

## 連想メカニズムを用いたスケジュールに関する質問応答システム The Question Answering System about the Schedule by using Association Mechanism

高淵 智史<sup>†</sup> 吉村 枝里子<sup>‡</sup> 土屋 誠司<sup>‡</sup> 渡部 広一<sup>‡</sup>  
Satoshi Takabuchi Eriko Yoshimura Seiji Tsuchiya Hirokazu Watabe

### 1. はじめに

近年、人間と円滑なコミュニケーションのとれる知的ロボットの実現が求められている。そこで、ロボットが人間の発言を理解し、発言の意図を捉えた会話を行うことが必要である。また人間の生活において、スケジュールは非常に重要である。スケジュールをコンピュータが把握し、会話形式でユーザが求める情報を提示出来れば、日々のスケジュール把握が容易になる。そこで、ユーザのスケジュールを理解し、それに関する質問に対して応答するシステムを構築する。このシステムは、予めユーザのスケジュールを格納した表(以下、スケジュール表と呼ぶ)を作成し、ユーザが予定に関する質問文(以下、予定質問文)を入力した際、スケジュール表から適切な答えを抽出し応答する。人の連想を模した連想メカニズムを与え、更に入力文の表す意図・意味・時間を正しく理解するシステムを用いることで、言葉の関連性を理解し、様々な表記で表現される文を正しく予定質問文と判断する。

### 2. 関連技術

#### 2.1 連想メカニズム

連想メカニズムとは、ある語から別の関連する言葉を連想したり、複数の言葉の間にある関連性を判断するといった人間が自然に行う思考をコンピュータで実現するために作られたものである。概念ベース<sup>[1]</sup>と関連度計算方式<sup>[2]</sup>を基に連想メカニズムは構築されている。

##### 2.1.1 概念ベース

概念ベースとは複数の国語辞書や新聞から自立語を抽出して自動構築された、語(概念)の特徴を表す語(属性)を集めたものであり、属性には重要度(重み)が定義されている。概念ベースには概念が87242語格納されている。

##### 2.1.2 関連度計算方式

関連度計算方式とは概念ベースを利用し概念と概念の関連の強さを定量的に評価するものである。関連度の値は0.0~1.0の連続値をとり、1.0に近づくほど関連が強い。

#### 2.2 意図理解システム

意図理解システム<sup>[3]</sup>は、入力文を挨拶・呼掛け、命令・依頼、疑問、情報の4種類に分類し、疑問であれば5W1H疑問文、YES/NO疑問文、N者択一疑問文など、その詳細について解析するシステムである。

#### 2.3 時間判断システム

時間判断システム<sup>[4]</sup>とは、単語や日常的な表現の文から日付、時間、季節を得ることができるシステムである。

<sup>†</sup>同志社大学大学院理工学研究科  
Graduate School of Science and Engineering, Doshisha University  
<sup>‡</sup>同志社大学理工学部  
Faculty of Science and Engineering, Doshisha University

時間を特定できる言葉を時語と呼び、それを基に入力の日付、時間、季節を判断する。事前に人手で用意した時間に関する知識及び概念ベースとシソーラスを用いて、知識のない言葉に関しても時間や季節を連想可能としている。シソーラスには、NTTシソーラス<sup>[5]</sup>を用いている。

### 3. 提案データベース

#### 3.1 スケジュール質問語データベース

『教える』や『伝える』等の、質問文で用いられることが多いと考えた10個の用言を質問語と定義し、それを格納したデータベースを、スケジュール質問語データベースと呼ぶ。

#### 3.2 予定語データベース

『暇』や『用事』等の、予定に関する文章で用いられることが多いと考えた15個の名詞や用言を予定語と定義し、これらを格納したデータベースを予定語データベースと呼ぶ。

### 4. スケジュールに関する質問応答システム

スケジュールに関する質問応答システムとは、入力文を解析し、それが予定質問文であるとシステムが判断した場合、その予定質問文の回答となる予定内容をスケジュール表から抽出し、質問に対する適切な応答を出力するシステムである。このシステムの流れを図1に示す。

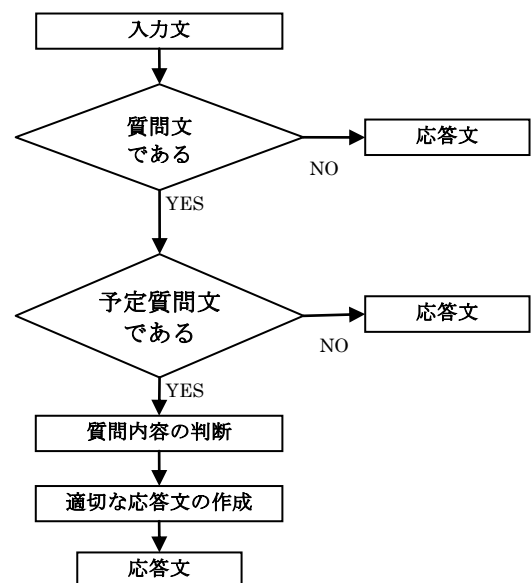


図1:質問応答システムの流れ

ここで用いるスケジュール表は、あらかじめ手入力したものを用いる。項目は、予定開始日、予定終了日、予定開始時間、予定終了時間、場所、メンバー、予定内容の7つである。スケジュール表の例を表1に示す。

表1 スケジュール表の例

予定開始日	予定開始時間	場所	予定内容
2013/7/25	9:00	東京	旅行
予定終了日	予定終了時間	メンバー	
2013/7/27	15:00	友人	

#### 4.1 質問文の判断

入力文が質問文かの判断は、入力文を意図理解システムに入力し、「疑問」、「命令」と判断した文を質問文の候補とする。「命令」と判断した文は、入力文の用言を抽出して、スケジュール質問語データベース内の質問語との関連度の算出を行い、閾値以上となる用言が存在すれば、質問文と判断する。この閾値は、ある概念  $X$  と  $X$  と類義や同義など、関連が非常に強い概念  $A$  との関連度を算出し、その平均とした。今回はそのような概念  $X$  と概念  $A$  の組み合わせ 500 組から閾値は 0.437 と算出された。

#### 4.2 予定質問文の判断

4.1 節で質問文と判断した文を When 質問文、When 以外の 5W1H 質問文、それ以外の質問文の 3 種類に分類する。

When 以外の 5W1H 質問文は、自立語ごとに時間判断システムに入力して時間を取得し、入力文の日時を決定する。これを現在日時と比較し、入力文の日付の方が過去であれば、予定質問文ではないと判断する。それ以外の文は、予定質問文であると判断する。

5W1H 以外の質問文は、自立語を取り出し、予定語データベースの語との関連度を算出する。そこで閾値以上の自立語があれば、入力文を予定質問文であると判断する。

When 質問文は、時間判断を行うための時語が無い場合、そのまま 4.4 節の適切な応答文作成を行う。

#### 4.3 質問内容の判断

予定質問文が何について質問しているのかを判断し、応答文で返すべき内容を決定する。具体的には、意図理解システムを用いて 5W1H のいずれについて尋ねる質問文なのかを判断し、これを基に 4.4 節で応答文を作成する。

When 質問文は場所、Who 質問文はメンバー、When 質問文は開始日時と終了日時、What・Why・How 質問文は予定内容が質問内容であると判断する。例えば「明日の予定は何ですか」は What 質問文であるため予定内容、「明日はどこに行きますか」は Where 質問文であるため場所について聞かれていると判断する。

5W1H 質問文以外の予定質問文については、全てのスケジュール項目について尋ねられていると判断する。

#### 4.4 適切な応答文の作成

4.3 節までの結果を基に、適切な応答文を作成する。まず、入力文に合致する予定がスケジュール表に入っているかを判断する。入力文が When 質問文以外の場合、4.2 節で判断した入力文の日時とスケジュール表の日時を照合し、適切なスケジュール内容を取り出し、「<予定内容>があります」や「<予定開始日>の<予定開始時間>から<予定終了時間>までです」のようなテンプレートを用いて応答する。

入力文が When 質問文である場合、どの日付の予定を抽出すれば良いか不明なことが多いため、スケジュール表

の項目で時間とは関係のない、場所、メンバー、予定内容を取り出し、テンプレートと組み合わせて応答する。

### 5. 評価と考察

#### 5.1 予定質問文の判断の評価

評価は予定に関する予定質問文 100 文を用いて行った。評価文を入力し、予定質問文であると判断出来ているかを評価する。4.4 節で応答文を返していれば正解、それ以外は間違いとした。質問文の内訳は、When 質問文が 30 文、Where 質問文が 20 文、Who 質問文が 20 文、What 質問文・その他の質問文が 30 文である。結果、予定質問文と判断されたものは 99 文であり、正解率は 99%となった。

#### 5.2 出力結果の評価

5.1 節と同じ評価文を入力し、正しいスケジュール項目を取得した応答文かを目視で評価する。適切な答えを出力すれば正解とし、不適切な答えを出力すれば不正解とした。結果、正解率は 55%となった。

#### 5.3 出力結果の評価

このような評価となった原因は主に「建国記念日」のような記念日が時間判断システムで定義されていないため、日時を取得出来なかったという点、「次の〇〇」といった表現の処理が出来なかった点、翌月の〇日に旅行の予定が入っている際に「〇日の旅行」という表現を入力すると、今月の〇日の予定を参照してしまう点等が挙げられる。時間判断システムの精度向上や、時間判断では対応の出来ない処理の追加、入力文とスケジュール表の内容の綿密な照合などが、システムの精度向上のために必要だと考えている。

### 6. おわりに

本研究では、スケジュール表から適切なスケジュール内容を取り出して応答するシステムを構築した。システムの評価は、応答出力に関しては 55%と高くない数字にとどまった。今後、ユーザが入力した文章からスケジュールを作成し、それに対する質問応答が出来れば、人間とコンピュータとの円滑な会話に繋がると考えられる。

#### 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] 荒木孝允, 奥村紀之, 渡部広一, 河岡司, “比較対象概念の共通属性を重視する動的関連度計算方式”, 同志社大学理工学研究報告, Vol.48, No.3, pp.14-24, 2007.
- [3] 大井健二, 渡部広一, 河岡司, “知能ロボットの意図理解と応答制御方式”, 言語処理学会第 8 回年次大会発表論文集, pp.275-278, 2003.
- [4] 岩瀬元秀, 渡部広一, 河岡司, “文の意味理解に基づく常識的時間判断システムの構築”, 電子情報通信学会技術研究報告, Vol.106, No.587, pp.1-8, 2007.
- [5] NTT コミュニケーション科学研究所監修, “日本語語彙体系”, 岩波書店, 1997.