

文の意味を考慮した常識的時間判断システムの構築
Construction of Common-sense Time Judgment System
Based on the Meaning of Sentences

吉川 義紀†
Yoshiki Yoshikawa

芋野 美紗子‡
Misako Imono

土屋 誠司‡
Seiji Tsuchiya

渡部 広一‡
Hirokazu Watabe

1. はじめに

情報化社会の発展とともに、今日の情報技術は目覚ましい進歩を遂げた。デジタル・ディバイドの問題を考えると、高齢者を含むあらゆるユーザが特別な訓練などをせずにご利用しやすいシステムの実現が望まれる。高度なシステムの1つにロボットがあげられる。ニュースなどでも取り上げられるように、コミュニケーションロボット需要が増えており、人間に違和感を与えず、人の相手をするにはコミュニケーションは自然な形で進められることが望まれる。

人と人がコミュニケーションや会話を行う場合を考えると、受け手は話し手の発話内容を受け取り、その発話は何を伝えたいかを理解することで会話が成立する。これは人間が文章やそれを構成する単語に関する「常識」を持っているからである。また、人が日常的に行う会話では時間や季節を用いた表現が頻りに用いられる。例えば、友人とどこかに出かける際には会話中で待ち合わせの日時を指定し計画を立てる。また、人間は「クーラーといえば夏」や「おやつといえば3時」というように、季節や時間帯を連想することができる。本稿では、このような季節・時間帯の連想を行う動作を「時間判断」と呼ぶ。我々は会話の中で「常識」を用いて時間判断を当たり前に行っている。そこで、人とコンピュータが円滑な会話を行うためには、時間判断ができるシステムが必要であると考えられる。本稿では、入力された文章の意味が考慮でき、最も重要な時間を出力できる常識的時間判断を行うシステムの構築を目的とする。

2. 関連技術

2.1 概念ベース

概念ベース^[1]とは電子国語辞書などを基にして機械的に構築された大規模な知識ベースである。概念ベース内で語は「概念」として定義されており、概念の意味的特徴を表す「属性」とその属性が概念にとってどれだけ重要かを表す数値である「重み」の対の集合より定義されている。

2.2 関連度計算方式

関連度計算方式^[2]とは概念ベースを利用し概念と概念の関連の強さを表現するものである。関連度の値は0.0~1.0の値を取り、その値が大きいほど関連が強い。

3. 構築する時間判断システム

入力された文章の意味を考慮でき、最も重要な時間を出力できる時間判断システムを構築した。システム全体の流れを図1に示す。

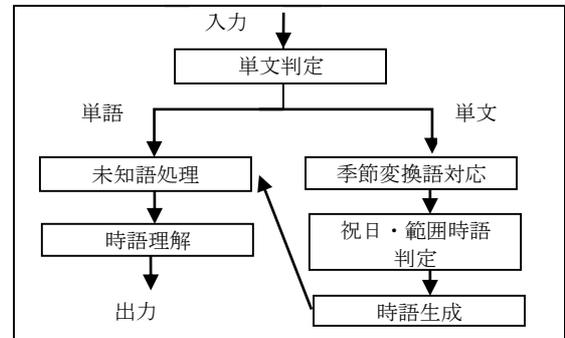


図1. システム全体の流れ

入力は主述関係を1つしか持たない単文もしくは単語とする。入力に対して形態素解析を行い、単語か文章かを判断する。用言か体言1つのみの場合は単語と判定する。また、主述関係が1つの場合は単文とする。

単文入力の場合、まず季節変換語対応(3.1節後述)という処理を行う。この処理は入力文が季節を変えるものであるか判定を行う。次に、祝日・範囲時語判定(3.2節後述)という処理を行う。この処理は入力文中に祝日や「～から～」といった期間を表す語(以下範囲時語)が存在するかを判定する。そして入力文を単語に変換する時語生成(3.3節後述)を行う。入力が「葉が落ちる」の場合、辞書的に意味が近い「落葉」という言葉に変換する。意味の近い単語が取得できない際は、入力文の自立語を取得する。

直接の単語の入力、または単文から生成した単語に対しては、先行研究^[3]である旧時間判断システム中の技術を用いる。先行研究では人が時間を連想できる語(以下時語)とその時語が持つ季節や時間帯(以下季節感)をアンケートで約600セット収集し時語知識ベースを構築した。この知識ベースは「桜=春」や「おやつ=3時」というように単語に対して常識的な季節や時間帯に関する知識を有する。また、単語が時語知識ベースに登録されていない語であるとき未知語処理により登録している語に変換する。最後に時語理解(3.4節後述)を行う。先行研究では、未知語処理をした単語を時語知識ベースで検索し、その語が持つ季節感を出力するものであった。本稿では入力が季節感を変えるものである場合は、単語から得られる季節感を変更する処理を行う。

3.1 季節変換語対応

入力が文章の場合、その文意によって単語の季節感を変えてしまう場合がある。例えば「ストーブをしまう」という入力の場合、人間は「冬」という季節を感じる「ストーブ」を「しまう」という動作から「冬の終わり」や「春」を感じる。このような「しまう」でみられる働きをする語を季節変換語と呼び、アンケートで収集した52語から季節変換語DBを構築した。表1に季節変換語DBの例を示す。この処理では、入力が季節変換語が存在した場合に対応する役割を判断し3.4節に述べる時語理解で役割の知識を用

†同志社大学大学院理工学部研究科
Graduate School of Science and Engineering, Doshisha University

‡同志社大学理工学部
Faculty of Science and Engineering, Doshisha University

いて季節の変換を行う。

表 1 季節変換語 DB の例

季節変換語	役割
しまう	季節を進める
望む	季節を戻す

3.2 祝日・範囲時語判定

まず、祝日判定では、入力文中に祝日が存在した場合、日付の取得を行う。旧時間判断システムでは祝日 DB は存在しなかったため新たに日本国内で法により決定された祝日をもとに DB を作成した。

次に、範囲時語判定では、入力文中に範囲時語が存在するかを判定し、その役割取得を行う。ここで範囲時語とは、例えば「今日から 2 日後」という入力における「から」や「後」といった言葉である。範囲時語 DB は時間表現に関する既存研究^[4]の定義によって決定された範囲時語を格納した。表 3 に範囲時語 DB を示す。

表 2 範囲時語 DB の例

範囲時語	役割
から	前開始地点
後	時間計算+

上記の例だと、今日の日付を時間の始まりとして、「後」に与えられた役割より、その日付に 2 日をプラスするという処理を行い、出力は 6/22~6/24 となる。

3.3 時語生成

時語生成では、入力された単文を単語に変換する処理を行っている。この変換処理には逆引き国語辞書^[5]を採用した。逆引き国語辞書とは、単語の意味を表現した入力に対し、対応した語を出力するものである。逆引き国語辞書では国語辞書に定義された語の意味を示す文を用い、概念ベースを用いて文章間の関連性を求めることで、文同士の意味の近さを考慮して文を一語に変換する。例えば入力が「葉が落ちる」の場合、「落葉」=「葉が落ちること」と辞書に意味が定義されており、入力と辞書的に意味の近い「落葉」に変換される。

3.4 時語理解

時語生成により生成された単語もしくは入力単語と、時語知識ベースを参照して、その単語が持つ季節感を取得する。例えば生成単語が「落葉」であれば、時語知識ベースに「落葉=秋」と登録されているため、「秋」と出力される。また 3.1 節で述べたように、季節変換語が持つ役割を理解し、その役割に従い、季節感を変える。入力が「ストーブをしまう」であれば、時語理解により「ストーブ=冬」が取得されるが、「しまう」に対する役割「季節を進める」が取得され、「冬から春」に変換し、季節を出力する。

4. 精度評価

旧時間判断システムと新時間判断システムについて精度評価を行った。評価を表 4 に示す。ここで、評価は被験者 5 名(各持ち点 2 点)の合計点が「8 点以上の場合には○、5 点以上 8 点未満は△、5 点以下は×」という評価を行った。なお、評価セットはアンケートにより収集し、単語 20 個、単文 80 文とした。

表 3 被験者ごとの結果

	○	△	×
旧時間判断システム	22%	10%	68%
本稿時間判断システム	58%	15%	27%

表 4 より精度は○が 36%上昇し、△は 5%上昇し、×は 41%下がった。

5. 考察

新旧時間判断システムは単語のみが入力の場合には同様の正解率であった。文章入力に対して入力文の意味を理解できなかったことが精度の差であると考えられる。

入出力例を表 5 に示す。まず、うまくいかなかった例は「バブル経済」等の具体的な固有名詞である。また常識的な時間判断の例として「星を見る」と「プラネタリウムで星を見る」という 2 文について考える。これらは「星」という時語から「夜(18 時~3 時)」という結果が得られる。しかし、常識的にプラネタリウムが夜中の 3 時まで営業しているとは考えにくい。よって出力結果は誤りであると考えられる。また、範囲を持つ時間とは「上旬」や「下半期」など様々な表現が存在する。しかし、本稿では範囲時語を「日本語時間表現の解釈法」という論文に定義された 7 語に限定してしまったため、それ以外の表現に対応できなかった。

表 4 入出力例

入力	出力	評価
梅雨が明ける	7 月	○
星を見る	18 時~3 時	○
プラネタリウムで星を見る	18 時~3 時	△
バブル経済っていつ?	出力なし	×

6. まとめ

本稿では既存の時間判断システムに対し、入力文の意味を考慮できる時間判断システムの構築を行った。3 章で述べた提案手法を実装することで表記揺れや季節を変える入力文への対応、範囲時語・祝日対応などの機能を追加し、その結果、36%精度が向上した。問題点としては固有名詞への対応が挙げられる。さらに、入力は単文のみとしているが、入力の制限は本来無くすべきものである。また「片づける」「しまう」のように同じ意味の季節変換語でも様々な表現が存在する。このような表記揺れにも連想を用いて対応する手法を検討する必要がある。

謝辞

本研究の一部は、科学研究費補助金(若手研究(B)24700215)の補助を受けて行った。

参考文献

- [1] 奥村紀之, 土屋誠司, 渡部広一, 河岡司, “概念間の関連度計算のための大規模概念ベースの構築”, 自然言語処理, Vol.14, No.5, pp.41-64, 2007.
- [2] 井筒大志, 渡部広一, 河岡司, “概念ベースを用いた連想機能実現のための関連度計算方式”, 情報科学技術フォーラム FIT2002, pp.159-160, 2002.
- [3] 野村理樹, 渡部広一, 河岡司, “時間の常識的判断メカニズムとその未知語処理”, FIT2003, pp.191-193, 2003
- [4] 溝渕昭二, 住友徹, 泓田正雄, 青江順一, “日本語時間表現一解釈法”, 情報処理学会論文誌, Vol.40, pp.3408-3419, 1997.
- [5] 三瀬慶久, “単一文の概念化に関する研究”, 修士論文, 2009.