

概念ベースを用いたクロスワードパズル作成システム

2S-7

帆苅 譲[†] 石川 勉[†]笠原 要^{††}[†]拓殖大学工学部情報工学科^{††}NTT コミュニケーション科学研究所

1. はじめに

知識が不完全でも、その不完全さに応じた概略的な判断が行えるシステムの実現を目ざし、研究を進めている。ここでは、不足知識を常識知識で補足することを想定しているが、その一つとして単語の意味の類似性を判別するための知識ベース(以下、概念ベース[1]と呼ぶ)の構築を行ってきた。本報告では、この概念ベースを利用したスリーヒントのクロスワードパズル[2]の解作成システムについて述べる。

2. 概念ベースの構成

概念ベースは、基本的には国語辞書等の単語(概念とする)の語義文について形態素解析を行い、そこに含まれる自立語を属性として選び、その自立語の語義文中での使用頻度をその属性値として構成する。この中で各概念 g は以下の形で表現される。

$$g = \{(p_1, q_1), (p_2, q_2), \dots\} \quad \dots (1)$$

ここで p_i は概念 g の特徴を現す属性、 q_i はその属性値であり、概念と属性の関係の深さを示している。具体的には、例えば概念「林檎」についての属性と属性値は、「果実: 0.328」「花: 0.277」「高木: 0.272」等となっている。この概念ベースには、概念として日常的に用いられる約4万の単語が収容されている。

3. システム構成

3.1 構成概要

本システムの構成を図1に示す。クロスワードパズル(以下パズル)の解を作成する部分と、解を構成する単語に対するヒントの候補群を抽出する部分からなる。入力パズルのパターンであり、出力はパズルの解とヒントである。パズル作成部では、概念ベース中の単語を文字数により分割した複数の集合の中から、入力されたパターンに当てはまるように適切な単語を選択し、解の作成を行う。ヒント抽出部では、作成部で得られたパズルの解とする単語に対して、概念ベースを用いてヒントの候補群とする単語を抽出する。

3.2 パズル構成法

以下、単語のはいる一連のマスを「枠」、枠どうしが

Forming crossword puzzles using a knowledge base of words

Yuzuru Hokari[†], Tsutomu Ishikawa[†], Kaname Kasahara^{††}

[†]Department of computer science, Takushoku university

^{††}NTT Communication science laboratories

交わるマスを「交点」と呼ぶ。また、パズル全体について全ての枠が交点を介して連結している形のパズルを「連結である」と定義する。

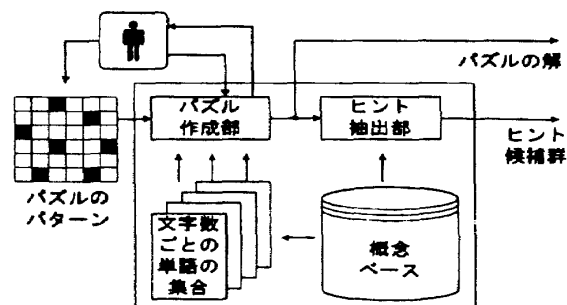


図1 システム構成

(1) 解作成法

パズル作成部では、交点の一致を取りながら枠を埋めていくが、この場合、毎回同じ位置から単純に埋めていく方法では、パズルの規模やパターンあるいは単語数の増加により、単語の比較回数は爆発的に増大する可能性がある。そこで、一般に単語を構成する文字数が多くなるほど、その文字数の単語数が少なくなるという特徴を利用する。表1に使用した概念ベースの単語の文字数と単語数の関係を示す。文字数が4以上の単語は語数が少なくなっていることがわかる。

表1 概念ベースにおける文字数と単語数の関係

文字数	単語数	文字数	単語数
2	1266	8	280
3	6966	9	92
4	12617	10	21
5	6277	11	8
6	2184	12	2
7	806	13	1

この特徴から、全ての枠の中で文字数の最も多い枠から埋めていくと、後の方で埋まる文字数の少ない枠ほど当てはまる単語数が多くなる。すなわち、単語の組み合わせのパターンが増え、解の見つかる確率が上がると予想できる。ここで採用したアルゴリズムは以下の通りである。なお、本アルゴリズムは、連結であるパズルを対象としているが、非連結の場合には、これを繰り返し実行すればよい。

[解作成アルゴリズム]

- ① パズルの空の枠の中で、最も文字数の多い枠を選択する。その枠と同一の文字数の集合の中からラ

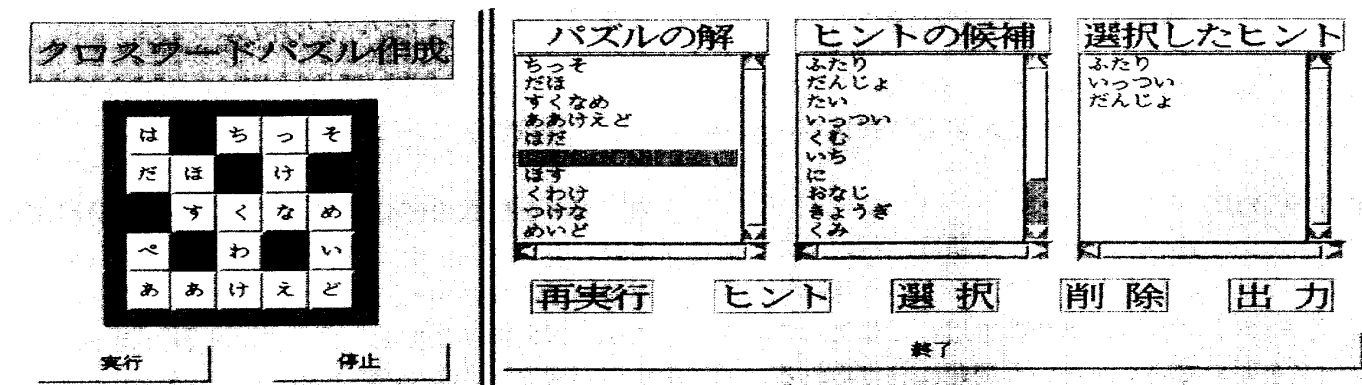


図2 クロスワードパズル作成システムの利用例

ランダムに単語を選び、その枠を埋める。

- ② 埋まっている全ての枠に対して、交点を介しつながっている空の枠の中で、最も文字数の多い枠を選択する。
- ③ ②で選択した枠について、その枠の交点が矛盾なく一致する単語を選択する。ここで、単語が見つければその単語を枠に埋める。すべての枠が埋まっていればパズルの作成を終了する。まだ空の枠が存在すれば②へ戻る。また、一致する単語が見つからなかったときは、バックトラック処理を行うために④へ進む。
- ④ 交点が一致しなかった空の枠について、交点を介してつながっている枠の中で、既に単語が埋まっている枠の中から文字数が最小の枠を選択する。選択した枠に対して、別の単語を選択する。このとき、その枠の他の交点が一致するように単語を選択する。ここで、別の単語が見つければ、その単語でその枠をうめ②へ戻る。見つからなかったときは、その枠を空にして、その枠について、このバックトラック処理を繰り返し実行する。

(2) 解修正法

作成されたパズルの解の中で、利用者の気に入らない単語が存在する場合がある。この場合には、その単語を別の単語に置き換える。置き換え方法としては、まず気に入らない単語を文字数ごとの集合から削除する。次にその単語が埋まっていた枠を空にする。そして上記アルゴリズムの②から、再度パズルの解の作成を行う。

3. 3 ヒント抽出法

ヒント抽出部では、概念ベースを用いヒントの抽出を行う。具体的には、解とする各単語を概念と考えて、その概念の属性を抽出し、その属性をヒントの候補群とする。例えば、パズルの解の中に「林檎」という単語があったときは、ヒントの候補群として「果実」「花」「高木」等の単語が抽出される。

4. 使用例

本システムの使用例を図2に示す。利用者は、まずパズルの作成パターンの塗りつぶしを行う。次に、結果が出力されるので、気に入らない単語があれば変更を行う。パズルが完成すると、それぞれの単語に対するヒントの一覧が表示される。利用者はその中からパズルのヒントとしたい単語を選択する。

本システムを用いて作成したパズルの例を図3に示す。この10×10のパズルの作成時間はSun Sparc Station5で12秒であった。

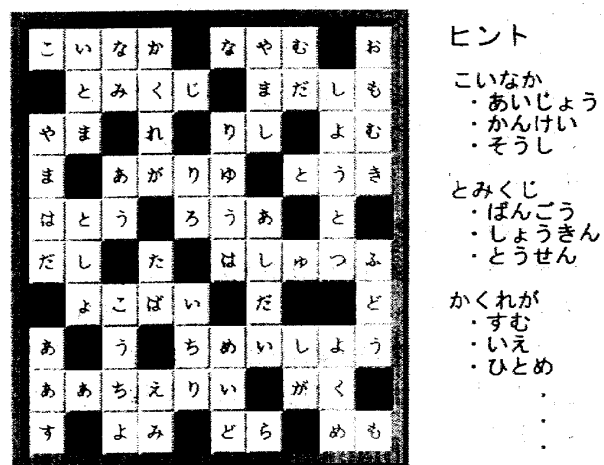


図3 解作成結果

5. まとめ

クロスワードパズルの解の作成およびそのヒントの抽出を行うシステムを作成した。解の作成では、文字数と単語数の関係に着目したアルゴリズムを用いることで、実用的な大きさのクロスワードパズルを10数秒で作成できた。

参考文献

- [1] 笠原、松沢、石川：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 Vol38 No7pp1272-1283
- [2] 金杉、松沢、笠原：アバウト推論の「言葉遊び」への適用、進学技法 NLC96-31(1996-10)