HappyRec:幸せ推薦システムの実装及び評価手法の検討

鈴木 健太[†] 田中涼太郎^{††} Panote Siriaraya^{†††} 栗 達^{††††} 中島 伸介^{††††}

†京都産業大学大学院 先端情報学研究科 〒 603-8555 京都府京都市北区上賀茂本山 ††京都産業大学 コンピュータ理工学部 〒 603-8555 京都府京都市北区上賀茂本山 †††京都工芸繊維大学情報工学・人間科学系 〒 606-8585 京都市左京区松ヶ崎橋上町 †††京都産業大学情報理工学部 〒 603-8555 京都府京都市北区上賀茂本山

E-mail: \dagger i2086060@cc.kyoto-su.ac.jp, \dagger †g1744717@cc.kyoto-su.ac.jp, \dagger † †spanote@kit.ac.jp, \dagger † † \dagger {lida,nakajima}@cc.kyoto-su.ac.jp

あらまし 心理学者 Seligman らが提唱するポジティブ心理学において、毎晩その日の幸せな出来事を3つ記録することにより幸福度の向上と抑うつ度の低下が起こるという研究が報告されている。しかし、日常生活で自覚できる幸せな出来事は決して多くはない。そこで、幸せな出来事を推薦することで、経験する幸せの回数を増やし、幸福度向上、抑うつ度の低下を狙えるのではないかと考えた。本稿では幸せ推薦システムの構築方法とシステム内で行う協調フィルタリングの予備実験、および提案システムに基づく評価実験結果について報告する。

キーワード 幸せ推薦システム, Web マイニング,幸福度,ウェルビーイング,地理情報, HappyDB

1. はじめに

心理学者 Seligman らが提唱するポジティブ心理学 [1] におい て、three good things という、毎晩その日の幸せな出来事を 3 つ記録するエクササイズが存在する. このエクササイズを行 うことによって,幸福度の向上と抑うつ度の低下が起こると報 告[2]されている.この研究に関して、関沢らの研究[3]で日本 人被験者による実験を用いて、同様の結果が報告されている. しかし、このエクササイズでは自力でその日の幸せな出来事を 思い出す必要があるが、自尊心の低い人やうつ病の人は普段の 生活の中で幸せな出来事を自覚することは容易ではない. 図1 は厚生労働省が公開している患者調査[4]の結果で、うつ病患 者数の推移を示している.調査が開始された平成8年から平成 28年で3倍ほどになっていることが分かる. これらから,幸 せな出来事を推薦することで,経験する幸せの回数を増やし, 幸福度の向上、抑うつ度の低下を狙えるのではないかと考え、 幸せ推薦システムの開発を開始した. 本稿では幸せ推薦システ ムの構築方法とシステム内で行う協調フィルタリングの予備実 験,提案システムの評価実験の計画について説明する.

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章では本稿で提案するシステムの概要を解説する。4章では協調フィルタリングを使用した幸せ情報の追加についての予備実験と結果について報告する。5章では幸せスポット推薦の評価実験について説明し、最後に6章でまとめを記述する。

2. 関連研究

本稿で提案するシステムは、心理学者の Seligman らが提唱するポジティブ心理学 [1] から着想している。この Seligman らの研究では、three good things という毎晩寝る前にその日あった幸せな出来事を 3 つその理由とともに書くエクササイズがある。このエクササイズを行うことで幸福度の向上と抑うつ

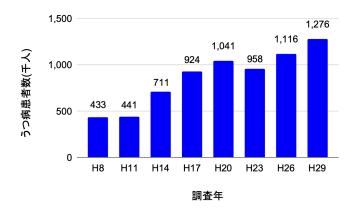


図 1 うつ病 (躁鬱病含む) 患者数の推移 (日本国内)

度の低下が起き、この効果の持続性を報告[2]している。また、Seligman らの実験に関して、関沢ら[3]は日本人の被験者で同様の実験を行い、幸福度が向上することを報告している。これらの研究では日頃起きる幸せな出来事を認知して記録する必要があるが、現実は難しい。そのため、本稿で提案する幸せ推薦システムによって身近で経験できる幸せな出来事を推薦することは有用だと考える。

幸せに関しての研究としては、幸せの瞬間を収集しデータベース化する試みとして Asai らの HappyDB [5] がある. これは欧米人の幸せな瞬間をクラウドソーシングにより約 10 万件のテキストデータを収集したものであり、人々がどのような瞬間に幸せを感じるのかを分析する上で貴重なデータである. 本稿では、このデータセットを使用して提案システムの構築を行う. 詳細は3章で説明する.

また、ユーザの感情を重視した情報推薦技術の研究として、Quercia らの研究[6] が存在する. この研究はソーシャルデータ (Flicker のメタ付き情報)を使用し、最短経路ではなく、ユー

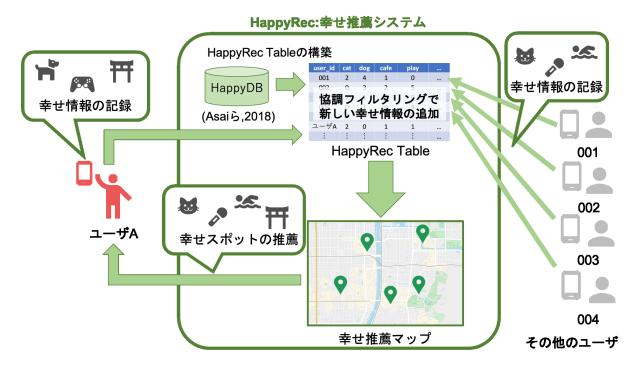


図 2 幸せ推薦システムの概要

ザが静かで、美しく、幸せに感じるであろうルートを推薦するというものである.しかし、この研究では、美しさ、静かさなどの全てのユーザに当てはまる感情をもとに推薦を行っている.本稿で提案する手法では、ユーザそれぞれに特化した推薦が可能になると考えられる.

これら以外にも、Yin らの研究[7]では、位置情報に注目した推薦システムを提案している。この研究では、ユーザの活動履歴を用いて、ユーザの興味のあることを抽出し、これをもとにイベントや場所の推薦を行っている。また、ユーザの興味だけではなく、ユーザにとって新しい都市を推薦する際は、その場所特有のイベントや場所を考慮することによって推薦の精度向上を図っている。この研究ではユーザの行動履歴から興味のあるものを推薦しているが、本稿で提案する手法ではユーザの幸せな体験の情報を用いることは斬新な視点だと考えている。

3. 幸せ推薦システムの開発

本章では、3.1 節で幸せ推薦システムの概要、3.2 節で HappyRec Table の作成について、3.3 節で幸せな体験ができる場所の推薦方法について説明する.

3.1 幸せ推薦システムの概要

本節では幸せ推薦システムの概要を図2を用いて説明する. 図左のユーザ A が幸せなスポットを推薦されるユーザである. 提案システムでは、利用ユーザが日頃感じる幸せな出来事を日記感覚でシステムに記録する. システムは投稿された幸せな情報を HappyRec Table として記録する. この記録された情報を使用して協調フィルタリングを用いることで、まだ経験したことのない未知の幸せな情報を追加する. この未知の情報も追加された HappyRec Table と類似度の高いスポットをユーザが幸せな体験のできるスポットとして提示する.

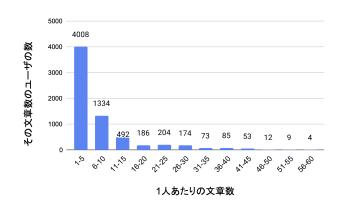


図 3 HappyDB の 1 人あたりの幸せな出来事の記録回数

3.2 HappyRec Table の作成

提案システムで使用する HappyRec Table の構築方法を説明する. 3.1 節で書いた通り、提案システムでは HappyRec Table を用いて協調フィルタリングで新しい幸せ情報の推薦を試みる. そのため、実際の利用ユーザの情報のみで構築した場合、新たな幸せ情報をユーザに推薦することが困難になるコールドスタート問題が生じることが考えられる. これを解決するため、予め HappyRec Table に一定量の幸せ情報を登録しておく必要がある. 今回は Asai らの公開している HappyDB [5] を使用する. HappyRec Table の構築の前に HappyDB について説明する.

HappyDB はクラウドソーシングによって集められた欧米人の幸せな瞬間を説明したテキストデータから構成されている. データは約1万人から集められ,全部で約10万件の文章が記録されている. 文章ごとに,その文章の幸せな出来事が,記録作業を行った時間から24時間以内のものか,3ヶ月以内のもの

表 1 HappyRec Table

			1.0		
userid	cat	dog	cafe	play	
001	2	4	1	0	
002	2	0	1	1	
003	4	3	1	4	
:	:	:	:	:	
Suzuki	0	3	2	5	
:	:	:	:	:	

_	user_id	cat	dog	cafe	play	
比較対象の ユーザ	001	2	(4)	3	0	
_	002	2	•	1	1	
_	÷	:		÷	÷	
推薦対象の ユーザ	Suzuki	3	₩ 4	3	1	
_	Tanaka	0	0	0	1	
	÷	:	:	÷	÷	

図 4 未知の幸せ単語の追加

かといった情報も付いている.

HappyRec Table の構築には HappyDB に入っている 24 時 間以内の幸せについて説明した文章のみ使用する. 直近3ヶ月 以内の幸せについてのデータではなく,24時間以内のデータの みを使用した理由は、直近3ヶ月以内の幸せとして記録されて いるものは、ユーザにとって長期間忘れない大きな出来事が多 く記録されており、直近24時間以内の幸せに関しては、その日 の出来事であり,毎日起きる小さな幸せが記録されていると考 えたためだ、HappyDBには24時間以内の幸せについて説明 した文章のデータは 6,634 人, 49,831 文章存在する. このデー タの内, ユーザ1人あたりの幸せな出来事の記録回数を集計し たところ、図3のようになった. 記録回数が少ないユーザが多 いことが分かる. 提案システムでは毎晩記録してもらうことを 想定しており, 1人あたりのデータ量が多くなることが想定さ れるため、HappyDB から使用するデータでは1人で30文章 以上記録している309人のデータを使用することにした. 文章 数としては 11,311 文章になる.

次に HappyRec Table について説明する. HappyRec Table の例を表 1 に示す。これは,ユーザの識別 ID とユーザが投稿した幸せの単語ごとの出現回数を記録したデータであり,表 1 の時は,ユーザ 001 の幸せ記録に cat が 2 回,dog が 4 回,cafe が 1 回出現し,play は出現していないことを示している.

構築方法としては、上で説明した HappyDB のテキストデータを文章ごとに形態素解析し、前置詞や冠詞などの stopwords を 抜いた単語の出現回数とユーザの識別 ID を使用して HappyRec Table を作成する.

3.3 幸せな体験のできる場所推薦

3.1 節, 3.2 節で作成した HappyRec Table を用いて,ユーザが幸せな体験ができる場所の推薦方法について説明する.

推薦には一部、協調フィルタリングを用いる. 具体的な方法 は図 4 を用いて説明する. 図中のユーザ $ID001\sim002$ のデータは HappyRec Table 構築時に使用した HappyDB のユーザで

表 2 正規化無しデータの上位 10 件に対する予測結果

ユーザ ID	単語	正解值	SVD 予測	SlopeOne 予測
230	life	100	4.09	1.87
120	event	52	1.73	0.46
6	Yesterday	36	1.21	0.71
74	life	36	4.42	2.07
40	home	33	2.59	1.29
39	Yesterday	32	2.03	0.75
40	Yesterday	32	2.56	0.89
53	research	32	3.2	0.19
132	life	32	1.63	2.08
74	moment	31	6.76	1.76

表 3 正規化無しの予測精度

	SVD	SlopeOne
RMSE	43.546	45.160
MAE	38.578	40.393

あり、Suzuki は実際の利用ユーザで、今回幸せな体験のできる場所を推薦されるユーザとする。Suzuki が既に記録している単語のみを使用した推薦では、新しい幸せの推薦ができなく、過去に記録したことのない幸せも推薦可能にするために協調フィルタリングを使用する。推薦されるユーザである Suzuki と類似度の高いユーザの幸せ情報の中から、Suzuki に存在しない項目を新たに Suzuki の幸せに関連の深い単語として、既に記録された単語に追加して推薦を行う。この作業には Python のライブラリである Surprise [9] を使用する。

以上の過程では、HappyRec Table を用いてユーザの幸せと 関連の深い単語の推薦までを行った、次に、この幸せに関連の ある単語とユーザの周辺のスポット情報を使用して幸せを感じ るであろうスポットの推薦方法を説明する.

ユーザの現在地周辺のスポットを OpenStreetMap [10] や Google マップといったサービスから複数箇所選ぶ. このスポットごとに付与されているスポットのカテゴリや口コミ情報などの単語を Word2vec を用いてベクトル化する. 同様にユーザの幸せに関連の深い単語を Word2vec を用いてベクトル化する. この際スポットごと, ユーザごとに単語を足し合わせることでそれぞれ1つのベクトルになるようにする. ユーザのベクトルと周辺のスポットごとのベクトルとの間でコサイン類似度を求め,類似度の高いスポット上位数件をユーザが幸せを感じるスポットとしてマップ上に表示して推薦を行う.

4. 予備実験

3.3節で説明した過去に記録したことのない幸せを協調フィルタリングで追加する処理について,3.2節で作成した HappyRec Table を用いて予備実験を行った.具体的には,HappyRec Table の値が大きい順に 10 箇所の値をそれぞれ,0 に変更して未知の値にし,これらの値の予測を SVD と SlopeOne を用いて行った.この予測には Python のライブラリである Surprise [9] を使用した.

今回は、HappyRec Table の値をそのまま使用したものと、

表 4 正規化ありデータの上位 10 件に対する予測結果

ユーザ ID	単語	正解值	SVD 予測	SlopeOne 予測
230	life	1	0.02	0
120	event	0.52	0	0
6	Yesterday	0.36	0.01	0
74	life	0.36	0.02	0
40	home	0.33	0.01	0
39	Yesterday	0.32	0.01	0
40	Yesterday	0.32	0.01	0
53	research	0.32	0	0
132	life	0.32	0.02	0
74	moment	0.31	0.02	0

表 5 正規化ありの予測精度

	P + 0 ===//4/G - /	> - 4 0/0/10/24
	SVD	SlopeOne
RMSE	0.452	0.463
MAE	0.404	0.416

HappyRec Table の最小値と最大値を使用して正規化した2種類のデータで予測作業を行った.正規化無しのデータでの予測結果を表2と表3に示す.正規化したデータでの予測結果を表4と表5に示す.

これらの結果から、予測精度が良くないことが分かる.しかし、本稿で提案するシステムでは3.3節で説明した通り、幸せな体験のできるスポットを推薦することを目指している.そのため、単語の予測性能より最終的に推薦するスポットの精度が重要なため、単語の予測性能の向上は今後の課題とし、本稿では3章で説明した手法での推薦を実施し性能の評価を行うことにした.

5. 評価実験

5.1 評価実験の概要

3章で説明した幸せ推薦システムが実際にユーザが幸せを感じるスポットを推薦することが可能かどうかを検証する被験者 実験を行う.

大学生 9 人の被験者に 2 週間,毎晩寝る前にその日あった幸せな出来事を 3 つ記録してもらった.記録の方法は,先行研究 [8] で開発しているスマートフォンアプリと Google スプレッドシートを使用して行った.この記録で収集された幸せな出来事を記録したテキストデータを使用して被験者ごとに幸せを感じるであろうスポットの推薦を実施し,nDCG を用いて評価する.

5.2 評価実験の流れ

最終的な評価の nDCG で使用する正解のランキングを作成するために、独自で作成した京都市内の 20 スポットを幸せに感じる順番に並び替えてもらうアンケートを実施した。このアンケートは各スポットごとにスポットの名称、Google マップのストリートビューへのリンク、スポットのジャンルを記載することで、どういった場所かを被験者に分かりやすく示した。このアンケートを2週間の実験の初日に回答してもらう。このアンケートで収集されたランキング情報をnDCG の正解ランキ



図 5 日本語データでの WordCloud



図 6 英語データでの WordCloud

ングとして使用する.

正解ランキングとは別に、被験者が2週間記録した幸せな出来事の情報を3章で説明した提案システムで分析することで、被験者ごとに幸せを感じるであろうスポットの予想ランキングを作成する.これらのランキング情報を使用してnDCGを求めて提案システムの評価を行う.

5.3 収集された幸せ情報について

被験者 9 人から集まった幸せな出来事を説明したテキストデータについて説明する。被験者には 2 週間毎晩寝る前にその日あった幸せな出来事を、出来事の説明と、それがなぜ幸せに感じたのかの説明のペアを 3 つ記録してもらった。日によって記録を忘れることもあったため、2 週間で合計 347 セットの記録が集まった。集まった文章データから WordCloud を作成したものを図 5 に示す。また、提案システムでは英語に翻訳した状態でも試すため収集された幸せ情報を Google 翻訳にかけ、翻訳語のデータでの WordCloud を図 6 に示す。

5.4 幸せを感じるスポットの推薦結果

収集したデータを用いて 3.3 節で説明した手法で実際に被験者へのスポットの推薦を実施した. 推薦は、日本語データとGoogle 翻訳で英訳したデータの 2 種類で行い、日本語データでは形態素解析器に MeCab を使用し、単語の分散表現の取得に

表 6 日本語データでの各手法ごとの nDCG の平均値

X 0 L7 L 7 COL 1 ACCO LOCK OF TALE				
手法	Top5	Top10	Top20	Top30
1 14	単語使用	単語使用	単語使用	単語使用
SVD	0.793	0.789	0.775	0.776
(変換無し)	0.133	0.103	0.110	
SlopeOne	0.802	0.781	0.776	0.775
(変換無し)	0.802	0.761	0.770	0.115
SVD	0.780	0.776	0.777	0.781
(最大値で正規化)	0.760	0.776	0.777	0.781
SlopeOne	0.798	0.801	0.776	0.774
(最大値で正規化)	0.796	0.601	0.770	0.114
SVD	0.796	0.787	0.778	0.774
(z 値で正規化)	0.790	0.767	0.116	0.114
SlopeOne	0.790	0.796	0.782	0.776
(z 値で正規化)	0.790	0.790	0.762	0.776
SVD	0.803	0.811	0.798	0.796
(tfidf に変換)	0.603	0.811	0.798	0.796
SlopeOne	0.786	0.702	0.774	0.774
(tfidf に変換)	0.760	0.783		

表 7 英語データでの各手法ごとの nDCG の平均値

表 7 英語アーダでの各手法ことの nDCG の平均値					
手法	Top5	Top10	Top20	Top30	
丁 伍	単語使用	単語使用	単語使用	単語使用	
SVD	0.837	0.837 0.851		0.858	
(変換無し)	0.001	0.651	0.850	0.696	
SlopeOne	0.839	0.055	0.848	0.853	
(変換無し)	0.039	0.855	0.040		
SVD	0.889	0.857	0.860	0.861	
(最大値で正規化)	0.009				
SlopeOne	0.883	0.853	0.864	0.856	
(最大値で正規化)	0.000	0.000	0.004	0.000	
SVD	0.870	0.865	0.852	0.852	
(z 値で正規化)	0.070	0.000	0.002	0.602	
SlopeOne	0.856	0.856	0.852	0.858	
(z 値で正規化)	0.000	0.000	0.002	0.000	
SVD	0.856	0.866	0.865	0.869	
(tfidf に変換)	0.000	0.800	0.005	0.809	
SlopeOne	0.823	0.854	0.851	0.848	
(tfidf に変換)	0.623	0.004	0.001	0.040	

表 8 協調フィルタリングを使用しない結果

TO MUMO 1100	/ C (X/11 0 6 1 / MI/N
手法	nDCG
日本語 (協調フィルタリング無し)	0.777
英語 (協調フィルタリング無し)	0.857



図7 日本語データの各手法の結果の比較

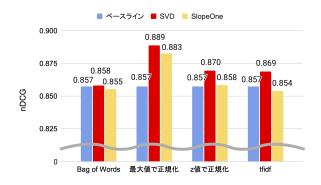


図 8 英語データの各手法の結果の比較

は Word2vec を用い,Word2vec のモデルは白ヤギコーポレーションが公開しているもの使用した.英語データでは形態素解析器には nltk を使用し,単語の分散表現の取得には Word2vec を用い,Word2vec のモデルは Google ニュースで学習されたものを使用した.日本語,英語のデータで単語の出現回数を記録した HappyRec Table を作成し,作成した HappyRec Table から最大値を用いた正規化,標準偏差に基づく z 値を用いた正規化,また,(1) 式 (2) 式を用いた,通常の tfidf 算出の文書頻度をユーザ頻度とした値のそれぞれに変換した HappyRec Table を作成した.

$$idf(t_i) = log(\frac{総ユーザ数}{単語 t_i を記述したユーザ数 + 1})$$
 (2)

作成したそれぞれの HappyRec Table で Surprise を使用し SVD と SlopeOne による協調フィルタリングを実施,協調フィルタリングによるスコアが高い順に 5 単語,10 単語,20 単語,30 単語の合計ベクトルをそれぞれ求め,これらを被験者のベクトルとしてスポットとの類似度を求めて類似度の高い順に被験者が幸せを感じるであろうスポットのランキングを作成した。作成したランキングと 5.2 節で作成した被験者による正解ランキングの間で nDCG を求めた。日本語データで実施したnDCG の結果を表 6 に示し,それぞれの手法で一番良い結果をまとめたものを図 7 に示す。英語データで実施した nDCG の結果を表 7 に示し。それぞれの手法で一番良い結果をまとめたものを図 8 に示す。また,ベースラインとして協調フィルタリ

ングを使用せず、被験者ごとの出現単語全ての合計ベクトルを 被験者のベクトルとする手法を行い、この結果を表8に示す.

図7と図8から、ベースラインである協調フィルタリングを行わない推薦結果と比べ、HappyRec Table を正規化などで変換し、協調フィルタリングを行った推薦方法の方が優れた幸せスポットランキングを作成することが可能だということが分かった。この理由としては、ベースラインでは幸せ記録の単語を全て使用してスポットとの類似度を求めており、被験者の幸せな出来事と関係の浅い単語も使用している。しかし、協調フィルタリングを利用する提案手法では、協調フィルタリングのスコアの上位単語のみを使用してスポットとの類似度を求める。これにより、被験者の幸せと関係の深い単語のみを利用することが可能となり、被験者が幸せに感じるスポットを的確に見つけることができたと考えられる。

また、図7と図8から日本語データの時は単語の出現回数を記録した HappyRec Table を tfidf で変換し、SVD で協調フィルタリングを行った結果がより良くなり、英語データの時は最大値で正規化し、SVD で協調フィルタリングを行った結果がより良くなったことが分かる。

6. おわりに

本稿では、ユーザが幸せな体験ができる場所を推薦する方法について説明し、提案システム内で行う協調フィルタリングをHappyDBのデータで実施し精度を確認した。また、2週間の実験を行い、提案システムで作成するスポットのランキングの評価をnDCGを用いて実施した。

今後は、提案システムで作成するランキング精度の向上と、より気軽に使用可能にするために、スマートフォンアプリの開発、より大人数での実験での評価を考えている.

謝辞

本研究の一部は,科研費基盤研究(B)(課題番号:17H01822,19H04118,20H04293)による.ここに記して謝意を表す.

文献

- Seligman, M.E.P. (2002). Authentic Happiness: Using the New Positive Psychology to Realize Your Potential for Lasting Fulfillment. The Free Press.
- [2] Martin E. P. Seligman and Tracy A. Steen, Nansook Park, Christopher Peterson. Positive psychology progress: Empirical validation of interventions, American Psychologist (60,410–421), 2005.
- [3] 関沢 洋一, 吉武 尚美. 良いことを毎日 3 つ書くと幸せになれるか?, RIETI Discussion Paper Series, 2013 年 11 月.
- [4] 厚生労働省,患者調査(https://www.mhlw.go.jp/toukei/list/10-20.html)
- [5] Akari Asai, Sara Evensen, Behzad Golshan, Alon Halevy, Vivian Li, Andrei Lopatenko, Daniela Stepanov, Yoshihiko Suhara, Wang-Chiew Tan, Yinzhan Xu, HappyDB: A Corpus of 100,000 Crowdsourced Happy Moments, 2018.
- [6] Quercia, D., Schifanella, R., and Aiello, L. M. (2014, September). The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. In Proceedings of the 25th ACM conference on Hypertext and social media (pp. 116-125). ACM.
- $[7] \quad Yin, \, H., \, Sun, \, Y., \, Cui, \, B., \, Hu, \, Z., \, and \, Chen, \, L. \, (2013, \, Au-1) + (2013, \, Au-1$

- gust). LCARS: a location-content-aware recommender system. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 221-229). ACM.
- [8] 田中涼太郎, Panote Siriaraya, 鈴木 健太, 中島 伸介. Happy-Button:日々の幸せ体験認識システムの開発. DEIM2020 P1-8.
- 9] Surprise, (http://surpriselib.com)
- [10] OpenStreetMap, (https://openstreetmap.jp/)