

2R-5 「ネットワーク型言葉遊び」の知識獲得への応用

金杉 友子† 笠原 要‡ 阿部 明典‡ 湯川 高志‡ 松澤 和光‡
† NTTアドバンステクノロジー(株) ‡ NTT(株)コミュニケーション科学研究所

1. はじめに

与えられた知識が不完全でも常識で補って概略的な答を導く「アバウト推論」[1]の研究を行っている。アバウト推論は、システム外部からの入力を言葉で行ない、その意味を工学的に取り扱って解釈するため、人間の常識に当たる膨大な知識を獲得する必要がある。

そこで、辞書から言葉の意味(「概念」)を自動獲得する方式を提案し、日常語4万の概念に関する知識ベース(「概念ベース」)を実際に構築した。更に概念間の類似性について評価を行った[2]。だが、辞書等に記述された知識が必ずしも「基本的な知識」とはいえない問題がある。

一方、概念ベースの工学的評価法の一つとして、「言葉遊び」を概念ベースを用いて自動的に行う方式を提案し[3]、その有効性の評価に、連想しりとり・クロスワードパズル・なぞなどを対象に自動解答の結果と人間の解答とを比較した。加えて評価の大規模化のため、ネットワークを介した言葉遊びをWWWで行う方式を検討[4]し、実験システムを公開した。[5]

現在多くの個人がホームページ(以下HP)を作成しており、言葉遊びはHP上で様々に展開されている。中には、アクセスするユーザの知識を取り込んでより面白い遊び方を追求するHPもあり、ユーザも遊びの一環として抵抗なく知識を入力している。ネットワークを用いた言葉遊びは、大規模な知識蓄積の新たな手段として有望といえる。

本稿ではネットワークを用いた言葉遊びを分類・分析し、その知見に基づく新たな知識獲得の構想について説明する。

2. ネットワーク型言葉遊び

2.1 現状と分類

8種のサーチエンジンを用い「言葉遊び」等の語で検索できた、97年6月時点のHP総数約60件の分類を試みた。その際、次の2点を考慮した。

1. ネットワークの特性の活用

リアルタイム性の追求や自作ソフト/システムの運

用といったネットワークの特性を活用したものか、自作発表や投稿募集のような従来の雑誌メディア等でも成立するものかを問題にした。

2. ユーザからの情報提供の必要性

投稿によるページ作成、自作システム改良のため単語登録等をユーザに求めるといった、HP存続にはユーザからの情報提供が必須であるか、自作発表の場としてのみのHPを利用してるかを問題とした。

以上を考慮して、言葉遊びのHPを4分類した。

表1: 言葉遊びHPの分類

	ネットワーク 活用する	ネットワーク 活用しない
ユーザに情報 提供求める	パターン1 16件	パターン3 39件
ユーザに情報 提供求めない	パターン2 5件	パターン4 2件

この分類からは、以下のようなことが言える。

1. ネットワークを生かしたHPの件数が少ない

ネットワークの特性をいかすには、手間や技術を要するためと思われる。リアルタイム性の追求には、HP作者の頻繁な改良が必要である。また、CGI, JAVA, Shockwave等を用いたインターラクティブなHP作成にはプログラム作成技術が要求される。

2. ユーザに情報提供を求めるものが多い。

これは、Web更新のサイクルが短いことに起因する。ユーザの興味を惹くには、頻繁な情報の更新が要求される。ユーザの常に新しい情報を見たがる要望に自作の発表だけで対応するのは難しい。

2.2 パターン1に関する考察

ここでは特にネットワークの活用と知識抽出を考えてパターン1を取り上げ、コミュニケーションの視点から更に2種類に分けた。

タイプ1: 人-マシンのコミュニケーション-6件
自作ソフトやシステムをユーザに使用させるもの。

- ・自動生成: 詩や俳句、ニュース等の自動生成。
- ・対話システム: 人工のキャラクタとユーザの対話。

ユーザの提供する情報はソフト/システムの改良用に単語や文章を登録するものが多い。ユーザの入力がシステムに活用され、入力された知識の蓄積自体が、遊びの質の向上になる特色を持つ。

Knowledge Acquisition with Neo Word Plays

Kanasugi.T†, Kasahara.K‡, Abe.A‡, Yukawa.T‡

, Matsuzawa.K‡

† NTT Advanced Technology

‡ NTT Communication Science Laboratories

‡ I-1, Hikarinooka Yokosuka-shi, Kanagawa 239 Japan

タイプ2：人-人のコミュニケーション-10件
ユーザ間のコミュニケーション。直前にアクセスしたユーザの記述に続けて情報を入れることで、言葉遊び自体が成立するもの。

- ・しりとり、伝言ゲーム
- ・共作：連歌、小説、漫才などの共作。

ユーザの情報提供は、言葉遊びを続けるのが目的である。知識を蓄積し質の向上を狙うものではない。

以上パターン1のHPについて細分類したが、我々のHPはタイプ1：人-マシンコミュニケーションに相当する。既に保有する知識ベースに更なる知識獲得を行うには、このタイプが最適と思われる。

3. ユーザからの知識獲得法

3.1 常識の獲得について

我々は概念ベース構築の際、辞書・百科辞典等の既存の電子化文書からの知識獲得を基本方針とした。これらの電子化文書からは、大量の語の知識が獲得できる上、それが識者によって編纂された確かなデータであることがその主な理由である。

だが、このような電子化文書は人間が読んで単語を理解することを目的とし、理解に必要な「基本的常識」はユーザに備わっているのが前提になる。例えば辞書の「馬」の説明文中には、動物であることは記述されていても、耳と目が2つあるというような自明の知識は書かれない。そこで、言葉の知識を獲得する別の手段として、「ネットワーク型言葉遊び」のユーザからの知識獲得を提案する。

3.2 言葉遊びを応用した知識獲得法の提案

まず、「ネットワーク型言葉遊び」のユーザから得られる知識の特色をあげる。

- (1) 固有名詞・新語等が得られやすい
- (2) 人間の連想や感覚的な表現の知識が得られる
- (3) 決まった形式のコーパス作成が容易
(文の格に対応する単語登録やあるキーワードに対する対話文作成など)
- (4) 遊びの一環として入力するため、知識と意識しない常識が獲得できる。
- (5) 大人数の知識蓄積が行なえる。

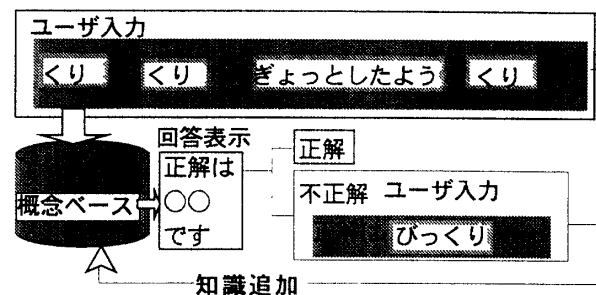
ユーザからの知識獲得は人間の判断に基づくため、電子化文書とはかなり異なった性質の知識が手に入ることが予想できる。このような知識を獲得する方式を以下に記す。

- (1) 電子化文書を用い(不完全な)常識を構築し、それを利用したネットワーク型言葉遊びHPを作成。
- (2) 上記HPをインターネット上で開放。
- (3) ユーザは言葉遊びをしつつ、情報の提供を行う。
- (4) 情報を知識として蓄積し、それを元に(1)を改良。
- (5) (2)に戻る。

次に過去に提案した「なぞなぞ」の言葉遊びシステム[3]を用いた、提案方式の応用例を考える。

「くりはくりでもぎよっと驚くりは?」(正解:びっくり)のような「AはAでもBなAは?」の形式のなぞなぞを概念ベースを用いて自動解答するシステムに適用する。Aには正解の読みの一部、Bには正解の語の特徴が入る。A,Bを空欄にしたテンプレートを作成し、HP上でユーザに空欄を埋めさせる。A,Bを手がかりに概念ベースを用いて自動解答を行い、答を表示。不正解だった場合は、正解を登録させる。

図1：なぞなぞへの応用例



この情報入力により、正解の語の特徴が概念ベースに追加できる。物の形状や色などあまり辞書等に記載されない特徴が得られると予想される。上記は現状のシステムで可能な応用例だが、将来的には対話等コーパス蓄積やテキスト以外の画像や音のデータの収集等、ユーザが遊び感覚で入力する様々なデータを蓄積する方向をめざしたい。

4. おわりに

本稿ではHP上約60の「ネットワーク型言葉遊び」を分類し、中でも多くのユーザに受け入れられている、人-マシンコミュニケーションのHPの性質について分析を行った。またこの形式の言葉遊びを用いて多数の人から大規模な知識を獲得する新しい構想を提案した。今後はこのシステムをインターネット上で開放し、その有効性の評価を行う予定である。

References

- [1] 松澤和光,湯川高志,笠原要,藤本和則: アバウト推論技術-常識を持ったコンピュータの実現を目指して-, NTT R&D, Vol.45, No.11(1996).
- [2] 笠原要,松澤和光,石川勉: 国語辞書を利用した日常語の類似性判別, 情報処理学会論文誌, Vol.38, No.7(1997).
- [3] 金杉友子,松澤和光,笠原要: アバウト推論の「言葉遊び」への適用, 信学技報, NLC96-31, pp.1-7(1996).
- [4] 松澤和光,阿部明典,金杉友子,永森千晴: 概念ベースを用いたネットワーク型「言葉遊び」の提案, 言語処理学会第3会年次大会, pp.577-580(1997).
- [5] <http://www.kecl.ntt.co.jp/banana>