

表に関する自然言語質問に答えることのできる 知的表理解システムの構築

古川 成道[†] 渡部 広一[†] 河岡 司[†]

†同志社大学大学院工学研究科

E-mail : †nfurukawa@indy.doshisha.ac.jp

人間は、表から多くの知識を獲得し、その表に関する質問に柔軟に答えることができる。コンピュータが表を理解することができれば、表から知識を得ることができ、コンピュータはより知的になると考えられる。また、表を理解できれば、その表に関する質問にも答えることができるようになる。本稿では、概念ベースや常識判断システムを用いて、表の構造や内容を解析し、自然言語による質問に対する応答が可能となる知的表理解システムを構築する。表に記載されている語句のみで構成される質問文だけでなく、表に記載されていない語句を含んだ質問文に対しても約 60% の正解率で答えることができる手法を提案する。

キーワード 概念ベース, 常識判断, 表理解, 質問応答, 自然言語

Construction of Intelligent Table Understanding System with Answering Questions of Natural Languages Concerning the Table

Narimichi FURUKAWA[†] Hirokazu WATABE[†] Tsukasa KAWAOKA[†]

† Graduate School of Engineering, Doshisha University

E-mail : †nfurukawa@indy.doshisha.ac.jp

We can answer questions concerning table by acquisition of many knowledge from the table. If computer can understand table it can acquire knowledge from table, and it is considered the computer is more intelligent, and it can answer questions about table. This paper constructs intelligent table understanding system with analysis of structure and contents of tables, and answering questions of natural language, using concept-base and commonsense judgment system. We propose a method of answering not only questions construction by words mentioned in table, but also include words not mentioned in table. The correct answer rate of 60% is gained by this method.

Keywords: *Concept-base, Commonsense Judgment, Understanding Tables, Answering Questions, Natural Language*

1. はじめに

近年、コンピュータという存在は、人間の日常生活や社会活動において必要不可欠なものとなった。人間にとってコンピュータは非常に便利な機械である。しかし、その便利になったコンピュータが多種多様な機能を持つために操作方法が複雑になり、人間がそれを使いこなすのに大変時間がかかってしまい、使いやすさの点で考えると便利とは言い難い。万人に使いやすいコンピュータを実現するには、普段、我々人間同士が行っているコミュニケーション(意思疎通)をコンピュータのインターフェースに取り入れることが必要となる。

この意思疎通ができると、人間とコンピュータが自然な会話を行えるようになり、人間はコンピュータを扱う際に何も操作する必要がなく、会話のみで行える。これによって、万人に使いやすいコンピュータを実現できる。この人間と自然なコミュニケーションがとれるような『知的さ』が、現在のコンピュータには欠けている。

この『知的さ』を得るためには、人間の持つ常識判断や知識などが必要となる。これらを得ることによって、コンピュータは柔軟な対応をすることができる。その知識を得る手段の1つとして、表から知識を得ることを考える。表には様々な情報や知識が含まれているため、表を理解することによってコンピュータはより知的になると考える。

本稿では、コンピュータに表を理解させる手法を提案する。与えられた表の構造やデータ、内容を理解し、その表に対して様々な自然言語質問に答えることのできる表理解システムを実現する方法を提案する。

2. 表理解システム

2.1 表の定義

本稿では、表を以下のように定義する。

1. $m \times n$ のセルから成るデータの羅列で、データの表記的な意味での二次元から成るもの
2. 罫線で囲まれていて、データの範囲が限定されており、ある程度定式化されているもの

「表」は『文章ではわかりにくい事柄などを、分類整理して、見やすくまとめたもの。リスト』であると大辞林第二版では書かれている。データベースよりももっと大域的なものであり、表はあくまで情報を見やすくするために形を変えただ

けのものと考えられる。表1はこの定義に沿って表をモデル化したものである。

表1 表のモデル

$X \backslash Y$	Y_1	Y_2	Y_3		Y_j	
X_1	r_{11}	r_{12}	r_{13}		r_{1j}	
X_2	r_{21}	r_{22}	r_{23}		r_{2j}	
X_i	r_{i1}	r_{i2}	r_{i3}		r_{ij}	

また、対象問題として、小学生が理解できる程度の表に限定し、小学校の教科書や参考書、問題集などから表を収集したものである。評価する。

2.2 表理解の定義

本稿では、表の理解を以下の5つで定義する。

- 1) 表を構成する要素概念を識別できる
- 2) 論理関係が識別できる
- 3) 言葉や単語の意味理解ができる
- 4) 検索質問文に対して、表の中からサブセットを抽出できる
- 5) 構成要素概念間の意味的論理関係が理解できる

1)は、表のタイトルや欄、項目やデータ、また合計や平均などの統計を表す部分やそのほかの一般概念をそれぞれ区別できることである。

2)の論理関係は表1においての、

$$X_i \in X \quad Y_j \in Y \quad X * Y = R$$

$$r_{ij} \in R \quad r_{ij} \in X_i \quad r_{ij} \in Y_j$$

を表す。“*”はXとYの意味から決まる論理関係とする。

3)では、表の中に存在するあらゆる言葉や単語、またその後の検索質問文における単語を、表記的ではなく意味的に理解する必要があることを表す。この部分が、表を“知的に”理解させている重要な点だといえる。

4)におけるサブセットとは、2)で示した、 $X * Y$ のことであり、表を参照するにあたって最も基本的と考えられる、3点をセットにしたものである。表を理解できるということを客観的に捉えるには、このようにサブセットを取り出すことを可能にすることが必要である。

5)の表構成要素概念間の意味的論理関係というのは、データとデータの意味的關係やデータと項目、

項目同士の関係を表す。例えば、「合計」という言葉があれば、そのデータ部分の和がそのセルには格納されており、「売り上げ」のデータを「値段」のデータで割れば「個数」のデータが理解できるということを表す。

2.3 システムの概要

上記のような表理解を踏まえて、表理解システム^[1]の概要を図 1 に示す。

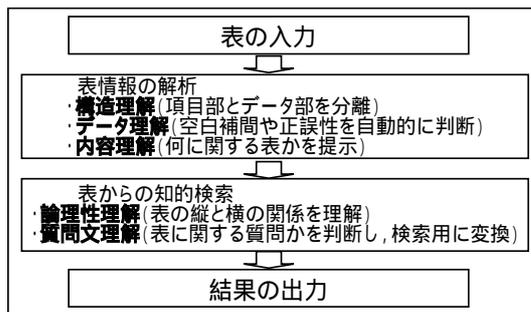


図 1 システムの概要図

本システムは、表計算ソフト Excel のセルによって表型に格納された情報を、コンピュータに与えて理解させている。これは将来的な展望として、知能ロボットに表を理解させるということを見据えているため、ロボットが見た画像の中から表部分のみを取り出し、そこからその表のデータを OCR 的に Excel のセルに変換することを想定しているためである。

システムの中核部分は大きく 2 つに分かれる。表がもつ情報をコンピュータが解析する部分と、表に対して尋ねた質問の答えを、コンピュータが取り出してくれる知的検索の部分である。2.2 節で表した定義を用いるならば、前者は 1) ~ 3) を、後者は前者で得た情報と共に 3) ~ 5) を行っているといえる。

3. 未知語処理

各処理を行う上で様々な知識ベースを用いるが、知識ベースには代表的な語のみを格納している。しかし、会話や文章で使用される語は様々であり、同じ意味であっても異なる表現で表される語が多々ある。会話や文章で用いられる語で知識ベースに格納されていない語は、未知語として扱い、概念ベースや関連度計算を用いて代表語の中で最も関連の強い語に置換する処理を行う。これを未知語処理という。この処理によって、ごく少量の知識から様々な語に対応することが可能となる。

3.1 概念ベース

概念ベース^{[2][3]}とは、ある単語（概念）とその意味的特徴を表す属性と重みの集合で構成されたものである。概念ベースには、約 9 万語の概念が格納されており、総属性数は約 254 万個である。属性数は概念によって異なるが、1 概念あたりの平均属性数は約 29 個である。ある概念 A に対して、その語の i 番目の属性を a_i 、重みを w_i 、概念 A の属性数を N 個とすると、概念 A は以下のように表される。

$$A = \{(a_1, w_1), (a_2, w_2), \dots, (a_N, w_N)\}$$

ここで、属性 a_i を概念 A の一次属性という。一次属性 a_i を一つ概念と見なせば、 a_i からさらにその一次属性を導くことができる。 a_i の属性 a_{ij} を概念 A の二次属性という。このように概念 A は n 次属性まで導くことができる。

3.2 関連度計算

関連度^[4]とは概念間の関連性の強さを定量的に表現したものであり、0 から 1 までの実数で表現される。概念間の関連性が強ければ関連度の値は大きく、弱ければ関連度の値は小さい。関連度が 1 になるのは二つの概念が同一のときだけである。関連度には、概念の表記特徴を利用した漢字関連度や、概念の属性の一致と重みを用いた意味関連度がある。本研究では、後者の意味関連度の計算方法（重み付き概念連鎖関連度計算方式）を使用した。

2 つの概念 A 、 B のそれぞれの一次属性を a_i 、 b_j とし、重みを u_i 、 v_j とすると、概念 A と概念 B は次のように表される。

$$A = \{(a_i, u_i) \mid i = 1 \sim L\}$$

$$B = \{(b_j, v_j) \mid j = 1 \sim M\}$$

ここで、 L は概念 A の属性数、 M は概念 B の属性数である。概念 A 、 B の重み付き一致度 $MatchW(A, B)$ は以下のように表される。

$$MatchW(A, B) = (s_A / n_A + s_B / n_B) / 2$$

$$s_A = \sum_{a_i=b_j} u_i \quad s_B = \sum_{a_i=b_j} v_j$$

$$n_A = \sum_{i=1}^L u_i \quad n_B = \sum_{j=1}^M v_j$$

重み付き一致度 $MatchW$ は、概念 A から見たとき

の概念 B の属性と一致した属性の重みの割合と、概念 B から見たときの概念 A の属性と一致した属性の重みの割合との平均である。

次に属性数が少ない方の概念を A ($L \leq M$) とし、概念 A の属性の並びを固定する。

$$A = \{(a_1, u_1), (a_2, u_2), \dots, (a_L, u_L)\}$$

概念 B の各一次属性を対応する概念 A の各一次属性との重み付き一致度 $MatchW$ の合計が最大になるように並び替える。二つの概念の一次属性は一対一対応させるために対応にあふれた概念 B の一次属性は無視する。

$$B_x = \{(b_{x1}, v_{x1}), (b_{x2}, v_{x2}), \dots, (b_{xL}, v_{xL})\}$$

二つの概念の重み付き概念連鎖関連度 $ChainW(A, B)$ は以下ようになる。

$$ChainW(A, B) = (s_A / n_A + s_B / n_B) / 2$$

$$s_A = \sum_{i=1}^L u_i MatchW(a_i, b_{xi}) \quad n_A = \sum_{i=1}^L u_i$$

$$s_B = \sum_{i=1}^L v_i MatchW(a_i, b_{xi}) \quad n_B = \sum_{j=1}^M v_j$$

4. 判断知識と常識判断システム

4.1 判断知識

4.1.1 シソーラス

シソーラス^[5]とは、一般名詞の意味的論理関係を木構造で表現したもので、約 13 万語が登録されている。

4.1.2 同義・反語辞書

同義・反語辞書^[6]とは、同義と考えられる語句を約 20 万組、反語と考えられる語句を約 1 万 7 千組格納されたものである。

4.2 常識判断システム

常識判断システム^[7]とは、曖昧な表現や抽象的な表現を受け取っても、人間の持つ様々な常識を用いて適切に判断することができるメカニズムである。常識には、量・数・時間・場所・感覚・感情に関するものを作っている。

4.2.1 感覚判断システム

感覚判断システム^[8]とは、「林檎は赤い」といったように、ものの特徴に関する常識を理解できるものである。語を入力するとその語から想起される形容詞・形容動詞が出力される。

4.2.2 時間判断システム

時間判断システム^[9]とは、「クリスマスは 12 月 25 日」「夏は 6 月から 9 月」といったように、時

間に関する常識を理解できるものである。時間に関する語を入力すると、期間や時間軸が出力される。

5. 表の解析

表を得たコンピュータが、その情報を取り出すためには、構造やデータの論理関係を理解しなければならない。そこで、

『構造理解』：得られた表から、その構造を正しく理解する。

- セルの集合体で入力された表の項目部分、データ部分、また統計語などが書いてある部分をコンピュータが判断し、境界やその範囲を正しく区分けする。

『データ理解』：データの論理関係を判断し、論理矛盾を検出・訂正する。

- 項目部とデータの内容、データの内容同士の関係性を明確にし、また理解させることで、誤りなどを検出する。

『テーマ理解』：表の構造やデータから、それが何の表であるかということ判断する。

- 表の意味や構造から、その表がどのような表（「 \quad に関する \quad の表」）であるかということを理解する。

の三つのことを行うことができれば、表を理解できていると考えてよいということになる。これはすなわち、与えられた表に対して、人間であれば言葉として表すことのできることを、コンピュータが同様に言葉として表すことができれば、コンピュータが表を理解していることと同義となる。

5.2 表の構造理解

コンピュータが表を理解する上で、まず必要となるのが、構造を理解することである。ただのデータの配列でしかない状態では、表を参照することも、意味を理解することもできない。抜き出された表部分のうち、どの部分が項目部分で、どの部分がデータ部分であるかを理解する必要がある（表 2）。

表 2 項目部分とデータ部分のモデル

	Y_1	Y_2		Y_j		← 項目部
X_1	r_{11}	r_{12}		r_{1j}		
						← データ部
X_i	r_{i1}	r_{i2}		r_{ij}		

RDB の検索などでは、項目部はフィールドとして、データ部分はレコードとしてすでに定義されているが、表の理解ではこの判断が必要となる。

項目部分とは、表が表しているものの意味的に抽象的な、また代表的な名称、すなわち項目が羅列している部分である。この部分は表の最も左、または最も上に現れやすい。データ部分とは、その表、または項目部分の具体的な内容を表している部分で、表の中で最も重要な中心の場所である。そして、統計語の部分とは、統計語とそのデータの部分であり、項目部やデータ部の両方に属することがある。

5.3 表のデータ理解

従来のデータベースを扱う研究と違い、入力された表は、その真偽が確かではない。構造を理解し項目部分とデータ部分を分けた中での論理関係の真偽を確認することで、正しくその表を理解できているといえる。その論理関係性を確認するためには、2.2 節における 2) で表す関係をチェックすることが必要となる。

表 3 項目やデータの関係性

好きな果物	人数	相対度数
バナナ	13	0.26
蜜柑	10	0.20
スイカ	27	0.54
合計	50	1.00

項目部分とデータ部分の論理関係は、抽象的な名称と、その具体的な内容である。すなわちその項目が具体名詞であればシソーラスにおける上位と下位の関係であるべきであり、また抽象名詞であれば、その抽象名詞の内容に合った名詞や形容詞、数詞、または特別な場合は動詞でないといけない。

それぞれの項目に対して、対応するデータ部分を取り出し、順に論理関係を見ていき、その際に、項目名が具体名詞である場合、抽象名詞である場合の二パターンに分け、それぞれ別の処理を行う。

項目名が具体名詞の場合は、データ部分の語それぞれの、シソーラスのルーツ(家系図)を取り出し、そのどこかの親ノードと項目名が同義であれば正しいという確認作業を行う。項目名がデー

タに数詞を扱うような(例えば『人数』のような)抽象的な語であれば、データ部分のデータ全てが、その内容に合った数字であるか、またはその抽象語を形容する形容詞(例えば項目名が『人数』ならば『多い』や『少ない』といった語)であるかどうかを確認する。もし間違っていればその検出を行う。

統計語が表の中に存在すれば、直交するデータの部分がその統計語の意味を持っている。

コンピュータは表内に統計語が存在することが分かっていても、その意味は理解することができないため、統計語の表記自体にその意味を持たせる必要がある。そこで統計語知識ベース(表 4)を作成している。レコード数は 10 件である。

表 4 統計語知識ベース

統計語	定義
合計	数の和
平均	合計 ÷ 個数
.	.
.	.
.	.
相対度数	数 ÷ 全体数

この知識ベースにより、統計語の表記がその意味を持ち、表内で関係性を表す記号となる。例えば、表 3 のように『合計』という言葉があれば、横に並んでいるデータ部分の『50』という値は合計を表す。それは、その『合計』 - 『50』というラインと直交するデータ部分、すなわち『13』と『10』と『27』の和であるということを理解する。

表のデータ理解は、データ部分のひとつひとつに対して、属する項目や、直交する他のデータとの関係性をあきらかにすることである。これらを正しく行うことで、正しく信頼性のある表の知識を得ることができ、また次の工程である表の内容判断を正確に行うことの足がかりとすることができる。

5.4 表のテーマ理解

入力された表の構造が理解できれば、その項目部とデータ部それぞれの関係性や意味内容から、表自体のテーマ・題材を判断することができる。人間であれば、その表を見ると、テーマが『 に関する の表』というように言うことができ

るが、これをコンピュータに自動で判断させる。すなわち表種の自動判別である。

具体的な方法として、先ほどの『 』に関する表』という表現のうち、『 』という部分と『 』という部分で別の処理をすることによって、実現を可能とする。つまり、
項目の関係性
データ同士の関係性
からみつけることが可能となる。

一般的な表は、データベース、統計表、分布表、タイムテーブル・スケジュール表、そして順位・順番表などに分かれるので、 にはこれらが該当する。この判別は表構造理解で分けることができたデータ部分の組み合わせによって行う。例えば、データ部分が、『範囲を表す数』と『数』から構成されていれば、その表は『分布表』であると判断することができる。このようにデータ部分の組み合わせパターンを表種知識ベースとして持つことで判断させる。表 5 は表種をまとめた知識ベースである。レコード数は 7 件である。

表 5 表種知識ベース

データ形態 1	データ形態 2	表種
範囲を表す数	数	分布表
時間名詞	事象	タイムテーブル スケジュール表
もの	統計語を含む 数	統計表
.	.	.
.	.	.
.	.	.

また、時間割表のような形式の表では、データ部分が一かたまりになっているため、そのデータ形態によって、表 5 と同様に知識ベースに格納して表種を指定している。例えば、データ部分が『値段を表す数』で形成されていると、それは『料金表』であると判断することができる。

これに対し、 の部分には、項目部分の内容から判断させる。例えば、項目が『氏名』、『年齢』、『住所』、『電話番号』から成っていれば、人間であればその表は『住所』に関する表である、というように、その項目群の中から最も重要な表のキーワードを得ることができる。

ここで項目群の中で最も重要な語を抜き出す

参考として、語の特定性が高い単語は項目群の中でも、表の中心となる単語になりやすい。他の項目すべてとなるべく深い関係を持つ語が、表題になりやすい項目である、ということが挙げられる。これらの仮定のもとに、項目群の中から最も重要な表のキーワードを選び出すことをコンピュータに行わせるために、本稿では *idf* (inverse document frequency) と、その他の語との関連度を用いる。

idf は語の特定性を表すための尺度として知られていて、文書集合全体にわたる分布を考慮して決定される重みを表し、特定の文書に集中して出現する語に対して大きな値が与えられる。*idf* は次の式によって求めることができる。

$$idf(\text{語}) = \log\left(\frac{N_{All}}{N(\text{語})}\right)$$

概念ベースにおける *idf* では、この式においての $N(\text{語})$ とは、概念ベース全ての概念の中で、三次属性まで展開した中にその語が含まれる概念の数を表し、 N_{All} は概念ベースの全ての概念の数である、87242 語が入る。

次に、参考の を考慮して、項目群の中で、他の語との関連度が高いものを選ぶ。項目一つに対し、その他の項目それぞれに対する関連度を求め、その平均を出す。この平均値が、その語の表全体に対する関係性を表す値だということが分かる。

idf と、表全体に対する関係性の二つの値を掛け合わせると、その項目の表に対する重要性を得ることができる。この値が高いものが項目として重要な部分なので、 にこの項目名が入る。

6. 表から知的検索

表の構造やデータを理解できれば、その表に関する質問にも答えられると考えられる。

知的検索の大きな流れとしては、

1. 質問文の意味理解：質問対象語と条件の取得
2. 表理解における質問文意味理解：質問対象語と条件を表に合わせて形式化
3. 表理解における質問文意味解釈：表から答えを抽出

となっている。

6.1 質問文の意味理解

質問文が与えられたとき、それに答えるために、

その質問文が何を求め、その条件はどのようなことであるかを理解しなければ答えを返すことができない。そこで、質問文から質問対象語（質問文が答えを求めている対象の語）とその条件を取得することを質問文の意味理解^[10]という。例えば、『赤い果物の値段の合計はいくらですか？』の場合、質問対象語は『合計』で、条件は『赤い果物の値段』となる。

構文解析ソフト ChaboCha^[11]を用いて、『？』に係る語や疑問詞、またその語の位置関係によって質問対象語とその条件を取得している。

6.2 表理解における質問文意味理解

『表理解における質問文意味理解』とは、取得した質問対象語と条件を表形式に形式化することと定義する。与えられた表から、質問文を形式化する表の構成は図 2 のようにする。

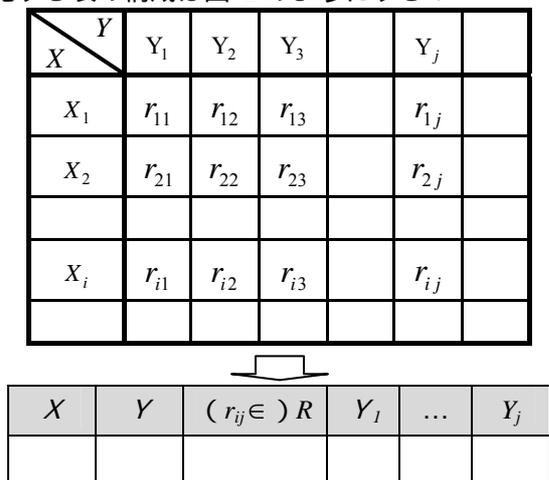


図 2 形式化の表

まず、質問対象語からどの部分から検索するかを調べ、形式化の表の該当する部分に『？』を格納する。次に、質問対象語の条件を自立語に区切り、質問対象語と同様にして、各自立語を形式化の表の該当する部分に格納する。(図 3)

図 3 の形式化の表において、『データ』とは図 2 の R の部分（データ部分）のことである。

具体的な処理について説明する。

まず、質問対象語と形式化の表の各項目と関連度計算を行い、最も関連度が高くなった項目の部分に『？』を格納する。

次に、質問対象語の各条件と形式化の表の各項目と関連度計算を行う。関連度の性質上同義とみなされる閾値 0.27 以上であれば、それは形式化の表の項目自身とし、0.27 未満であれば、形式化

の表に格納するものであると判断する。

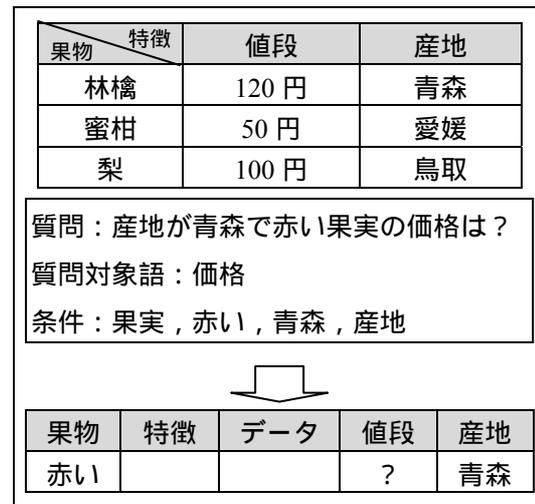


図 3 形式化の例

図 3 の場合、条件の一つ目の『果実』は項目の『果物』と同義であり、条件の二つ目の『赤い』は 0.27 未満なので、『果物』の部分に格納される。

このように、質問文を形式化できれば、表理解において質問文を意味理解できたとみなすことができる。

6.3 表理解における質問文意味解釈

質問文を形式化した後、実際に表から答えを抽出することを『表理解における質問文意味解釈』と定義する。

各条件の語句を順番に表の一番上の列から表記一致で検索する。もし、見つければその列の『？』の格納されている項目の値を答えの候補として表から取得する。表記一致で見つからなければ、同義語辞書や関連度計算によって同義のものを検索し、あれば表記一致の時と同様にする。また、『赤い』や『丸い』のような感覚語が条件の場合、感覚判断システムを用いて、その感覚語が出力されれば、同様にする。

もし、質問対象語が合計や平均といった統計語の場合、条件から取得した答えに統計処理を行い、答えを返す。

7. 評価

上記で述べた表からの知的検索について、評価実験を行った。

図 3 で表したような Web や数学のテキストから

集めた 15 件の表に対して、

- 1) 表に記載されている語句のみで構成された質問文
- 2) 表に記載されていない語句も含まれている質問文

をそれぞれ 100 文ずつ入力し、正しく答えを返すか評価した。結果を図 4 に示す。左が 1) の結果で 95% の正解率を得た。右が 2) の結果で 60% の正解率を得た。

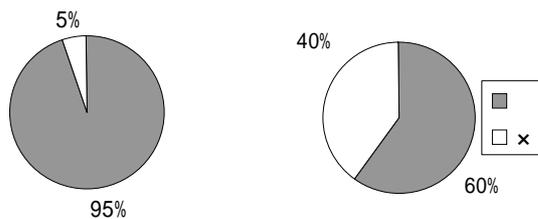


図 4 評価結果

表に表記されている語句だけで質問するとほぼ正しく答えることができた。また、表に表記されていない語句を用いても、関連度計算や常識判断システムによって連想することで 6 割の正解率を得た。

1) の失敗例としては、合計や平均などの統計語が含まれている表において、統計語の列を他のデータの列と同様の扱いになってしまったものがあった。

2) の失敗例としては、動詞の処理がうまくできていないものや、記号と動詞の対応づけがうまくできていないものがあった。

動詞の処理や対応づけをうまくできれば、かなりの正解率の向上を期待できると考えられる。

8. おわりに

本稿では、人間と同じようにコンピュータに表を理解させる手法を提案した。表の解析と知的検索の二つの部分から、表を理解させることを実現した。

表の解析では、構造理解・データ理解・内容理解によって表を理解することを実現した。知的検索では、概念ベース等を用いて様々な質問文を表形式に形式化し、答えを取得することを実現できた。

今後は、四角形ではない表や項目が上端や左端以外にある表のような、特殊な表にも対応していく必要がある。

謝辞

本研究は文部科学省からの補助を受けた同志社大学の学術フロンティア研究プロジェクトにおける研究の一環として行ったものである。

参考文献

- [1] 権 東旭, 古川 成道, 渡部 広一, 河岡 司, “常識判断に基づく知能ロボットのための表理解システム”, 信学技報, NLC2004-11, pp.1-6, 2004.
- [2] 眞鍋 康人, 小島 一秀, 渡部 広一, 河岡 司, “概念間の関連度やシソーラスを用いた概念ベースの自動精錬法”, 同志社大学理工学研究報告, Vol.42, No.1, pp.9-20, 2001.
- [3] 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, pp.1272-1283, 1997.
- [4] 渡部 広一, 河岡 司, “常識的判断のための概念間の関連度評価モデル”, 自然言語処理, Vol.8, No.2, pp.39-54, 2001.
- [5] NTT コミュニケーション科学研究所監修, 「日本語彙体系」, 岩波書店, 東京, 1997.
- [6] 小島 一秀, 渡部 広一, 河岡 司, “連想システムのための概念ベース構成法-属性信頼度の考え方に基づく属性重みの決定”, 自然言語処理, Vol.9, No.5, pp.93-110, 2002.
- [7] 土屋 誠司, 小島 一秀, 渡部 広一, 河岡 司, “常識的判断システムにおける未知語処理方式”, 人工知能学会論文誌, Vol.17, No.6, pp.667-675, 2002.
- [8] 渡部 広一, 堀口 敦史, 河岡 司, “常識的感覚判断システムにおける名詞からの感覚想起手法”, 人工知能学会誌, Vol.19, No.2, pp.73-82, 2004.
- [9] 土屋 誠司, 奥村 紀之, 渡部 広一, 河岡 司, “連想メカニズムを用いた時間判断手法の提案”, 自然言語処理, Vol.12, No.5, pp.111-129, 2005.
- [10] 古川 成道, 渡部 広一, 河岡 司, “概念ベースを用いた知的検索における曖昧な質問文の意味理解”, 人工知能学会全国大会, 2D1-10, 2004.
- [11] Taku Kudo, Yuji Matsumoto, “Fast Methods for Kernel-Based Text Analysis”, ACL 2003 in Sapporo, Japan, 2003.
<http://cl.aist-nara.ac.jp/~taku-ku/software/cabocho/>