストデータとユーザの現在のコンテキストをベクトルで表現し、これらのベクトル間の類似度を評価できると考えた。類似度の高いデータをユーザのコンテキストであると推定できることから、精度の高い推定が可能となると考えた。

5. 提案方法

5.1. コンテキストの定義

本研究では Dey らのコンテキストの定義[3]を参考にし、RDF グラフでのコンテキストの定義を提案する。コンテキストを「RDF グラフにおいて、ルートがユーザであるすべてのリソースとリテラル」と定義する。

5.2. ユーザのコンテキスト推定システムの構成

ユーザのコンテキスト推定システムを図1に示す。なお、

本研究ではコンテキストの推定までを取り扱い、サービス提供については研究の対象外とする。

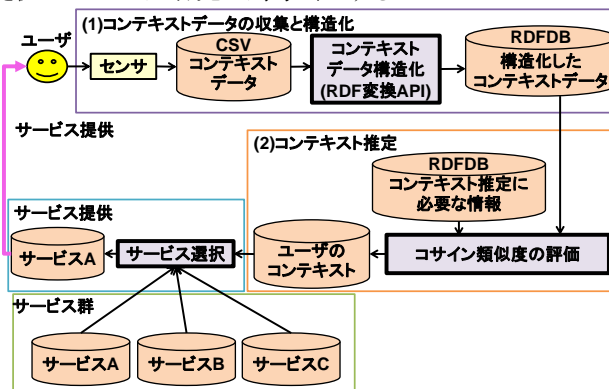


図1 ユーザのコンテキスト推定システムの構成

(1)コンテキストデータの収集と構造化

サービス提供者が推定したいコンテキストに合わせて、センサでコンテキストデータを収集する。センサで収集するコンテキストデータはCSV形式とする。

収集したコンテキストデータを自動でRDFに変換し構造化する。構造化したコンテキストデータは構造化したコンテキストデータのRDFDBに格納する。

(2)コンテキスト推定

構造化したコンテキストデータとコンテキスト推定に必要な情報のコサイン類似度を評価し、ユーザのコンテキストを推定する。コンテキスト推定に必要な情報はサービス提供者が予めコンテキスト推定に必要な情報のRDFDBに格納する。

5.3. コンテキスト推定システムの振舞い

コンテキスト推定システムの振舞いを図2に示す。

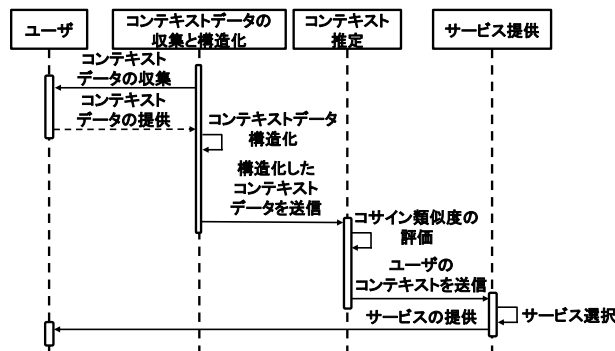


図2 コンテキスト推定システムの振舞い

5.4. RDFによるコンテキストデータの構造化

RDFでコンテキストデータを構造化するために、コンテキストデータのデータモデルを図3に示す。また、コン

キストの定義より、図 3 のコンテキストを主語としたリソース、またはリテラルが目的語に記述できる。



図 3 コンテキストデータのデータモデル

5.5. RDF によるコンテキスト推定に必要な情報の構造化

RDF でコンテキスト推定に必要な情報を構造化するために、コンテキスト推定に必要な情報のデータモデルを図 4 に示す。要素とは、コンテキスト推定に必要な情報を抽出するために必要なデータであり、要素とコンテキストが一致するならば、同一の URI で記述する。



図 4 コンテキスト推定に必要な情報のデータモデル

5.6. ユーザのコンテキスト推定のためのデータの抽出

SPARQL を用いて、構造化したコンテキストデータの RDFDB に格納したデータからコンテキストを抽出する。抽出したコンテキストを基に、コンテキスト推定に必要な情報の RDFDB に格納したデータを抽出する。

5.7. ユーザのコンテキスト推定

(1) 重み付け

SPARQL で抽出したコンテキストとコンテキスト推定に必要な情報をベクトル化するために重み付けが必要がある。コンテキストの重み付けはサービス提供者がサービスの内容にあった設定をする。

(2) 重み付けに基づくベクトル化

SPARQL で抽出したコンテキスト推定に必要な情報を評価するために、SPARQL で抽出したコンテキストとコンテキスト推定に必要な情報を重み付けに基づいてベクトル化する。

(3) コサイン類似度の評価方法

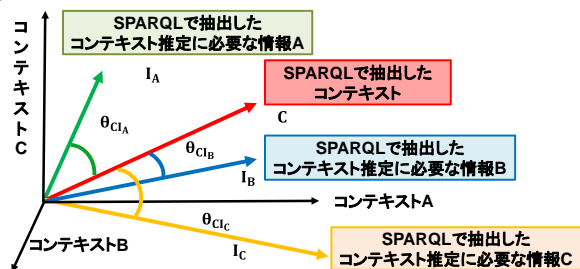


図 5 コサイン類似度の例

$$\cos\theta_{CI_n} = \frac{C \cdot I_n}{\|C\| \|I_n\|} \quad (1)$$

ベクトル化を行った SPARQL で抽出したコンテキストとコンテキスト推定に必要な情報を基に、コサイン類似度の評価を行う。SPARQL で抽出したコンテキストとコンテキスト推定に必要な情報のコサイン類似度をそれぞれ評価し、コサイン類似度が十分に推定可能である値以上であった SPARQL で抽出したコンテキスト推定に必要な情報をユ

ーザのコンテキストと推定する。なお、この値は推定する内容に合わせて設定する。コサイン類似度の例を図 5 に示し、コサイン類似度の求め方を式(1)に示す。

6. シナリオへの適用

6.1. 実行環境

本シナリオを実行するための図 1 に示すコンテキスト推定システムの実行環境を表 1 に示す。

表 1 実行環境

OS	Windows8.1
CPU	Intel® Celeron® CPU G1820 , 2.70GHz
メモリ	4.00GB
RDF 変換 API	MetaBrige[5]のメタデータ変換 API
SPARQL エンドポイント	http://localhost:8890/basket http://localhost:8890/store http://localhost:8890/recipe
RDFDB	PostgreSQL

6.2. シナリオの概要と目的

(1) シナリオの概要

シナリオとして、ユーザは夕食にカレーを作ることを目的としており、本シナリオでは、ユーザが作る料理を「カレー」と推定する。

(2) シナリオの目的

本シナリオでは、ユーザが購入しようとしている商品から作るようとしている料理を推定することで本提案の妥当性を示すことを目的とする。

(3) 前提条件

1) コンテキストデータの取得

ユーザが買い物かごに商品を入れるたびに、センサが商品の商品 ID を読み取る。読み取った商品 ID を CSV 形式のファイルで保存する。CSV 形式のファイルを Turtle 形式に変換する API である MetaBridge を使用し、先ほどのファイルを Turtle 形式の RDF に変換する。

2) コンテキスト推定システムの本シナリオへの適用

シナリオに適用するために、RDFDB に配置するデータを表 2 に示す。

表 2 RDFDB へのデータ配置

RDFDB	配置 RDFDB	格納データ
構造化したコンテキストデータ	買い物かご RDFDB	商品 ID
	商品 RDFDB	商品名, 商品の情報, 商品の分類
コンテキスト推定に必要な情報	レシピ RDFDB	料理 ID, レシピ名, レシピの分類, レシピの材料, 食材の分類, 食材の分量

本シナリオでは、レシピ RDFDB に格納されているレシピをクックパッド[1]を参考にいくつか選択した。レシピは、「ルーを使ったカレー」に使われている食材を基準にし、他のレシピで基準の食材が使用されている割合を基準に選択した。表 3 に例を示す。

表 3 選択したレシピの分類と基準の食材の使用割合

レシピ名	レシピ分類	割合[%]
ルーを使ったカレー	カレー	100
チーズカレー	カレー	88
クリームシチュー	シチュー	75
スパイスを使ったカレー	カレー	40
カレースパゲティ	パスタ	25
トマトパスタ	パスタ	0

6) コンテキスト推定システムの実行制御

本シナリオでは、構造化した買い物かごの中の商品の中で食品のデータが 3 個目になったときに推定を開始し、食品が 1 個増えるごとに推定を繰り返す。ユーザが会計を行った時点で本システムを終了する。本シナリオ終了時のユーザの買い物かごの中の食品を含む商品とその分類、センサが読み取った順番を表 4 に示す。

表 4 本シナリオでセンサが読み取った商品

順番	商品名	分類
1	にんじん	食品
2	じゃがいも	食品
3	たまねぎ	食品
4	鶏もも肉	食品
5	カレールー	食品
6	たまご	食品
7	醤油	調味料
8	牛乳	食品
9	水	その他
10	ビール	酒
11	はみがき粉	生活用品
12	ガム	菓子

7) コンテキスト推定方法

本シナリオでは、買い物かごの中の食品とレシピの食材からコサイン類似度を用いてユーザが作る料理を推定する。本シナリオではコサイン類似度 0.60 以上であった場合、十分な推定ができたと仮定する。

6.3. ユーザのコンテキスト推定に必要なデータの収集

(1) シナリオ

ユーザが夕食の食材等をスーパーで買い物をしている。推定 1 回目のとき、ユーザの買い物かごには、「にんじん」、「じゃがいも」、「たまねぎ」が入っている。

(2) ユーザのコンテキスト収集

センサが買い物かごの商品を読み取り、CSV 形式のファイルで保存する。

(3) コンテキストデータ構造化

収集した CSV 形式の買い物かごのデータを Turtle 形式の RDF に変換する。ユーザのコンテキスト収集からコンテキストの構造化まで自動で行う。

(4) SPARQL を用いたデータの抽出手順

本シナリオで扱う構造化したコンテキストデータとコンテキスト推定に必要な情報の RDFDB から推定で扱うデー

タの抽出手順を図 6 に示す。

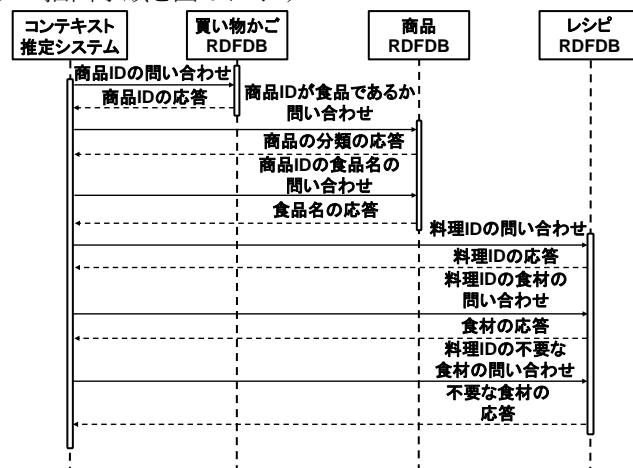


図 6 SPARQL を用いたデータの抽出手順

1) 買い物かごの RDFDB に買い物かごを主語とする目的語の商品 ID を SPARQL で問い合わせる。その結果、買い物かごにある商品の商品 ID を得る。

2) 商品 RDFDB に商品 ID を主語として目的語が食品であるかを SPARQL で問い合わせる。True が返ってきた商品 ID を抽出する。このとき、買い物かごの食品が 3 個未満の場合は、推定は開始せず、コンテキストの収集を続ける。

3) 商品 RDFDB に 2) で得た商品 ID を主語とする目的語の食品名を SPARQL で問い合わせる。その結果、食品である商品 ID の食品名を得る。

4) レシピの RDFDB に目的語に 3) で得た食品名を持つ主語の料理 ID を SPARQL で問い合わせる。その結果、買い物かごにある食品を使った料理 ID を得る。

5) レシピの RDFDB に主語を料理 ID とする目的語の食材を SPARQL で問い合わせる。その結果、4) で得た料理 ID が持つ食材を得る。

6) レシピの RDFDB に料理 ID を主語とする目的語の食材が類似度の評価に不要かを SPARQL で問い合わせ、不要ならば True を返す。その後、4) で得た料理 ID が持つ食材のうち不要とされたものを除く。

6.4. ユーザのコンテキスト推定のための類似度評価

コンテキスト推定方法は、食材の重要度と食品群を考慮したレシピ間類似度の算出方法[2]で扱われていた重要度の算出方法を用いて重みの評価する。また、レシピ間類似度の算出方法を用いて買い物かごの食品とレシピの食材を用いて類似度評価する。

(1) 重みの評価方法

本シナリオでは、重みとして重要度という値を用いる。これは、その食材が料理においてどの程度使用されているかを表す数値である。これを、Web 上の食材単体及び食品群の出現割合から算出する。食品群については五訂増補日本食品標準成分表[6]を参照にした。

(2) シナリオの類似度評価方法

レシピ RDFDB からレシピ名「ルーを使ったカレー」を例

として類似度評価方法を説明する。

買い物かごに入っている食品をベクトル A 、レシピの RDFDB で推定に必要な食材をベクトル B とする。各ベクトルの次元は、それぞれの要素の和集合とする。この和集合はレシピの集合と買い物かごの集合を接続したものである。要素の値はその食材の重要度である。ベクトル A 、ベクトル B の例をそれぞれ式(2)、式(3)に示す。

$$A(\text{カレールー, 鶏もも肉, たまねぎ, じゃがいも, にんじん}) \\ = (0, 0, 0.58, 0.41, 0.37) \quad (2)$$

$$B(\text{カレールー, 鶏もも肉, たまねぎ, じゃがいも, にんじん}) \\ = (0.31, 0.42, 0.58, 0.41, 0.37) \quad (3)$$

この2つのベクトル間類似度を求める式を式(4)に示す。

$$\cos\theta_{BA} = \frac{B \cdot A}{\|B\| \|A\|} = 0.84 \quad (4)$$

同様に他のレシピでもコサイン類似度で評価を行った結果が表5の1回目の列に当たる。推定が終わった時点で、類似度が0.60以上の料理IDを類似度の高い順番に返す。この料理IDを用いて SPARQL でレシピの分類を問い合わせる。よって、1回目の推定結果は「カレー」、「クリームシチュー」となる。

本シナリオでは、推定は最後の食品である商品名「牛乳」が買い物かごに入ったときの5回目までである。これら5回分の推定結果を表5に示す。

表5 本シナリオで扱った5回分のコサイン類似度

レシピ名	1回目	2回目	3回目	4回目	5回目
ルーを使ったカレー	0.84	0.94	1.00	0.80	0.76
チーズカレー	0.65	0.73	0.78	0.62	0.59
クリームシチュー	0.71	0.74	0.69	0.57	0.67
スパイスを使ったカレー	0.64	0.72	0.68	0.54	0.52
カレースバゲティ	0.17	0.16	0.26	0.22	0.21
トマトパスタ	0.00	0.00	0.00	0.00	0.00

7. 評価と考察

7.1. 妥当性の評価

(1) ユーザのコンテキストデータ構造化

ユーザのコンテキストデータを RDF で表現することで、意味に基づいた検索が可能となった。本提案をシナリオに適用した結果、買い物かごに入れた商品とレシピサイトのレシピの材料を同一の URI で記述することで、同一の意味定義が可能となった。これにより、ユーザのコンテキストを基にコンテキスト推定に必要な情報を SPARQL で検索可能となった。

(2) ユーザのコンテキスト推定

ユーザのコンテキストの要素とコンテキスト推定に必要な情報の要素をコサイン類似度で評価することで、ユーザのコンテキストが推定可能となった。また、コンテキストを構成している要素をカテゴライズし、要素が持つ重みを付加することでより推定の精度が向上した。本提案をシナリオに適用した結果、表5からすべての推定でレシピで扱う食品が得られた場合は類似度が増加し、それ以外の場合は減少することを確認した。また、ユーザのコンテキスト

推定結果が正しいことも確認した。本シナリオのユーザの目的はレシピの分類で「wd:カレー」であるが、表3の「スパイスを使ったカレー」と「クリームシチュー」の基準の食材の使用率を比較すると、0.35の差で「クリームシチュー」が大きい。しかし、本提案方法で推定を行った結果、表5の3回目の推定結果より差は0.01に減少し、精度が向上した。よって、本提案は動的なユーザのコンテキストを正しく推定可能であるという点で有効であると評価できる。

7.2. 考察

本研究で収集したコンテキスト推定に必要な情報の要素の中で「カレールー」の重要度が他の食材と比べ低くなった。これはクックパッドで扱われているレシピに「カレールー」を使った「カレー」が少ないためと考えられる。ユーザのコンテキスト推定の精度向上のため、扱うデータは偏りがないように収集する必要がある。

8. 今後の課題

- (1)推定したユーザのコンテキストが実際のユーザのコンテキストに合致するか確認する必要がある。また、ユーザのコンテキストに合致しない場合は、その結果を修正するシステムを構成する必要がある。
- (2)複数の項目の推定を行うために、扱うユーザのコンテキストやコンテキスト推定に必要な情報を、推定する内容に合わせて抽出する必要がある。

9. まとめ

本研究では、コンテキストアウェアサービスを提供するために RDF を用いてユーザのコンテキストを推定する方法を提案した。推定では、ユーザのコンテキストを収集し、RDFを用いて構造化し、構造化したユーザのコンテキストとコンテキスト推定に必要な情報に重み付けをし、コサイン類似度を用いて評価する方法を提案した。また、提案方法をシナリオへ適用することで妥当性を示した。

10. 参考文献

- [1]Cookpad Inc., レシピ検索 No.1/料理レシピ載せるならクックパッド, <http://cookpad.com/>.
- [2]福本 亜紀, ほか, 食材の重要度と食品群を考慮したレシピ間類似度の算出方法, 和歌山大学大学院システム工学研究科, 和歌山大学システム工学部, 2012.
- [3]藤波 香織, 分散コンテキストアウェアシステムに関する研究, 早稲田大学院理工学研究科博士論文, 2005.
- [4]市川 裕也, ほか, コンテキストアウェアサービスモデルの設計方法の提案と評価, 2014年度南山大学情報理工学ソフトウェア工学科卒業論文, 2015.
- [5](一社)メタデータ基盤協議会, メタデータ変換 API チュートリアル, https://www.metabridge.jp/tutorial_mdapi.html.
- [6]文部科学省, 五訂増補日本食品標準成分, http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu3/toushin/05031802.htm.
- [7]W3C, RDF 1.1 Primer, <http://www.w3.org/TR/rdf11-primer/>.
- [8]W3C, SPARQL Query Language for RDF, <https://www.w3.org/TR/rdf-sparql-query/>.