

経験の分散表現による意味類似性判定と経験価値を提供する地物検索

Semantic Similarity Calculation based on Distributed Representation of Experiences and Search for Geographic Objects Providing Experiential Value

松村 優也[†]

Yuya Matsumura

大島 裕明[‡]

田中 克己[†]

吉川 正俊[†]

Hiroaki Ohshima

Katsumi Tanaka

Masatoshi Yoshikawa

1. はじめに

本研究では、「安眠する」、「女性とちよつとお酒を飲む」のような経験を動詞句で表現したものを入力とし、その経験ができる地物を検索する手法を提案する。

情報検索を行うユーザの意図の一つには、ある経験を想定し、その経験を享受できる地物を知りたい、というものがある。たとえば、「コーヒーを飲む」、「本を買う」、「安眠する」、「女性とちよつとお酒を飲む」などの経験を想定する。そして、そのような経験を享受できる地物を知ろうとするものである。そのような、「ある経験を享受できる地物」を検索することを本研究の目的とする。

地物に関する検索において、Google Maps*などの地図ベースで情報検索を行う地理情報検索システムが普及している。地理情報検索システムでは、主に、地物を表す語が検索クエリとして用いられる。たとえば、「京都タワー」のような、具体的な地物名で検索すると、地図上に京都タワーの位置が表示される。「京都府京都市左京区吉田本町」のような、地物の住所も検索クエリとして一般的に用いられる。また、「喫茶店」や「居酒屋」のような、ある種の地物を表す語を検索クエリとして用いることも可能である。このような検索クエリの場合、検索結果としてクエリの語と関連が強い地物が比較的精度よく地図上に表示される。

従来の地理情報検索システムにおいては、「コーヒーを飲む」や「本を買う」、「安眠する」、「女性とちよつとお酒を飲む」というような経験の動詞句表現をクエリとして検索を行った場合、検索精度はあまり高くない。そのため、ある「経験」が可能な地物を検索したい場合、ユーザはその経験の動詞句表現を検索クエリとはせず、たとえば、「コーヒーを飲む」や「本を買う」などの経験を享受したい場合にはそれぞれ、「喫茶店」、「本屋」といった地物を表す名詞を思い浮かべ、これらの語を検索クエリとして用いる。実際、これらの検索を行った場合、それぞれの経験が可能な地物が検索結果として得られる。

一方で、「安眠する」、「女性とちよつとお酒を飲む」のような経験は、適切な地物を表す語を思い浮かべることが困難である。「眠る」や「お酒を飲む」という経験であれば、それぞれ「ホテル」、「居酒屋」などの、その経験を享受できる地物を表す語を思い浮かべることができる。しかし、すべての「ホテル」が「眠る」だけでなく「安眠する」ことに適しているわけではなく、すべての「居酒屋」が「女性とちよつとお酒を飲む」ことに適しているわけではない。このような、適切な地物を表す語を思い浮かべることが困難な場合、ユーザは地理情報検索システムに適切なクエリを入力することができないため、地物の検索をうまく行うことができない。

そこで本研究では、このような適切な地物を表す語を思い浮かべることが困難な経験であっても、それが可能な地物の検索を可能にすることを目的とする。たとえば、「女性とちよつとお酒を飲む」のような経験の動詞句表現を入力とし、たとえば「京都市 四条河原町」付近であれば、「Ace cafe」や「鱈」のような、その経験を享受できる地物を出力するという問題に取り組む。

本稿では特に、入力で扱う経験の動詞句表現について、「しつぽりお酒を飲む」や「女性とちよつとお酒を飲む」などの、「お酒を飲む」という経験に関係するものに焦点を当てる。また、出力である地物を、バーや居酒屋などの飲食店に限定する。

経験の動詞句表現を入力として、その経験を享受できる地物を出力するという問題に取り組むため、地物（飲食店）に対するユーザによるレビュー投稿サイト上のコンテンツに注目する。このようなレビュー投稿サイト上には、ある地物における、実際にその地物においてサービスを利用したユーザの経験に基づくレビューが投稿されている。そのようなレビューを解析することで、実際その地物においてサービスを利用してみないと知ることができない、経験財についての知識を得ることができる。たとえば、「しつぽり飲むのにおいしいお店です。」というような記述がレビュー中に存在すれば、そのレビューが投稿されている地物において、「しつぽり飲む」という経験が可能であると考えられる。このような、地物と経験の

[†] 京都大学, Kyoto University

[‡] 兵庫県立大学, University of Hyogo

*Google Maps: <https://maps.google.co.jp>

対応関係を十分に抽出することができれば、地物による経験の検索が可能となると考えられる。

しかし、既存のレビュー投稿サイトを用いるのみでは十分な地物と経験の対応関係を抽出することができない。まず、地物ごとに投稿されているレビュー数には大きなばらつきがあり、数百の投稿がなされているものから、数件の投稿しかされていないものまである。この場合、後者の地物に対する情報は不十分になってしまう。また、経験によってはレビュー内に記載が少ないようなものもある。この場合、そのような経験についての情報は不十分になってしまう。このように、既存のレビュー投稿サイトを用いるのみでは、地物と経験の対応関係を十分に抽出することができない。

そこで本研究では、経験の意味類似性に注目する。たとえば、「ちょっと飲む」と「少し飲む」という経験は、経験の意味そのものが類似するので、「ちょっと飲む」ことのできる地物においては、「少し飲む」こともできるし、逆に、「少し飲む」ことのできる地物では「ちょっと飲む」こともできると考えられる。この場合、2つの経験には意味類似性が存在するとする。一方で、「女性と飲む」と「おしゃれに飲む」という経験は、経験の意味そのものは類似していない。しかし、それぞれの経験を享受できる地物の状況は似ていると考えられる。このような場合においても、「女性と飲む」ことのできる地物においては、「おしゃれに飲む」ことができ、「おしゃれに飲む」ことのできる地物においては、「女性と飲む」ことができると考えられる。本研究では、このように、経験の意味そのものが類似しているものだけではなく、経験を享受できる地物の状況が類似する経験どうしにも意味類似性が存在するとする。このような経験の意味類似性を判定することによって、地物と経験の対応関係を補う。そうして得られた地物と経験の対応関係を用いることで、経験から地物の検索を実現する。

そこで本研究では、このような経験の意味類似性を判定する手法を提案する。経験の意味類似性を判定するため、レビュー中の経験を分散表現で表し、経験どうしの類似度を算出する。この際、レビュー中の経験に関する記述の周辺に記載された、その経験が可能な状況を表すような語の重みを大きくした上で経験を特徴づけ、経験を分散表現で表す。そうすることで、経験の意味そのものだけではなく、経験を享受できる地物の状況が似ているものも類似するとした、経験どうしの類似度を算出することができると考える。こうして得られた経験どうしの類似度を用いて、経験の意味類似性判定を行い、地物と経験の対応関係を補う。こうして、経験による地物の検索を実現する。

最後に、これまでレビューから得られた地物と経験の

対応関係と、経験の意味類似性判定を用いた経験による地物の検索手法について提案する。

本稿の構成は以下の通りである。2章では、関連研究を紹介する。3章では、経験の意味類似性判定の手法について説明する。4章では、経験の意味類似判定を用いた、経験価値を提供する地物の検索手法について説明する。5章では、実際にユーザに需要のある経験の収集手法について説明する。6章では、今後の課題について述べる。

2. 関連研究

内田ら [3] は、たとえばデジカメの重さや大きさ、形などの、使わなくても評価できる属性である探索的属性から、持ち運びやすさや耐久性など、使って初めて評価できる属性である経験的属性を推定し、経験的属性によるオブジェクトの検索を行った。使って初めて評価できるもので検索を行うという点では、サービスを利用して初めて分かる経験で検索を行う本研究と類似する。一方で、本研究では、レビュー中に記述のない経験について、その可否を推定するのではなく、経験どうしの類似性を判定することで補うという点で異なる。

倉島ら [4] は、人間の経験を(状況、行動、主観)からなる情報と定義し、文章中から(時間、空間、動作、対象、感情)を抽出し、人間の経験に関する興味深い知識を発見する手法を述べている。人間が主体となる動作、およびその対象を抽出し、マイニングすることで新たな知識を得ようとしている点で本研究と共通する部分は大きい。倉島らの研究では、データマイニング分野で提案されている、ルールの「興味深さ」指標を用いた相関ルール抽出を行い、抽出した経験の中から状況に特有な行動と感情を表現するルールを興味深いものとして発見しているのであり、経験を享受できる地物を検索可能にすることを目的とする本研究とは異なる。

経験間の関係を抽出する研究も多く行われている。

竹内ら [5] は、動詞がとる項の間の関係に動詞の意味を分解することによって多元的な意味記述を行う語彙概念構造 (lexical conceptual structure (LCS))[2] のもつ動詞間の包含関係に着目し、より詳細化した分類体系を構築することで、動詞項構造の語義を多元的に記述する動詞項構造ソーラスを提案した。

橋本ら [6] は、誰かが「離婚した」ならばその人は以前に「結婚した」はずである、というような含意関係が成立する動詞ペアを web から大規模に獲得するための類似度尺度を提案した。橋本らは、同じような文脈に現れる語は同じような意味を持つという分布類似度仮説に基づき、条件付確率を考慮した上で、類似度尺度を提案した。

本研究においては、経験そのものの意味だけではなく、経験が可能な地物の状況に着目することで経験間の類似性を判定している点で、前述の研究とは異なる。

3. 経験の分散表現による意味類似性判定

本章では、経験の分散表現による意味類似性判定の手法を述べる。本研究における、経験の意味類似性とは、「ちよつと飲む」と「少し飲む」のように、意味そのものが類似している経験間と、「女性と飲む」と「おしゃれに飲む」のように、経験を享受できる地物の状況が類似する経験間に存在するものである。このような経験の意味類似性を判定することによって、レビューから得られた地物と経験の対応関係を補う。そうして得られた地物と経験の対応関係を用いることで、経験から地物の検索を実現する。

そうした経験の意味類似性を判定するため、経験を、その周辺語で特徴づけた分散表現で表し、類似度を算出する。経験を分散表現で表す際、その経験が可能な地物の状況を表すような語の重みを大きくした上で経験を特徴づける。そうすることで、意味そのものだけではなく、経験が可能な地物の状況が似ている経験どうしも類似するとした類似度を算出することができると考えられる。

3.1 経験

本研究において、経験とは、動詞、動詞の目的語、少なくとも1つの修飾語句からなる組とし、経験 e 、動詞 v 、目的語 o 、少なくとも1つの要素を持つ修飾語句の集合 M を用いて以下のように表現する。

$$e = (v, o, M)$$

ここで、 o はなくともよいとする。 o がいない場合は次のように、 $Null$ と表す。

$$e = (v, Null, M)$$

たとえば、「安らかに眠る」、「女性とちよつとお酒を飲む」という経験はそれぞれ、

$$e = (\text{眠る}, Null, \{\text{安らかに}\})$$

$$e = (\text{飲む}, \text{お酒}, \{\text{女性と}, \text{ちよつと}\})$$

と表される。

3.2 経験の分散表現

本節では、経験の類似判定を行うため、経験を分散表現で表す手法について説明する。ここで、本稿では、「女性とちよつとお酒を飲む」のような「お酒を飲む」という経験に伴うようなもののみを扱う。よって、常に、 $o = \text{お$

酒、 $v = \text{飲む}$ である。また、このような経験は、「お酒を」という目的語がなくとも意味はほとんどの場合変わらない。たとえば、「女性とちよつとお酒を飲む」と「女性とちよつと飲む」、「おしゃれにお酒を飲む」と「おしゃれに飲む」という経験の意味はそれぞれ同じである。よって、これ以降は簡単のため、「女性とちよつとお酒を飲む」という経験を、「女性とちよつと飲む」のように、 $o = Null$ として表記する。

また、経験の類似性判定では、「女性と飲む」や「ちよつと飲む」のような経験を構成する修飾語句が1つのもののみ扱う。「女性とちよつと飲む」のような、複数の修飾語句からなる経験については、4.章の、経験を入力とし地物を実際に検索する際に対処する。

経験を分散表現にするため、 word2vec^\dagger を利用する。 word2vec は与えたテキストコーパスを学習し、各単語をその周辺語で特徴づけ分散表現にする。この分散表現を用いることで、各単語の意味を捉え、単語間の類似度を算出することができる。

しかし、本研究で扱う経験は1単語からなるものではない。たとえば、「女性と飲む」という経験は、 word2vec の学習のためにテキストコーパスを形態素解析した際、「女性」、「と」、「飲む」という3語に分けられる。そのため、レビューに対して単に word2vec を用いるのみでは「女性と飲む」という経験を分散表現で表すことはできない。

そこで本研究では、「女性」という語と「飲む」という語がレビュー中で近い位置に存在した場合、それぞれを「女性と飲む」という経験を表すような記号に置換する。このように、レビュー中にある1つの経験を構成する語が近い位置に存在する場合に、それらをその経験を表す記号に置換する。そうすることで、形態素解析により複数の語に分けられる経験についても、1つの語として学習を行い、分散表現で表すことが可能となる。

たとえば、以下の文がレビュー中に存在する場合を考える。

「女性と いい 雰囲気 の カウンター 席 で 飲む。」

この文は形態素解析により、以下のように単語ごとに分解される。

「女性/と/いい/雰囲気/の/カウンター/席/で/飲む」

ただし、“/”は単語の区切りを表す。ここで、「女性と飲む」という経験が「女性」と「飲む」という2つの語によって構成されると考える。 Ex_0 という記号が「女性と飲む」という経験を表すものとする。すると、上記の

[†]<https://code.google.com/archive/p/word2vec/>

表 1: 経験が可能な状況を表す語とその経験の例

経験が可能な状況を表す語	経験
カウンター	おしゃれに
飲み放題	大勢で
落ち着いた	しっぽりと
いい雰囲気	女性と
立ち飲み	ひとりで
始発	朝まで

文は,

「 Ex_0 /と/いい/雰囲気/の/カウンター/席/で/ Ex_0 」

のように置換される。このように、学習コーパスの文章中の経験をすべて記号で置換した後に、word2vecによる学習を行う。そうして得られた記号 Ex_0 の分散表現を、「女性と飲む」という経験の分散表現として扱う。

また、以下のように、同時に複数の経験を構成する語が文中に存在した場合を考える。ただし、「ちょっと」と「飲む」は2語で「ちょっと飲む」という経験を構成し、 Ex_1 という記号が「ちょっと飲む」という経験を表すものとする。

「女性/と/いい/雰囲気/の/カウンター/席/で
/ちょっと/飲む」

このような場合は、以下のように、複数の経験を構成する語は複数の経験を表す記号に置換する。

「 Ex_0 /と/いい/雰囲気/の/カウンター/席/で
/ Ex_1 / Ex_0 / Ex_1 」

以上の手法を学習コーパスであるレビューに適用することで、複数の語からなる経験を分散表現で表す。こうして得られた経験の分散表現を用いて、経験間の類似度を算出し、経験の類似性判定を行う。

3.3 経験が可能な状況を表す語

経験を分散表現にする際、経験が可能な状況を表す語の重みを大きくする。そうすることで、意味そのものだけでなく、経験が可能な地物の状況が似ている経験どうしも類似するとした類似度を算出することができると考えられる。経験が可能な状況を表す語の例を表??に示す。

本稿では、レビュー中で「飲む」という語に直接係っている語は、「お酒を飲む」に関係する経験が可能な状況を表す語であると考え、そのような語を抽出する。

抽出した語は3.2節で述べた、経験を分散表現で表す際に、重みを大きくして経験を特徴づけるようにする。

表 2: 予備実験に用いた経験のすべて

ちょっと飲む	さっと飲む	一杯飲む
ゆっくり飲む	しっぽり飲む	静かに飲む
わいわい飲む	楽しく飲む	一人で飲む
ひとりで飲む	二人で飲む	みんなで飲む
大勢で飲む	がつつり飲む	しこたま飲む
ジョッキで飲む	久しぶりに飲む	初めて飲む
おしゃれに飲む	オシャレに飲む	美味しく飲む
たくさん飲む	朝まで飲む	旅先で飲む
昼間に飲む	昼から飲む	深夜に飲む
友達と飲む	連れと飲む	彼女と飲む
上司と飲む	同僚と飲む	先輩と飲む
息子と飲む	サシで飲む	軽く飲む
ワイワイ飲む	独りで飲む	皆で飲む
安く飲む	いっぱい飲む	朝から飲む
女性と飲む	取引先と飲む	会社帰りに飲む
大人数で飲む		

表 3: 「ちょっと飲む」と類似する経験上位5件とその類似度

類似する経験	類似度
ゆっくり飲む	0.646
たくさん飲む	0.592
いっぱい飲む	0.582
女性と飲む	0.570
楽しく飲む	0.559

3.4 経験の類似性判定の予備実験

本節では、これまで述べた経験の類似性判定の手法に対して実施した予備実験について示す。

まず使用したのは表2に示す46件の経験である。また、word2vecにおける学習コーパスとしては、食べログのロコミ検索[‡]において、検索対象のエリアを「京都府」に設定し、上記46件の経験で検索して得られた9,769件のレビューを用いた。

その結果の一部を表3と表4に示す。それぞれ、「ちょっと飲む」及び「静かに飲む」という経験と類似する上位5件の経験とその類似度である。

4. 経験価値を提供する地物検索

本章では、経験価値を提供する地物の検索手法について述べる。経験の動詞句表現を入力として、その経験を享受できる地物を出力するという問題を解決することが本研究の目的である。そのような問題を解決するため、

[‡]食べログ ロコミ検索: <https://tabelog.com/rst/rvwlst/>

表 4: 「静かに飲む」と類似する経験上位 5 件とその類似度

類似する経験	類似度
ひとりで飲む	0.724
しっぽり飲む	0.687
おしゃれに飲む	0.670
ゆっくり飲む	0.660
オシャレに飲む	0.658

「経験-地物行列」という、ある経験がある地物において可能であるという、経験と地物の対応関係を集約した行列を作成する。十分な対応関係を持つそのような行列を作成することができれば、経験から地物を検索するという本研究の目的は達成されると考えられる。

そのような経験-地物行列の作成にあたり、地物に対してユーザがレビューを投稿するサイト上のコンテンツに注目する。このようなレビュー投稿サイト上には、ある地物における、ユーザが実際にその地物を利用した経験に基づくレビューが投稿されている。そのようなレビューを解析することで、そのレビューが投稿された地物においてどのような経験を享受することができるのかという情報を得る。

本章では、レビュー投稿サイトを利用して経験-地物行列を作成し、経験から地物の検索を行う手法について説明する。

4.1 地物及びレビューの収集

地物及びレビューは飲食店へのレビュー投稿サイトである食べログ[§]を用いる。このようなレビュー投稿サイト上のコンテンツには、ある地物における、ユーザが実際にその地物を利用して経験した内容に関する情報が記載されている。たとえば、レビュー中に「しっぽり飲むのにおいしいお店です。」という文があれば、そのレビューが投稿されている地物において、「しっぽり飲む」という経験が可能であることが分かる。そのようなレビュー投稿サイト上のレビューを収集し、レビューが投稿された地物において、どのような経験を享受することができるかを分析する。

食べログでは、飲食店が複数のジャンルに分けられている。[¶] その中でも本稿では、「お酒を飲む」という経験に伴う経験のみを扱うことから、以下のジャンルに分類されている飲食店を地物として収集する。

- レストラン > 居酒屋・ダイニングバー
- バー・お酒

[§]食べログ: <https://tabelog.com/>

[¶]食べログジャンル一覧: https://tabelog.com/cat_lst/

表 5: 経験-地物行列

	g_1	g_2	...	g_k	...	g_n
e_1	s_{11}	s_{12}	⋮
e_2	s_{21}	s_{22}	⋮
⋮	⋮
e_j	s_{jk}	...	⋮
⋮	⋮
e_m	s_{mn}

また、それらの飲食店に投稿されているレビューを収集する。

4.2 経験-地物行列の作成

本節では、5.2 節及び 4.1 節で述べた方法で収集した、経験、地物、地物について言及されたレビューを用いて、経験-地物行列を作成する方法について述べる。

経験-地物行列とは、ある経験がある地物において享受できるという情報を集約した行列である。この行列のある行は、ある 1 つの経験を表し、ある列は、ある 1 つの地物を表す。行列のそれぞれの成分の大きさは、その行が表す経験が、その列の表す地物において可能である度合いを表す。ここで、全経験集合を $E = \{e_1, e_2, \dots, e_m\}$ 、全地物集合を $G = \{g_1, g_2, \dots, g_n\}$ とすると、経験-地物行列は表 5 のように表される。ただし、 $e_j \in E, g_k \in G (j = 1, 2, \dots, m, k = 1, 2, \dots, n)$ である。行列中の成分 s_{jk} は、経験 e_j が地物 g_k において可能である度合いを表す。たとえば、経験-地物行列の一部に表 6 のような部分があるとする。この場合、経験 $e = (\text{飲む}, \text{Null}, \{\text{ちよつと}\})$ という経験が地物 $g = \text{Ace cafe}$ において可能な度合いが 4.0 である。

経験-地物行列の作成方法について説明する。まず、行列の各行が表す経験については、3. 章と同様に、経験を構成する修飾語が 1 つのもののみを扱う。

次に、行列の各列が表す地物については、4.1 で収集した地物を用いる。その中でも、経験-地物行列の作成には、1 つ以上のレビューが投稿されている地物のみを用いる。

行列の各成分の値については、それぞれの地物に投稿されたユーザのレビュー中で、それぞれの経験について言及されているレビューの頻度を計測することで定める。行列中の成分 s_{jk} の値は、地物 g_k に投稿されているレビューのうち、レビュー文章中で経験 e_j についての記述されているレビューの数とする。ここで、経験 e_j についての記述がされているとは、レビュー文章中に、経験 e_j の修飾語 m と動詞 $v = \text{飲む}$ という単語がともに含まれ

表 6: 経験-地物行列の例

	...	鱈	A c e c a f e	鳥 貴 族 木 屋 町 店	...
⋮
(飲む, Null, {ちよつと})	...	1.0	4.0	2.0	...
(飲む, Null, {女性と})	...	2.0	3.0	0.0	...
⋮

ていることを指す。たとえば、地物 $g = \text{Ace cafe}$ に投稿されているレビューのうち、経験 $e = (\text{飲む}, \text{Null}, \{\text{ちよつと}\})$ という経験を含む、すなわち、「ちよつと」と「飲む」という単語をともに含むレビューが4つであれば、表6の例ようになる。

以上のように、収集した経験、地物、及びレビューを用いて、経験-地物行列を作成する。

こうして作成した経験-地物行列を用いて、経験 e_j を享受できる地物を検索する際には、 $s_{jk} > 0$ となるような地物 g_k を、 $s_{jk} > 0$ の大きい順にランキングして出力する。たとえば、表6で値の記載がある 2×3 の部分が経験-地物行列の全体であるとする、「女性と飲む」という経験を享受できる地物を検索する場合、入力 $e_j = (\text{飲む}, \text{Null}, \{\text{女性と}\})$ であり、出力はランキング上位から、 $g = \text{Ace cafe}$ 、 $g = \text{鱈}$ となる。

しかし、現状の経験-地物行列はスパースであるという問題がある。地物ごとに投稿されているレビュー数にはおおきなばらつきがあるため、ある地物を表す列には値があまり入っていないということが起きる。また、経験によってはレビュー内に記載が少ないようなものもある。そのような経験を表す行には値があまり入らない。上記の理由から、現在の経験-地物行列はスパースである。

そのため、現在の経験-地物行列を用いた、経験による地物の検索は十分ではない。この問題を解決するため、3章で行った、経験の類似性判定を用いて、経験と地物の対応関係を補う。

4.3 経験の類似性判定による行列の拡張

3章で行った、経験の類似性判定を利用して、経験と地物の対応関係を補うことで前節で作成した経験-地物行列を拡張する。そうすることで、行列がスパースであるという問題を解消し、十分な、経験による地物の検索を行おうと試みる。

前節で作成した経験-地物行列の j 行目の成分を表す

ベクトルを

$$\vec{e}_j = (s_{j1}, s_{j2}, \dots, s_{jk}, \dots, s_{jn})$$

と表す。ここで、行列の j 行目が表す経験と類似度が高い上位3つの経験が表される行の成分を表すベクトル e_{j1}, e_{j2}, e_{j3} について、以下のような計算を行う。

$$\vec{e}_j' = \text{sim}_{j1} * e_{j1} + \text{sim}_{j2} * e_{j2} + \text{sim}_{j3} * e_{j3}$$

ただし、 $\text{sim}_{j1}, \text{sim}_{j2}, \text{sim}_{j3}$ は、行列の j 行目が表す経験と類似度が高い上位3つの経験との類似度である。この処理をすべての行に対して行い得られた $\vec{e}_j' (j = 1, 2, \dots, m)$ を用いて経験-地物行列を再作成する。

こうして新たに得られた経験-地物行列を用いて、経験による地物の検索を行う。

また、前述のように経験どうしの類似性を判定して、地物と経験の対応関係を補うというのは、情報検索の分野でよく用いられる、LSA(Latent Semantic Analysis; 潜在意味解析)[1]などの次元圧縮の手法に近い。レビューから得られた地物と経験の対応関係を数値化し、各行が経験を、各列が地物を表すような行列で表した場合、スパースな行列が作成される。この行列に対して、LSA等の次元圧縮の手法を適用することで、実質的に似た経験どうしをクラスタリングして扱うことができる。こうすることで、地物と経験の対応関係を補い、経験による地物の検索が実現される。このような次元圧縮の手法を用いて、地物と経験の対応関係を補う手法についても実験、及び評価を行い経験の類似性判定による手法との比較を行う。

5. 経験の収集

本章では、実際にユーザが検索したいという需要のある経験を、CQAコーパスより抽出する手法について説明する。

特に本稿では、「しつぱりお酒を飲む」や「女性とちよつとお酒を飲む」などの、「お酒を飲む」という経験に関するようなものに焦点を当てる。また、出力である地物を、バーや居酒屋などの飲食店に限定する。

そのようなユーザに検索の需要がある経験を抽出するため、CQAコーパスを利用する。CQAコーパスには、「四条河原町で女性とちよつと飲むのいい場所を知りませんか?」というような、ある経験を享受できる地物を尋ねているような質問と、それに対する回答が多数存在する。そのような質問を分析することで、実際にユーザが享受できる地物を求めているような経験を抽出する。

表 7: 経験の享受できる地物を尋ねる質問文の例

- (1) 彼女と楽しく飲める居酒屋があれば教えてください。
- (2) 渋谷で静かに飲める場所ありますか？
- (3) 新宿か下北沢で女一人でも飲める場所ありますか？
- (4) 2人で落ち着いて飲める、個室の居酒屋を探しています。

5.1 使用する質問

本研究では CQA コーパスとして Yahoo!知恵袋^{||}を利用する。Yahoo!知恵袋ではそれぞれの質問がカテゴリごとに分類されている^{**}。その中でも本稿では、以下に示すカテゴリに分類されている質問を対象に、パターンマッチングによる経験の抽出を行う。

- 暮らしと生活ガイド > 料理
- レシピ > 飲食店

5.2 経験の抽出

本節では、CQA コーパスを利用した経験の抽出方法について述べる。このような CQA コーパスには、「四条河原町で女性とちょっと飲むのいい場所を知りませんか？」というような、ある経験を享受できる地物を尋ねているような文を含む質問文章が多数存在する。そのような質問文章中の質問文の中で、「お酒を飲む」という経験に伴うような経験の享受できる地物を尋ねているものをいくつか表 7 に例示する。このような質問文を分析することで、実際にユーザが経験を享受できる地物を求めているような、経験を抽出する。

本手法では、CQA 上の質問文章から経験を抽出するため、対象となる質問文章に対し以下のような前処理を施す。まず、質問文章を読点や記号などの区切りとなる位置で文に分割する。次に、それぞれの文に対して形態素解析と係り受け解析を行う。こうして、各文は閉路のない有向グラフで表現される。この有向グラフの各ノードは文の各文節に対応しており、ノードは複数の単語のシーケンスで構成される。また、各エッジは文節の係り受け関係に対応している。さらに、有向グラフを複数の有向道に分解することによって、1 文を複数の有向道の集合で表現する。この処理で得られた各有向道とあらかじめ人手で用意しておいた抽出パターンとのマッチングを行うことで、質問から経験を抽出する。

質問文章からの経験の抽出パターンを正規表現で表したもののすべてを表 8 に列挙する。パターン中の $[a/b]$

表 8: 経験の抽出パターンと抽出される経験

抽出パターン	抽出される経験
$[x/\text{名詞}][\text{と}/\text{助詞}][\text{飲む}/\text{動詞}]$	$\{x \text{ と} \}$
$[x/\text{名詞}][\text{で}/\text{助詞}][\text{飲む}/\text{動詞}]$	$\{x \text{ で} \}$
$[x/\text{名詞}][\text{に}/\text{助詞}][\text{飲む}/\text{動詞}]$	$\{x \text{ に} \}$
$[x/\text{動詞}][\text{て}/\text{助詞}][\text{飲む}/\text{動詞}]$	$\{x \text{ て} \}$
$[x/\text{形容詞}][\text{飲む}/\text{動詞}]$	$\{x \}$
$[x/\text{副詞}][\text{飲む}/\text{動詞}]$	$\{x \}$
$[x/\text{名詞}][\text{と}/\text{助詞}][\text{飲める}/\text{動詞}]$	$\{x \text{ と} \}$
$[x/\text{名詞}][\text{で}/\text{助詞}][\text{飲める}/\text{動詞}]$	$\{x \text{ で} \}$
$[x/\text{名詞}][\text{に}/\text{助詞}][\text{飲める}/\text{動詞}]$	$\{x \text{ に} \}$
$[x/\text{動詞}][\text{て}/\text{助詞}][\text{飲める}/\text{動詞}]$	$\{x \text{ て} \}$
$[x/\text{形容詞}][\text{飲める}/\text{動詞}]$	$\{x \}$
$[x/\text{副詞}][\text{飲める}/\text{動詞}]$	$\{x \}$

表 9: 抽出されたの経験の例

- (1) $e = \{ \text{彼女と, 楽しく} \}$
- (2) $e = \{ \text{渋谷で, 静かに} \}$
- (3) $e = \{ \text{新宿か, 下北沢で, 女一人で} \}$
- (4) $e = \{ \text{二人で, 落ち着いて} \}$

という表現は、「b という品詞の a という語」を表している。たとえば、「飲む/動詞」という表現は、「動詞である“飲む”という語」を表す。パターン中の変数 x に当てはまる単語は経験の修飾語として抽出される。

表 7 の例に対して、この経験の抽出パターンを適用した場合に抽出される経験を表 9 に示す。

6. 今後の課題

本稿では、「女性とお酒を飲む」のような経験を動詞句で表現したものを入力とし、その経験ができる地物を出力するという問題に取り組んだ。そのような問題を解決するため、経験の分散表現による意味類似性判定と経験価値を提供する地物検索という手法を提案した。現在、予備実験のみしか実施できていない状況であるため、今後本格的な実験へと取り組んでいくつもりである。

謝辞

本研究は JSPS 科研費 16H02906, 15H01718, 25240050 の助成を受けたものです。また、本研究の実施にあたり、ヤフー株式会社より提供いただいた Yahoo!知恵袋のデータを利用しました。ここに記して謝意を表します。

^{||} Yahoo!知恵袋: <https://chiebukuro.yahoo.co.jp/>

^{**} Yahoo!知恵袋 カテゴリ一覧: https://chiebukuro.yahoo.co.jp/dir/dir_list.php

参考文献

- [1] S. Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. Indexing by latent semantic analysis. *JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE*, Vol. 41, No. 6, pp. 391–407, 1990.
- [2] R. Jackendoff. *Semantic structures*, Vol. 18. MIT press, 1992.
- [3] Shinryo Uchida and Katsumi Tanaka. Object ranking by learning and inferring pairwise preferences from user reviews. *Master Thesis of Kyoto University*, 2017.
- [4] 倉島健, 藤村考, 奥田英範. 大規模テキストからの経験マイニング. *電子情報通信学会論文誌*, Vol. 92, No. 3, pp. 301–310, 2009.
- [5] 竹内孔一. 意味の包含関係に基づく動詞項構造の細分類. *言語処理学会年次大会論文集*, pp. 1037–1040, 2008.
- [6] 橋本力, 鳥澤健太郎, 黒田航, デサーガステイン, 村田真樹, 風間淳一. WWW からの大規模動詞含意知識の獲得. *情報処理学会論文誌*, Vol. 52, No. 1, pp. 293–307, 2011.