

1. はじめに

チャットにおいてコンピュータとの会話を成立させること目的としたチャットボットがある。チャットボットは会話の内容を理解するのではなく、主に会話の成立を目的とした単純なアルゴリズムで応答を行っている。この応答文の生成は主に、次の3つの形が検討されている。①辞書型、②ログ型、③マルコフ文生成型の3つが検討されている^[1]。

①辞書型は、あらかじめ応答文を辞書として持ち、辞書内に格納されている応答文の中から会はないように沿ったものを応答文として出力する。②ログ型は、過去のログを検索し、利用者の発言と類似するログを見つけて、応答文の材料にする。③マルコフ文生成型は、手持ちの情報(単語・語句)をマルコフ連鎖で組み合わせ応答文を生成する。

辞書型は一定の文しか返せなく、マルコフ型では不自然な文章になりがちである。そこで本研究では辞書型とマルコフ型を併用したチャットボットの試作を行い、自然な文と不自然な文がどの程度出るかを考察する。

2. マルコフ文生成型

ある出来事が過去の出来事に影響を受けて確率的に生じるとするマルコフ確率過程を利用した文生成型である。本研究は、最も簡単な単純マルコフ過程(出来事が直前の過去の出来事のみ影響を受ける)を使用した生成方法^[2]で行う。

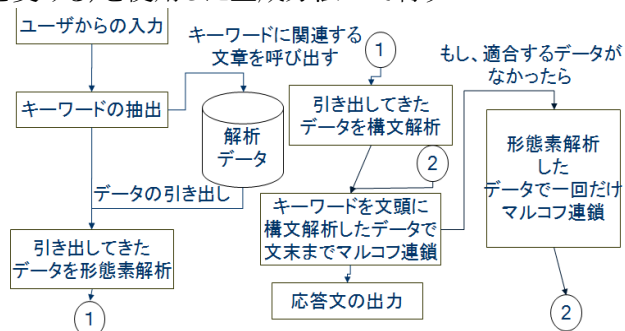


Fig.1 システム構成フローチャート

3. システム構成

新聞記事から辞書型とマルコフ型を使用した応答文を生成する Chatbot を試作した。試作システムのシステム構成(フローチャート)は、Fig.1 の通り。まず入力文からキーワードを抽出し、そのキーワードに適する文すべてをデータベースから抜き出す。その後、抜き出した文を形態素・構文解析し、構文解析を行ったデータを使用してマルコフ連鎖を行い、応答文の出力を行う。もし構文解析のデータ中

に次へのマルコフ連鎖が無いなら形態素で検索を行う。今回、データベースの形態素解析、及び構文解析は黒橋、河原氏が開発した juman, knp を使用した^[3, 4]。また本研究のデータベースは平成 17 年のスポーツ新聞記事を使った先行研究^[5]のデータベースを利用した。

4. 結果と考察

結果例を Table.1 に示す。

Table.1 入力文に対する応答文の例

入力文(キーワード)	出力文
松坂は先発で出場した。(松坂)	松坂投手に立ち上がった観衆から大きな拍手が送られた。
4番は中村で決まりだろう。(中村)	中村紀が6番一塁でオープン戦に出場可能な支配下選手契約を目指す中村だけど開幕三塁手の座を奪い取るしれない。
ベッカムのけがは完治したのだろうか。(ベッカム)	ベッカムは右ひざ内側の靭帯を伸ばしているという。
松井の打数は安定している。(松井, 打数)	打数無安打だった。
初戦で桑田と戦うらしい。(戦, 桑田)	戦に出られると思っていなかったそれですぐに結果を出すところはさすが。
寺島的前提条件は打席に立つことだ。(寺島, 前提)	前提としてそれを途中で放棄することを契約するのは選手の保有権はおかしいと巨人の清武代表は話した。

全体として会話は成り立っていないが、出力の日本語は文として成り立っているものが多い。形態素と構文ではキーワードが多くなると精度が悪くなることがある。これは関連の少ない2つのキーワードを設定してしまったためであろう。

5. おわりに

・入力に対し見掛け上、それなりの応答を行う。ただ出力連鎖の後、1度出力されたものは別の文になおすことも必要であろう。また抽出文が増えると時間が非常にかかり、プログラムの高速化が必要になる。

6. 文献

- [1]森部敦, 他:“自動会話システム(人工無能)の開発とその応用”, 信学会, 信学技報, IE2005-29, pp.11~16, 2005 年
- [2]小高 知宏:“はじめての AI プログラミング C 言語で作る人工知能と人工無能” オーム社 2006
- [3]黒橋, 河原:“日本語形態素解析システム JUMAN”, <http://nlp.kuee.kyoto-u.ac.jp/nl-resource/juman.html>
- [4]黒橋, 河原:“日本語構文解析システム KNP”, <http://nlp.kuee.kyoto-u.ac.jp/nl-resource/knp.html>
- [5]志賀明雄:“新聞記事から応答文を生成するチャットボット”, H17 年度, 都立航空高専卒業論文。