

単語間の類似度が低い単語を使う人工無能の開発

Development of a chatbot program using low-similarity words

CS24 中野卓人
指導教員 大島真樹

1. はじめに

今日までの人工無能による会話システムは、語彙や一定の文をデータベースにとりまとめ、ユーザが発した言葉・文章に対しデータベースから適切な回答を選択し、それを返すというものがある。この返答において、今までの研究資料等を見ると、より正確な答えを返すことの研究が行われている。[1]

しかし、より人間らしい会話を実現するのであれば、『間違える』という行動を盛り込むことが必要であると考えられる。その間違いにも様々な種類があるが、今回は『わざと行う』間違いに注目したい。

わざと行う間違いとは、駄洒落や冗談といったように、現在の会話とは直接関係無い単語でも、間接的に関係のある単語を会話に出すことで、従来関係の無いはずの単語で会話を成立させる術である。

2. 研究方法

例えば、『うどん』という単語から、50個の単語を得られたとする。その得られた単語と、入力した『うどん』という単語との関連度合い(類似度)を調べ、ソートする。その結果から上位に来る単語を利用した人工無能を作成すれば、それなりにまともな返答をする人工無能[2]となるだろう。更に、ここであえて下位の単語を利用することにより、会話に変化が現れることと思う。しかし、『うどん』という単語の検索結果として表示された単語である以上、下位の単語も一応うどんと関連した単語である。

そういった単語を利用することによって、どのような会話になるのか、検討していく。

プログラムとしては、辞書型の人工無能を作成し、それに Web サイトの検索動作を自動制御するプログラムを追加する。これにより、入力された会話文から単語を抜き出した時、その単語を利用してそれに関連した単語を拾うことが出来る、人工無能となる。

3. 検証方法

類似度の低い単語を利用した人工無能を作成するわけであるが、これを人間らしい会話をする人工無能であるかどうかを判断するのは難しい。

会話を持続させるという点においても、会話続行の条件をまず厳密に定義できない。このあたりは主観に頼ることになるが、数人に実際会話してもらい、その会話が流れとして不自然かどうか判断してもらう。

4. 結果

表1 類似度表

単語A	単語B	単語C
うどん	麺	ラーメン
そば	麺	ラーメン
みかん	柑	ウンシュウミカン
炒飯	インディカ	ジャポニカ
パン	製パン	フジパン

上記の表は reflexa[3]と呼ばれる連想検索エンジンを利用し、検索を行ったもの。単語Aを入力し、得られた結果を単語Bとし、その単語Bを再度入力し、得られた結果を単語Cとして表にまとめた。

5. 今後の発展

上記の表では、単語Aから見て、類似度の高い単語をB、低い単語をCとして見ると、AとCの関係は必ずしも「関連なし」とはならない。しかし、一見関係なさそうな物も存在し、これは類似度の低い単語が生成できたと我々は考える。

本研究の結論の一つとして、人工無能の分野に新たな文章生成法を提案した。類似度を割り出す方法や、その確立、類似度を利用した単語選択法などには、まだ様々なパターンがある。今後はこれらを比較検討し、どのような方法が一番効率的なのを見極めていきたい。

文献

- [1] ”人工無脳は考える”,
<http://www.ycf.nanet.co.jp/~skato/muno/>
- [2] ”自動会話システム(人工無能)の開発とその応用”, 森部敦, 毛利公美, 森井昌克電子情報通信学会技術研究報告, S, オフィスインフォメーションシステム 105(283), 11-16, (2005)
- [3] ”連想検索エンジン reflexa”,
<http://labs.preferred.jp/reflexa/>