

遺伝的アルゴリズムを用いた

会話型ご当地キャラクタによる地域活性化手法の提案

函館工業高等専門学校専攻科生産システム工学専攻

中島真太郎

内容

1	背景	1
2	研究目的	2
3	既存研究と本研究との違い	3
4	システム概要	4
5	GA について	6
6	協調フィルタリングについて	7
7	提案する会話アルゴリズム	8
7.1	遺伝子プールの初期化	8
7.2	遺伝子の適応度	12
7.3	GA による文章作成	14
8	実行例	15
9	考察	18
9.1	GA パラメータの調整	18
9.2	適応度計算の見直し	18
9.3	品詞データベースの拡張	18
9.4	音声入力機能と GPS 機能のシステムへの導入	18
9.4	外国語への対応	19
10	まとめ	20
	—参考文献—	21
	付録 C 言語ソースコード(会話アルゴリズム)	別紙

1 背景

近年、東北大震災や原発事故、消費税アップ等の影響で、特に地方の経済や観光産業は不調である。そのような事もあり、国内の観光に対する消費額は、低迷期を過ぎてはいるが東北大震災前の水準までまだ回復できていない[1][2]。この現状から今後は各都市間の観光における競争が激しくなると予想され、地域活性化に向けたより積極的なアプローチが行われると考えられる。

一方、その地域活性化の手段の一つとしてご当地キャラクタが近年よく利用されている[3]。ご当地キャラクタとは、ゆるキャラとも称され、イベントや各種キャンペーン、地域おこし、名産品の紹介などのような地域全般の情報PR、企業・団体のコーポレートアイデンティティなどに使用するキャラクタのことである。またそういったかわいらしいイラスト全般を指す場合もある。

この最初のご当地キャラクタは1980代に発祥したと言われ、現在までに多くのご当地キャラクタがデザインされ、自治体はこのご当地キャラクタを用いて消費を観光客に促したり、地域の知名度を向上させたりといったような活動を多く行ってきた。

その一方で、観光客が現地に行っても必ずご当地キャラクタと会えるわけではないといった問題がある。これはご当地キャラクタの地域活性化へのアプローチが十分に生かされているとは言えないという現状がある。

2 研究目的

本研究では、GA（遺伝的アルゴリズム）[4][5][6][7]を用いた会話型ご当地キャラクターによる地域観光ガイドサイトを構築して背景で述べた現状の解決を図る。

構築するサイトは現地情報をリアルタイム発信し、観光客がいつでもどこにいても「クラウド上にいる会話機能を搭載したご当地キャラクター」とコミュニケーションを取ることが出来る。またGAによる会話アルゴリズムを用いることで、ユーザの入力に対して柔軟な出力を行う事が出来る。さらに協調フィルタリング[7][8]の考えを応用し、他の多くの観光客の嗜好情報を学習することで、利用者と嗜好の類似した観光客の情報を提供することが出来る。

本研究を進める事によって、観光客はスマートフォンやタブレット端末等を利用してクラウド上にいるご当地キャラクターと時間や場所を問わず自然言語によってコミュニケーションを行うことが出来る。それにより観光客に対してより自然な観光ガイドを提供して観光消費を促すため地域活性化につながる。

3 既存研究と本研究との違い

本研究と類似している研究として、埼玉県北本市の「とまちゃん」がある[9]。埼玉県北本市は、自治体初の「人工知能」技術を活用したサイトナビゲータをリリースした。住民が親近感を持ちやすい、ご当地キャラ「とまちゃん」に人工知能技術を実装することで、インターネットを通じての地域活性化や観光情報を発信することを目的にして開発された。また、「とまちゃん」は埼玉県北本市が富士ゼロックスシステムサービス株式会社と共同で開発された。

観光客は「とまちゃん」と会話形式でさまざまな情報を収集できる。「とまちゃん」は会話アルゴリズムとして、会話推測機能を搭載した自然言語対話システム「Jackalope Engine (ジャッカロープエンジン)」[10]を利用している。これにより自治体は「とまちゃん」を通じて自治体内の名産品やイベントなどを積極的に発信することで地域の活性化を図ることを可能にし、住民はパソコンやスマートフォンなどで「とまちゃん」との気軽な会話を通じて、市役所への電話やキーワード検索などをしなくても知りたい情報を得ることが可能になった。

ただし「とまちゃん」はある決まった定型的な返事しかしないという問題がある。そこで本研究では会話アルゴリズムとしてGAを採用し、ユーザとの間でより自然で柔軟な会話の実現を試みた。

また類似している研究として、Siri (シリ) も挙げられる[11]。Siri は、iOS 向け秘書機能アプリケーションソフトウェアであり、自然言語処理を用いて質問に答え、観光スポットの推薦などを行う。しかし本研究との明確な違いとして、ユーザの嗜好に沿った推薦を行わない。本研究では協調フィルタリングを応用したユーザの嗜好性に沿った推薦を行っているため、その点で相違している。

4 システム概要

本研究で構築した観光ガイドシステムを図 1 に示す。今回構築したシステムではサンプルとして「MYCAL」という自作のご当地キャラクタをシステム中央部に配置したが、このキャラクタを変えるだけで自治体ごとのシステムが簡単に作れるようなデザインとした。また本システムの中央部と下部にテキストボックスを配置し、観光客が会話文を入力する事によってコミュニケーションを行えるようにした。中央部のテキストボックスは自然言語でのシステム出力である。下部はユーザが入力するテキストボックスであり、ユーザはキーボード等で自然言語でのテキストを入力する。ユーザが文章を入力した後システムが文章を出力するので、会話をしているかのようにコミュニケーションを取ることができる。

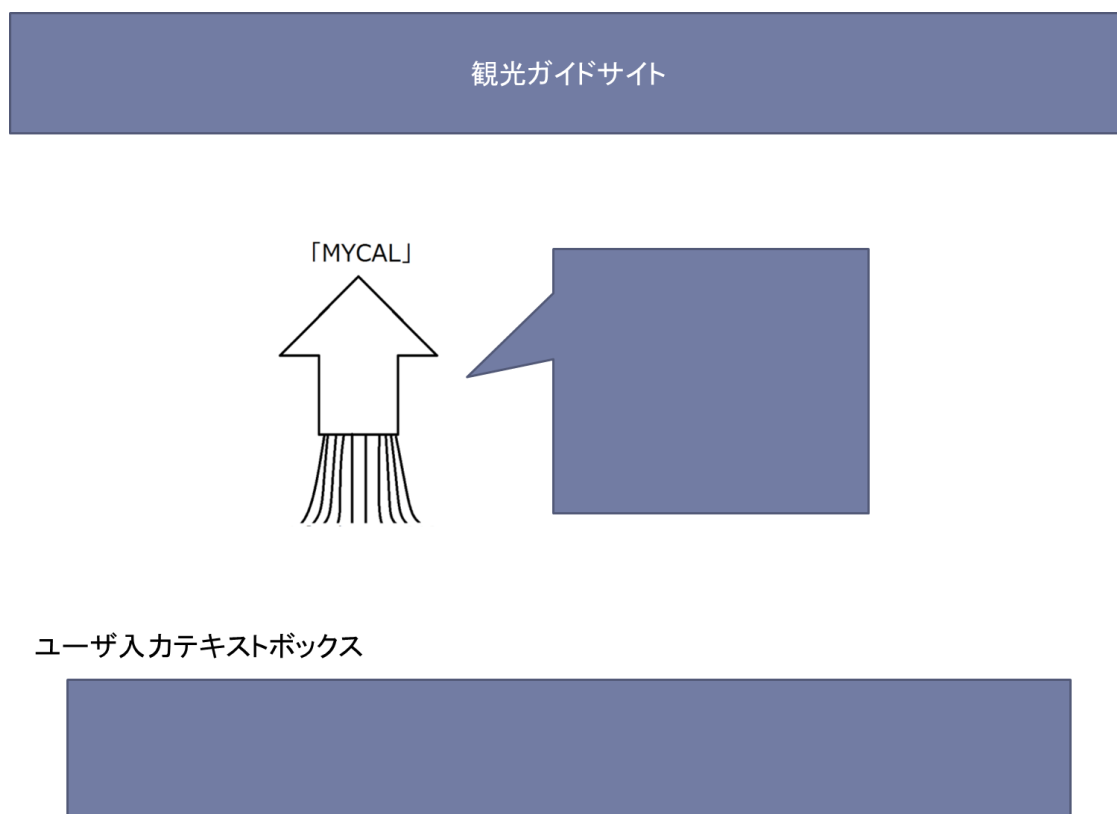


図 1 構築した観光ガイドシステム

本研究で構築した観光ガイドシステムは、観光客に対して観光スポットや飲食店、役所などの位置や、アクセス方法等の地域情報を提供する。

さらにこのシステムはユーザが自治体の観光情報やおすすめスポットを学習させることによって、他のユーザからの質問に返答することができる。学習した情報はDBに保存され、ユーザのロコミを用いた観光ガイドを可能となるので、提供する情報の信憑性が向上する。

また、システムはユーザの入力した日本語から日本語らしさを学習し、ユーザの入力した口調や言い回しを学習することで、自治体毎のご当地キャラクタに個性を持たせることを可能にした。

5 GA について

遺伝的アルゴリズム (図 2) [4][5][6][7] はデータ (解の候補) を遺伝子で表現した「個体」を複数用意し、適応度の高い個体を優先的に選択して交叉や突然変異などの操作を繰り返しながら解を探索するアルゴリズムである。適応度は適応度関数によって与えられ、遺伝子の自然淘汰を行う。適応度関数[12]とは一種の目的関数であり、遺伝的アルゴリズムにおける解の最適度を定量化し、特定の遺伝子が他の遺伝子より優れていることを示すために用いられる関数である。

最適な遺伝子、あるいは少なくとも他よりも比較的最適な遺伝子で、「交叉」「突然変異」などの遺伝的操作が行われ、より適応度の高い個体を含んだ世代を生み出す。

また、遺伝子の表現の仕方によっては組合せ最適化問題や NP 困難な問題などのさまざまな問題に適用可能なアルゴリズムでもある。

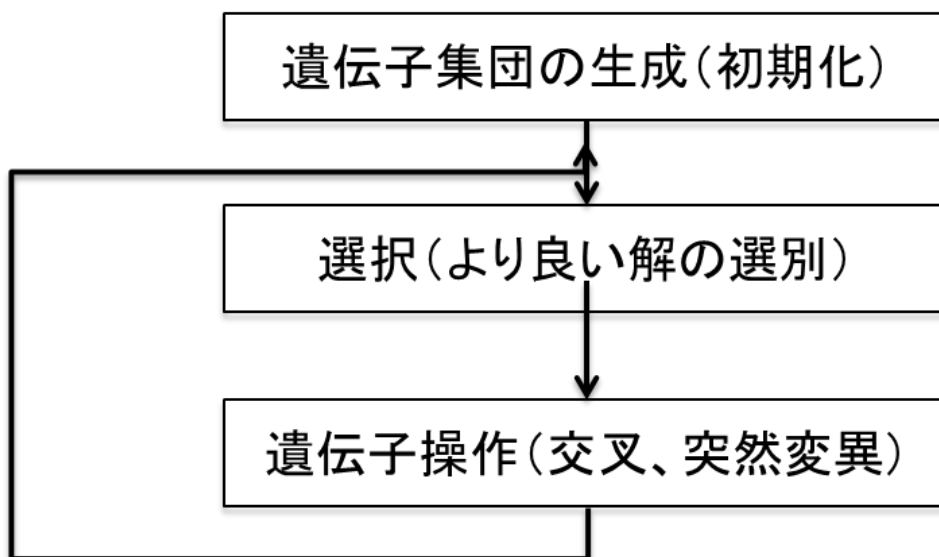


図 2 遺伝的アルゴリズムの概要

6 協調フィルタリングについて

本論文で提案するシステムにおいて、適切なキーワードを抽出するために協調フィルタリング[7][8]のアルゴリズムを応用した。協調フィルタリングとは多数のユーザの嗜好情報を蓄積し、あるユーザと嗜好の類似した他のユーザの情報を用いて自動的に推論を行う要素技術である。本システムは名詞と動詞を適切なキーワードとして認識してテキストマイニング[13][14][15]を行っている。

7 提案する会話アルゴリズム

本研究で用いている GA と協調フィルタリングによる会話アルゴリズムについてこの章で説明する。

7.1 遺伝子プールの初期化

初めにランダムに遺伝子を多数生成して遺伝子プールを初期化する。一つの遺伝子は複数の遺伝子座の集合で表現される。今回は各遺伝子座に ID と No の 2 つの情報を持たせ、ID と No を品詞データベースから参照する事によって各遺伝子座は形態素を表す (図 3)。ここで形態素とは意味を持つ言葉の最小単位の事である。ID は形態素の種類(品詞)を意味しており、1 から 5 の整数を名詞、動詞、形容詞、助詞、助動詞に対応させた。No は各品詞データベースに格納されている形態素の識別ナンバーである。遺伝子の生成例を図 4 に示す。

なお、今回は文章からの形態素の切り出しに Mecab を用いた(図 5) [16]。Mecab は奈良先端科学技術大学院大学にて開発されたオープンソースの形態素解析エンジンである。

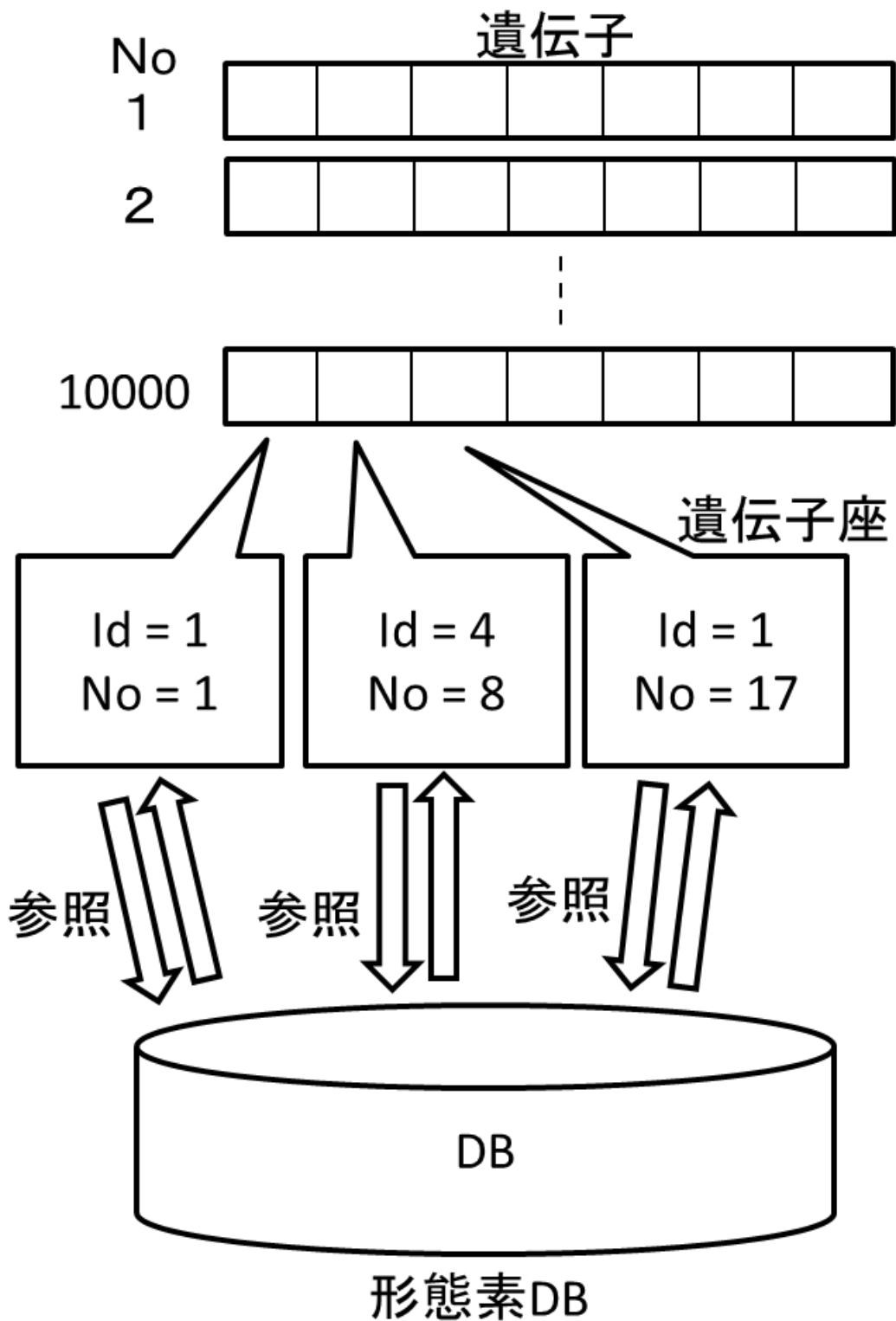


図3 遺伝子プールの初期化

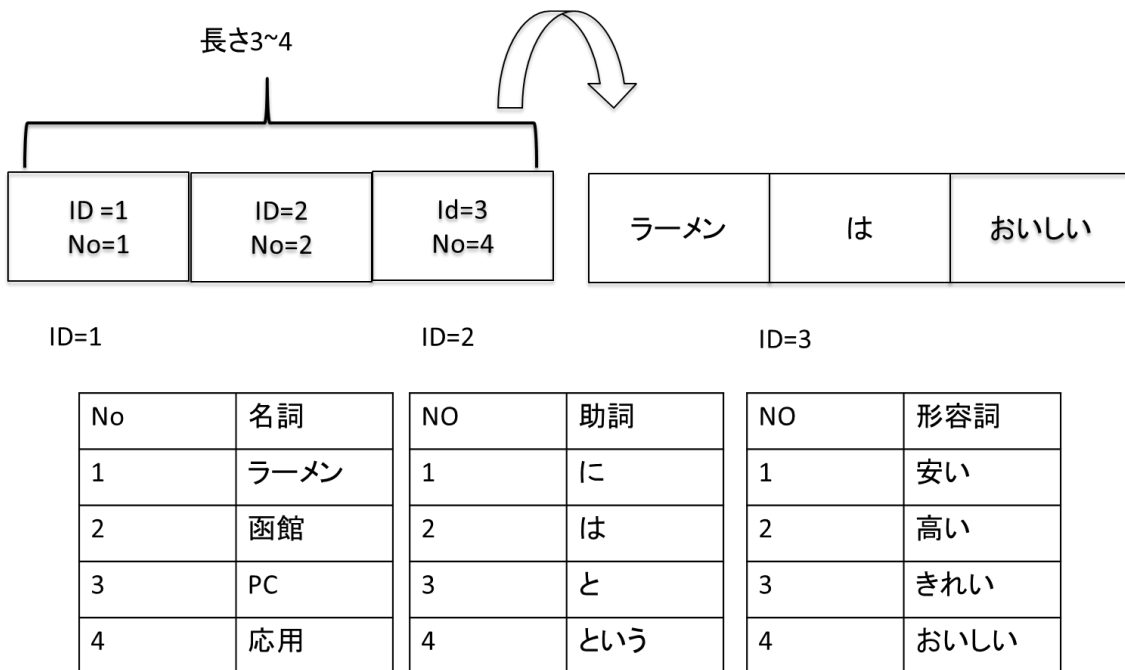


図 4 遺伝子の生成例

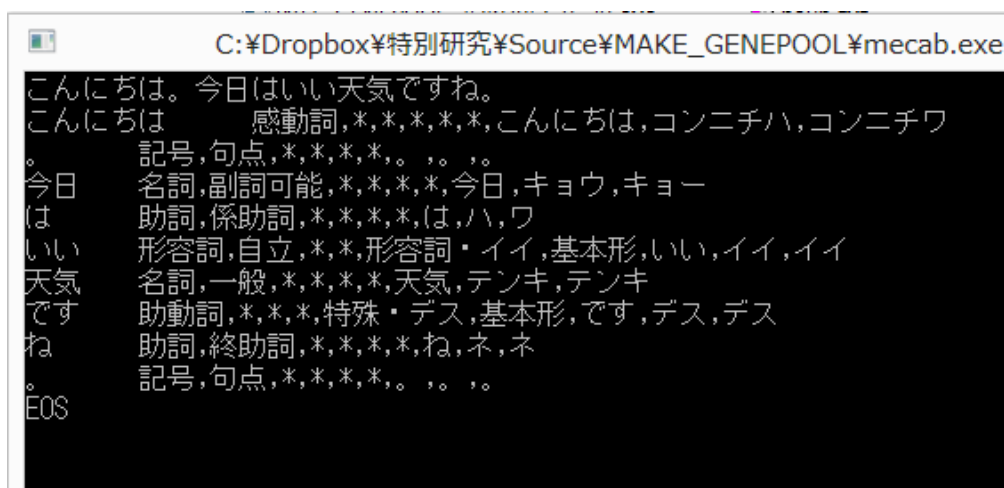


図 5 mecab の実行画面

また、本システムはある観光情報について複数のユーザーが書き込みを行うことが想定される。そこでロコミを用いて信憑性を向上させるために、入力キーワードと関連性のある適切なキーワードを過去の入力ログから抽出して初期遺伝子に出力する。

適切なキーワードの抽出には協調フィルタリング[7][8]のアルゴリズムを応用した。本システムは名詞と動詞を適切なキーワードとして認識してテキストマイニングを行っている。具体的には適切なキーワードを初期遺伝子に出力するため、ユーザーの入力から抽出したキーワードを基にユーザーの過去の入力ログから条件付き確率を計算してキーワード間の関連性を求め、入力キーワードと関連性の高い出力キーワードを確率に従って出力する(図6)。図6の例では「遊園地」が1/2、「プール」「公園」が1/4の確率で出力される。

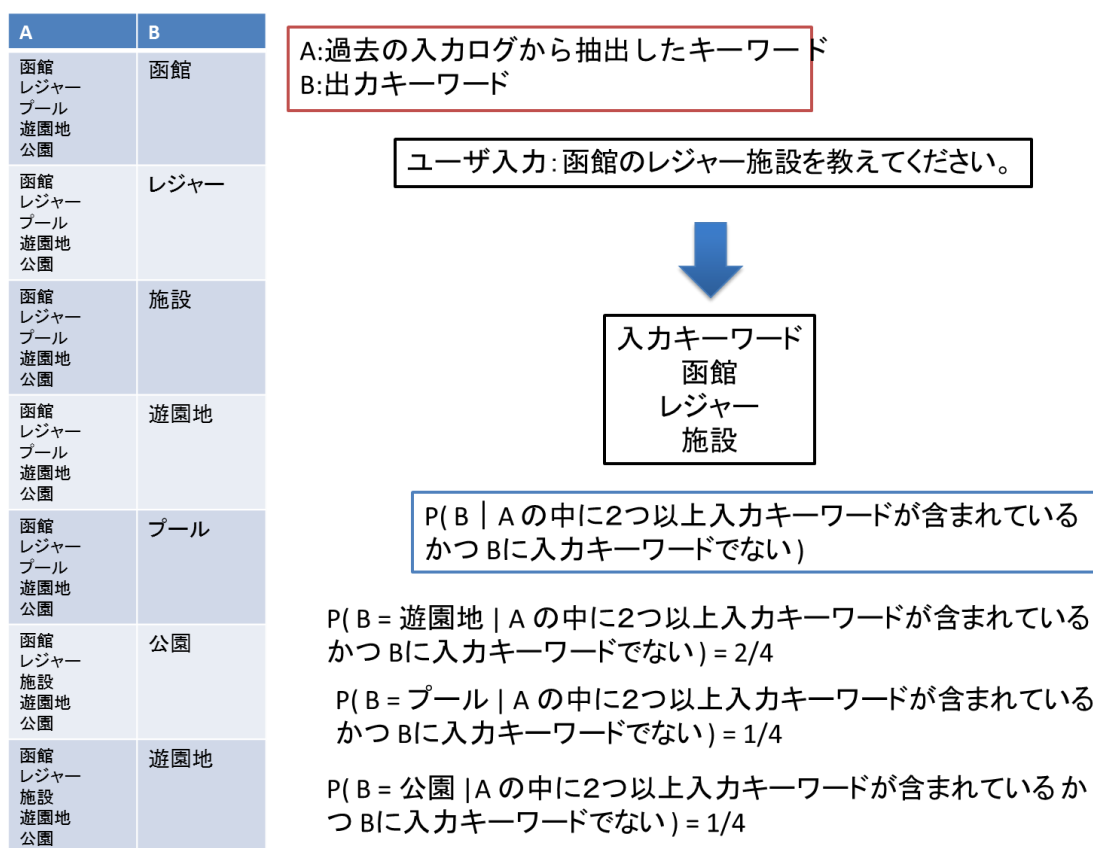


図6 条件付き確率を用いた協調フィルタリングの例

7.2 遺伝子の適応度

本システムで用いている遺伝子の適応度の評価項目は以下の二つである。

- (1) 日本語らしい文章であるか
- (2) ユーザの入力に対して関連性の高いキーワードが含まれているか

詳細を以下で説明する。

- (1) 日本語らしい文書であるか

複数のユーザの入力文章から、隣り合った二つの形態素のつながりのパターンを取り出して形態素の 2-gram を作成する(図 7) [17]。同時にこの形態素の 2-gram につきパターンの出現回数も記録する。複数のユーザが文章を入力するたびにパターンと出現回数が学習・更新される。

形態素table		出現回数
明日	の	--- 26
の	天候	--- 5
天候	は	--- 13
は	晴れ	--- 53
晴れ	です	--- 22
です	。	--- 91

図 7 形態素の 2-gram の例

この形態素の 2-gram に記録されている形態素のつながりパターンと、遺伝子中の各形態素のつながりパターンとのマッチングを行うことで適応度の評価を行う (図 8)。形態素の 2-gram に記録されているパターンの出現回数が多いほどそのパターンは日本語らしい文章であることを意味しているため、出現回数の多いパターンを多数複数含んでいる遺伝子の適応度を高くする。逆に出現回数の多いパターンを含んでいない遺伝子は日本語らしい文章でないと判断して適応度を低くする。

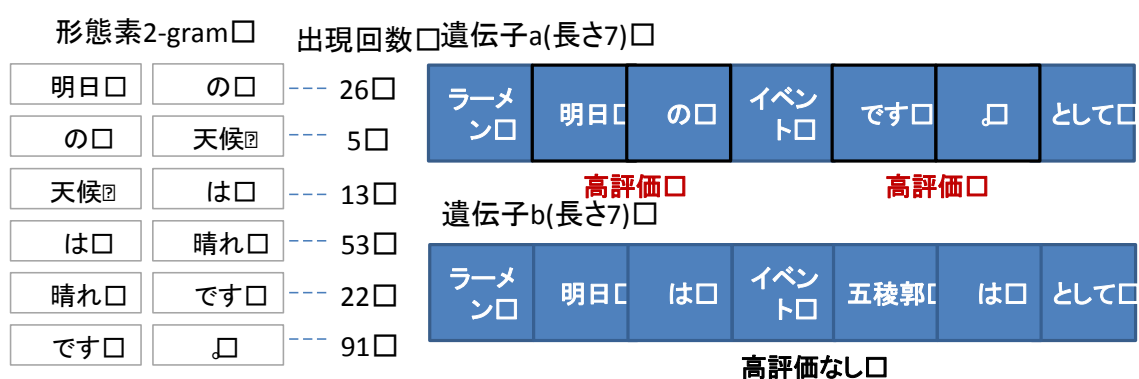


図 8 2-gram を用いたパターンマッチング

(2) ユーザの入力に対して関連性の高いキーワードが含まれているか

遺伝子プールの初期化の際に求めたキーワード間の関連性を示す条件付確率 (図 6) を用いて、入力キーワードに関連するキーワードが含まれている遺伝子の適応度を高くする。そうでない遺伝子は適応度を低くする。

7.3 GAによる文章作成

最後に GA を用いて出力文章を作成する。一般的な遺伝子操作と同じように選択、交叉、突然変異を一つのサイクルとして複数回繰り返し、最終的に最も適応度の高い遺伝子を出力する。なお選択と交叉には様々な方式があるが、今回は[18][19]を参考にして選択にはトーナメント方式、交叉には一点交叉を採用した。

トーナメント法とは、あらかじめ決めた数だけ集団の中からランダムで個体を取り出し、その中で最も適応度の高い個体を選択する方式である。トーナメントサイズを変更する事で選択圧をコントロールできるという長所がある(図9)。

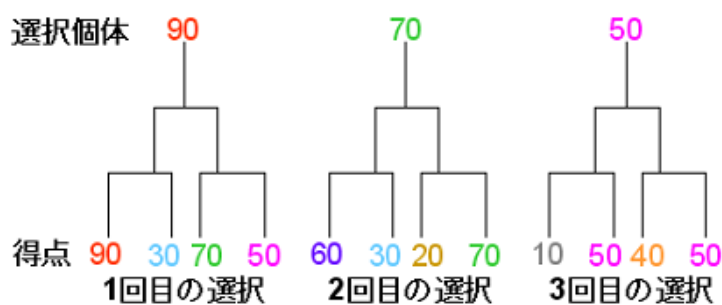


図9 トーナメント法

また一点交叉とは、遺伝子が交叉する交叉点をランダムで一つ選び、その場所より後ろを入れ換える方式である(図10)。

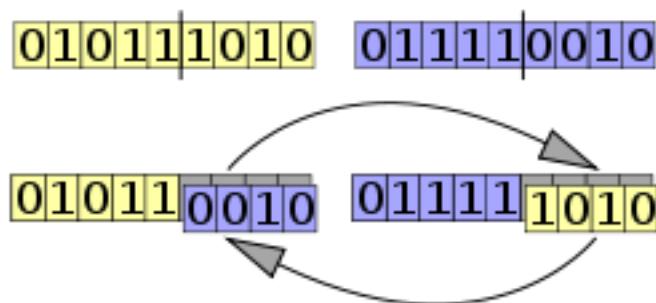


図10 一点交叉

8 実行例

実際にシステムを実行した例をこの章で示す。ここで、今回用いた GA の各パラメータを表 1 に示す。

表 1 各種パラメータ

パラメータ	遺伝子数	繰返し回数	突然変異率	選択率	交叉方法	選択方法
値	10000 個	10000 回	2%	20%	一点交叉	トーナメント

実行例を図 11 と図 12 に示す。図 11 がサイトへの入力例、図 12 が出力例である。また実行結果の一覧を表 2 に示す。

表 2 から、本システムは概ね日本語らしい文章を返答しているが、まだ完全な日本語ではないと思われる。例えば、1 番目と 3 番目の出力例の様に主語と述語をつなぐ「が」や「に」などの助詞が抜けていたり、4 番目の例の様に動詞が抜けていたりする。さらに問題点として出力される文章が過去の入力ログに依存しやすいといった点も挙げられる。

観光ガイドサイト

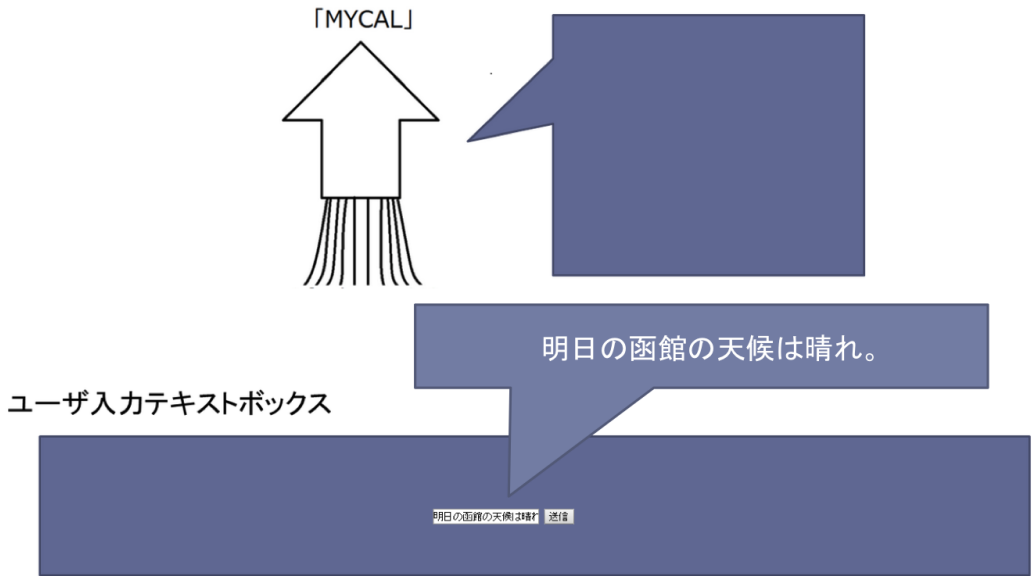


図 11 実行例（入力例）

観光ガイドサイト

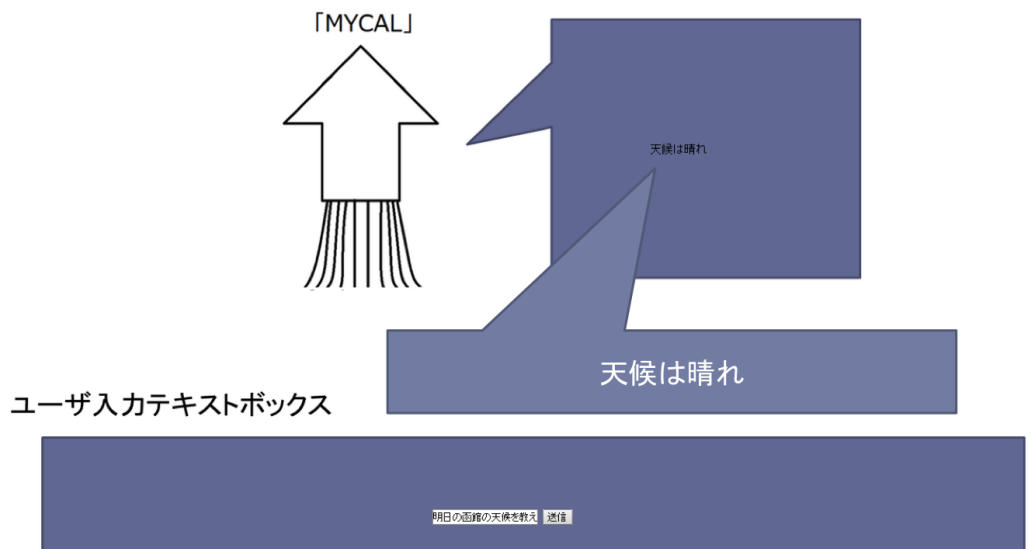


図 12 実行例（出力例）

表 2 実行結果

学習	ユーザの質問	システムの出力
ローソンのみかんはおいしいです。	おいしいみかんは。	ローソンあります。
王将はおいしいラーメン屋です。	おいしいラーメン屋は。	王将はうまい。
函館のレジャー施設としてプールや図書館や遊園地があります。	函館のレジャー施設を教えてください。	遊園地あります。
今日の五稜郭のイベントは祭りです。	今日の五稜郭のイベントは。	祭り。

9 考察

実行結果から、システムの性能を向上させる為に以下の様な課題が挙げられる。

9.1 GA パラメータの調整

GA においてパラメータの設定は実行結果に大きく変化を与える要因であるため、最適なパラメータの設定が必要不可欠である。したがって、より良い文章が導出されるようなパラメータ値を模索していく必要がある。

9.2 適応度計算の見直し

本システムの出力をより日本語らしい文章にするために適応度の計算方法を改良する必要がある。例えば、助詞や動詞が抜けている、出力される文章が過去の入力ログに依存しやすい等の問題があった。従って更なる精度向上のために改良していく必要がある。

9.3 品詞データベースの拡張

現在、システムが保有している品詞データベースは名詞、動詞、形容詞、助詞、助動詞の5つで構成されている。これらの品詞の種類を拡張し、語彙を増やしていく事でシステムが出力する文章の精度が向上するものと考えられる。たとえば形容詞と動詞の意味を持つ形容動詞や、より会話を自然に行う為に必要な感動詞や接続詞の追加が考えられる。

9.4 音声入力機能と GPS 機能のシステムへの導入

現在システムはユーザがスマートフォン等のキーボードでテキストを入力する形になっている。しかしキーボード等の入力では誤って文章入力を行うなど入力するのに比較的多くの時間を要する。この問題に対してシステムへの音声入力を可能にすることで入力に対する煩わしさの解消を目指す。さらにユーザの位置情報に最適な観光ガイドを行うために、スマートフォン等の GPS 機能を利用することも検討する。

9.4 外国語への対応

近年、円安が進んでいることから外国人観光客が増えている[20]。そこで本システムを外国人からも使用してもらうため翻訳機能を付加することを考えている。

10 まとめ

本研究では、地方の経済や観光産業を振興するために、GA を用いた会話型ご当地キャラクターによる地域観光ガイドシステムを構築した。構築したシステムは現地情報をリアルタイム発信し、観光客がいつでもどこにいても「クラウド上にいる会話機能を搭載したご当地キャラクター」とコミュニケーションを取ることが出来る。また GA による会話アルゴリズムを用いることで、ユーザの入力に対して柔軟な出力を行う事が出来る。さらに協調フィルタリングの考えを応用し、他の多くの観光客の嗜好情報を学習することで、利用者と嗜好の類似した観光客の情報を提供することが出来る。

本システムを利用する事によって、観光客はスマートフォンやタブレット端末等を利用し、クラウド上にいるご当地キャラクターと時間や場所を問わず自然言語によってコミュニケーションを行うことが出来る。それにより観光客に対してより自然な観光ガイドを提供して観光消費を促すため地域活性化につながる。

結果として、求める機能を持った観光ガイドシステムを構築することができ、おおむね満足できる出力結果が得られたが、一方では完全な日本語が出力されていない場合があった。そのため、今後は会話アルゴリズムの見直しなど、システムの洗練化を行う必要があると思われる。

—参 考 文 献—

- [1] 国土交通省 観光省, 統計情報,
<http://www.mlit.go.jp/kankocho/siryou/toukei/index.html>
- [2] 国土交通省 観光省, 旅行消費額の概況,
<http://www.mlit.go.jp/common/000210907.pdf>
- [3] 月刊地域作り一般財団法人地域活性化センター, ご当地キャラクターによる地域活性化,
<http://www.chiiki-dukuri-hyakka.or.jp/book/monthly/1208/html/f13.htm>
- [4] 棟朝雅晴, 遺伝的アルゴリズム, 森北出版株式会社(2008)
- [5] 伊庭斉志, 遺伝的プログラミング, 東京電機大学出版局(2001)
- [6] 森直樹, Java で学ぶ遺伝的アルゴリズム, 共立出版(2007)
- [7] 小高知宏, はじめての AI プログラミング, 株式会社オーム社(2010)
- [8] マイケル・J.A. ベリー, データマイニング手法—営業、マーケティング、CRM のための顧客分析, KAIBUNDO(2006)
- [9] 埼玉県北本市, 「とまちゃんホームページサポート」
<http://www.city.kitamoto.saitama.jp/index.shtml>
- [10] Enzan Trades Inc, Jackaloe Engine,
<http://www.enzantrades.co.jp/business/opensystem/product/jackaloe/>
- [11] Apple, siri, <https://www.apple.com/jp/ios/siri/>
- [12] 菅沼研究室静岡理工科大学, 遺伝的アルゴリズムと最適化,
http://www.sist.ac.jp/~suganuma/kougi/other_lecture/GA/GA.htm
- [13] 石井一夫, よくわかるデータマイニング, 日刊工業
- [14] 元田浩, データマイニングの基礎, Ohmsha
- [15] 石村貞夫, 多変量解析によるデータマイニング, 共立出版
- [16] 奈良先端科学技術大学院大学, Mecab Japanese morphological analyzer,
<https://code.google.com/p/mecab/>
- [17] ナレッジワークス株式会社, N-gram と形態素解析,
<http://scope.knowledge-works.co.jp/2012/06/n-gram%E3%81%A8%E5%BD%A2%E6%85%8B%E7%B4%A0%E8%A7%A3%E6%9E%90/>
- [18] Jiro Kamiura, GA/DGA のパラメータ(4) — 交叉 (交叉手法, 交叉率) ,
<http://mikilab.doshisha.ac.jp/dia/research/person/jiro/reports/GAparams/GAparams>

04.html

[19] Marek Obitko, GAのパラメータ,

<http://www.obitko.com/tutorials/genetic-algorithms/japanese/recommendations.php>

[20] 日本経済新聞, 訪日外国人客、初めて年 1000 万人突破 円安で割安感 13 年 20%増,

http://www.nikkei.com/article/DGXNASFS20038_Q3A221C1000000/