

レイアウトとコンテンツの知識を用いた仮説駆動型 文書画像理解

黄瀬浩一† 百田賢一††* 杉山淳一††**
馬場口登††† 手塚慶一††

印刷文書に記録された情報を計算機を介して有効利用するためには、文書中の文字を認識し、論理構造により構造化するという文書画像理解が不可欠となる。一般に、文書の論理構造は、章、節などの論理オブジェクトの木構造として表現され、文書のレイアウトとともに、コンテンツとも深い関連性を持つものである。したがって、文書画像理解のロバスト性を向上させるためには、いずれか一方ではなく、両者を知識として蓄え、協調的に使用することが望ましい。本論文では、仮説駆動型の処理戦略の導入により、この目標の達成を試みる。本手法では、まず文書のレイアウトに関する知識を用いて、論理オブジェクトの領域を抽出する。ここで、抽出結果に複数の可能性が残る場合には、互いに矛盾する仮説として生成する。生成された仮説は、依存関係を保持する依存関係データベースにより記録・管理され、処理の制御に使用される。また、処理の途中で発見された仮説の矛盾は矛盾データベースに記録され、探索空間の削減に用いられる。次に仮説の依存関係、矛盾を考慮しつつ、コンテンツに関する知識を用いて仮説を検証する。本手法では、単語の接続性、単語列の矛盾という二つの概念を導入し、知識を表現する。縦書き名刺100枚を対象とした仮説生成検証実験から、文字の抽出率93.0%、論理オブジェクトの抽出率92.6%、コンテンツの同定率86.8%を得、本手法の有効性を確認した。

Hypothesis Driven Document Image Understanding with Knowledge about Layout and Content

KOICHI KISE,† KEN-ICHI MOMOTA,††* JUN-ICHI SUGIYAMA,††**
NOBORU BABAGUCHI††† and YOSHIKAZU TEZUKA††

Document image understanding is a task to extract logical objects (e. g. chapter, title) laid out in a document image. One of the central issues of understanding is how to deal with uncertainty of extraction. To cope with this problem, we utilize the strategy of hypothesis generation and verification. All possible candidates of logical objects are generated based on the knowledge about layout of a document, as hypotheses, and tested using the knowledge about validity of content for logical objects. This process is controlled based on the logical dependency and inconsistency between hypotheses using *dependency database* and *inconsistency database* mechanisms. From the experimental results for 100 visiting cards, we confirm our method is effective and robust for understanding of documents with complicated structure.

1. はじめに

印刷文書として蓄積された大量の情報を有効利用す

るためには、計算機への自動入力が必要である。文書の持つさまざまな情報のうち、テキスト部分の担う情報(テキスト情報)に限って考えても、柔軟な検索、知的通信、再編集など、高度かつ広範囲に及ぶ利用形態を可能とするためには、テキスト情報を単に抽出するだけではなく、構造化する必要が生じてくる。例えば、対象が図1に示す名刺の場合には、図2(a)のような文字を羅列したようなテキスト情報ではなく、図2(b)のような構造化したテキスト情報を得る必要がある。ここでは、図2(b)のような構造化されたテキスト情報を文書のコンテンツ(content)と呼ぶことにする。一般に、文書のコンテンツの構造は論理構造と

† 大阪府立大学工学部
College of Engineering, University of Osaka
Prefecture

†† 大阪大学工学部
Faculty of Engineering, Osaka University

††† 大阪大学産業科学研究所
I.S.I.R., Osaka University

* 現在, NTT (株)
Currently with NTT Corp.

** 現在, 富士通研究所(株)
Currently with Fujitsu Laboratories

いわれ、論理オブジェクトをノードとする木構造として表現される¹⁾。このような構造化を含めた処理は、文書画像理解と呼ばれることが多い。

さて近年、文書画像理解の一環として、文書画像の構造解析あるいは文書画像解析と呼ばれる手法が、活発に提案されている^{2)~5)}。これは、論理オブジェクトがどのようにレイアウトされるかについての知識（レイアウトに関する知識）を用いて文書画像を解析し、論理オブジェクトに対応する領域を抽出する処理である。この処理により領域が正しく抽出可能ならば、領域内の文字を認識することにより、構造化したテキスト情報を得ることができる。しかし、文書のレイアウト構造にさまざまな変動がある場合には、抽出結果が一意に定まらないという問題点がある⁶⁾。

一方、論理構造はコンテンツの構造であるため、コンテンツにより特徴付けることもできる。名刺を例にとると、住所、肩書という論理オブジェクトは、それぞれ住所、肩書を表す文字列、例えば“大阪府吹田市…”や“課長”などに対応する。そこで、文書のレイアウトに関する知識に加え、論理オブジェクトがどのようなコンテンツを持ち得るかに関する知識（コンテンツに関する知識）を利用することにより、上記の問題を解決するというアプローチが考えられる。しかしこのためには、コンテンツの考慮に伴う処理量の増大にどう対処するのか、またレイアウトとコンテンツという異質の知識を処理上でどのように融合するのか

など、新たな問題を解決する必要が出てくる。

この問題を解決する一方策は、文書画像の構造解析により得られた結果を仮説として扱い、コンテンツに基づいて検証するという仮説駆動型の処理戦略を導入することである⁶⁾。仮説駆動型の処理では、仮説の検証という目的のみにコンテンツを考慮すればよいことから、処理量の軽減が期待でき、また2種類の処理を自然に融合することも可能となる。駱らは、同様の観点から図書目録カードを対象とした処理手法を提案している⁷⁾。しかしこの手法は、論理オブジェクトごとに詳細に定められた制御知識を用いて仮説を検証するため、対象文書ごとに制御知識を変更しなければならず、汎用性に問題点が残されていると推察される。

そこで本論文では、汎用性の向上を目的とする仮説駆動型の文書画像理解手法を提案し、名刺を対象とした実験から本手法の有効性を検討する。本手法の特徴は、生成された仮説を依存関係に基づいて管理し、さらに仮説間の矛盾を発見することにより効率的に検証する点にある。このとき、仮説の依存関係と矛盾は、依存関係データベースおよび矛盾データベースにより統一的に管理される。このことにより、処理全体の見通しが良くなるほか、制御知識の追加、変更を考慮する必要がないため、手法の汎用性が向上すると考えられる。

“大阪電気株式会社 総務部 課長 大阪太郎 支社
大阪府豊中市… 電話… 本社 大阪府吹田市… 電話…”

(a) 構造化されていないテキスト情報

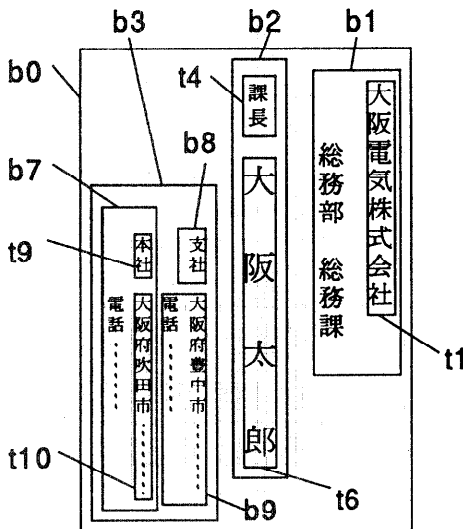
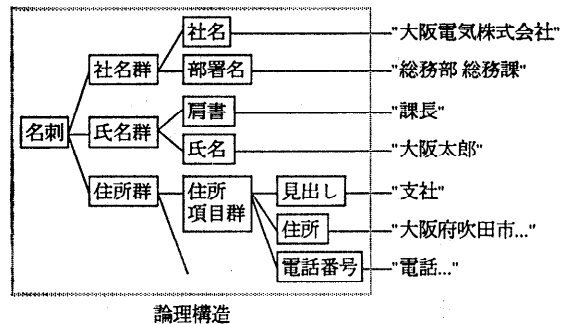


図1 名刺

Fig. 1 An example of a visiting card.



論理構造

(b) 構造化されたテキスト情報(コンテンツ)

図2 名刺のコンテンツ

Fig. 2 Content of a visiting card.

(a) Unstructured text information.

(b) Structured text information.

2. 文書画像理解のための知識

まず、本手法で用いるレイアウトに関する知識、コンテンツに関する知識について簡単に述べる。

2.1 レイアウトに関する知識²⁾

レイアウトに関する知識とは、文書画像の構造解析に用いる知識である。図1のような縦書き名刺に対して、本手法では、図3に示すように、対象文書の論理構造、レイアウト構造、および両者を結合する対応規則により知識を表現する。

(1) 論理構造

論理構造は、氏名、社名などの論理オブジェクトから構成される木構造である。図3(c)に示すように、名刺に記述されている情報は、個人に関する情報(氏名群)、個人の所属に関する情報(社名群)、所属の連絡先に関する情報(住所群)の3種類に分類できる。住所群には、本社、支社などに対応した一つ以上の連絡先(住所項目群)が記述される。ここで、本社、支社などを住所項目群の見出しと呼ぶ。また、論理構造の葉となる論理オブジェクトを項目と呼ぶ。

(2) レイアウト構造

レイアウト構造は、矩形領域(レイアウトオブジェクト)の包含関係を表す木構造である。本手法では、レイアウトオブジェクトとして、ブロック領域、文字列領域の2種類を考えている。ここで、ブロック領域とは、複数の文字列領域を囲む矩形領域を指し、文字列領域とは、縦、横のいずれか一定方向に並ぶ文字を囲むような矩形領域をいう。図3(a)では、*b*はブロック領域、*t*は文字列領域を表す。また、レイアウトオブジェクト間の包含関係は、縦、横のいずれか一定方向に並ぶ子のレイアウトオブジェクトを、親のレイアウトオブジェクトが囲むように設定される。

本手法では、レイアウトオブジェクト間の部分全体関係、類似差異関係と呼ぶ関係記述を用いて、レイアウト構造を記述する。部分全体関係とは、レイアウトオブジェクト間の包含関係を規定する関係記述である。一方、類似差異関係とは、二つのレイアウトオブジェクト間の相対的な配置等を表す関係記述である。これらの関係は、矩形領域の縦幅、横幅などを基本として定められた特徴量、およびセンタリング、揃えなどのレイアウトを規定する記号の2種類により記述される。

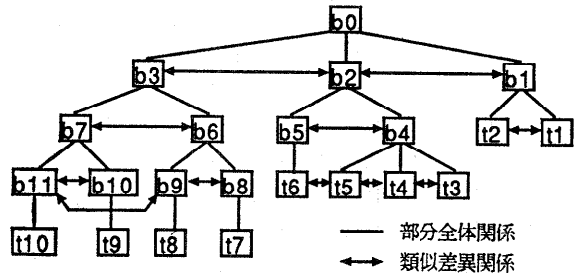
レイアウト構造を知識として記述する際の最も重要な点は、レイアウト構造の変動にいかに対処するかで

あると考えられる。本手法ではレイアウト構造の変動を、レイアウトオブジェクトの個数変動、形状変動、関係変動に分類して表現する。個数変動は区間値を用いて表す。例えば、図3(a)の*b8*は、名刺によって存在する場合と存在しない場合があるため、個数変動は[0, 1]となる。一方、形状変動、関係変動は、それぞれ部分全体関係、類似差異関係の記述として表現される。具体的には、特徴量の取り得る値の範囲を区間値として表す。

その他に記述される知識には、レイアウトオブジェクトが並ぶ方向(横または縦)と順序などがある。

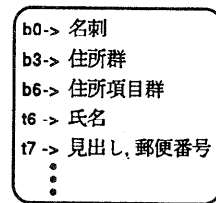
(3) 対応規則

対応規則は、レイアウトオブジェクトから論理オブジェクトを導くルールとして表現される。一般に、こ



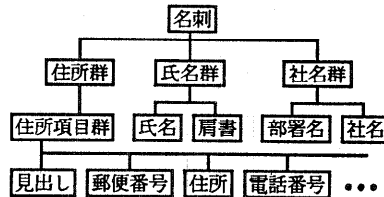
(a) レイアウト構造

(a) Layout structure.



(b) 対応規則

(b) Rules of correspondence.



(c) 論理構造

(c) Logical structure.

図3 レイアウトに関する知識
Fig. 3 Knowledge about layout.

のルールは、

$$x \Rightarrow y_1, \dots, y_n \quad (1)$$

の形式を持ち、レイアウトオブジェクト x が論理オブジェクト y_1, \dots, y_n を含む可能性があることを意味する。ただし、必ずしもすべての y_i を含む必要はないとする。例えば、図 3 (b) の b_3 は住所群と 1 対 1 に対応する。また t_7 は、見出し、郵便番号のいずれか一方に対応する場合もあり、両方を含む場合もある。

2.2 コンテンツに関する知識

図 2 (b) に示したように、名刺のコンテンツは、名刺に記述されたテキスト情報が、名刺の論理構造により構造化されたものである。ここで、名刺以外の論理オブジェクトについても、その論理オブジェクトを根とする部分木を考えると、同様にコンテンツを定めることができる。コンテンツに関する知識とは、各論理オブジェクトがどのようなコンテンツを持ち得るかを規定した知識であり、単語の接続性、単語列の矛盾という二つの概念に基づいて記述される。

(1) 単語の接続性⁹⁾

論理構造の葉となる論理オブジェクト (項目) のコンテンツの妥当性を表現するために、単語の接続性を導入する。ある二つの単語が読み順に沿って意味的に接続するとき、これらは接続性を持つという。すると、項目の妥当なコンテンツは、接続性を持つ単語の並び、すなわち単語列として表現できる。例えば、住所のコンテンツ “大阪府吹田市山田丘 2-1” は、単語

と接続性により、

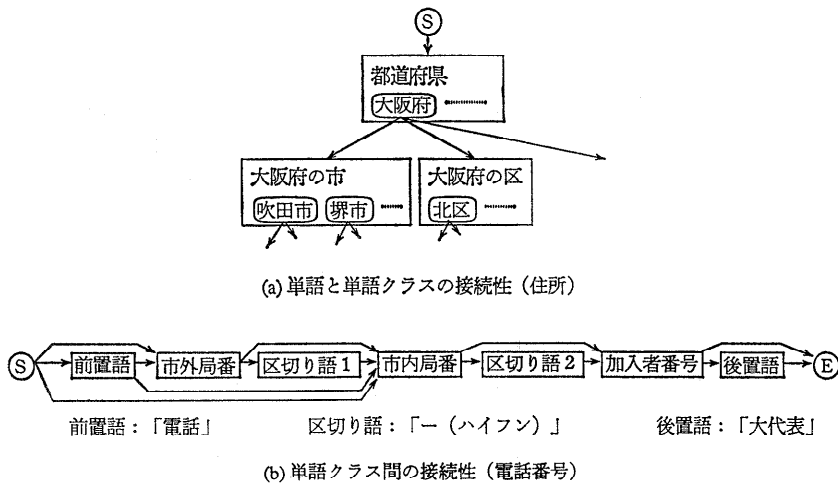
$$\textcircled{S} \rightarrow \text{「大阪府」} \rightarrow \text{「吹田市」} \rightarrow \text{「山田丘」} \rightarrow \text{「2」} \rightarrow \text{「-」} \rightarrow \text{「1」} \rightarrow \textcircled{E}$$

のように表すことができる。ここで、単語を「」、接続性を \rightarrow により表現している。また、 \textcircled{S} 、 \textcircled{E} は、単語列の始端、終端を表す。実際には、適当な単語のクラス (単語クラス) を設定し、接続性を記述している。住所の場合は単語が木状に接続することから、図 4 (a) に示すように単語と単語クラス間に接続性を記述する。その他の項目では、図 4 (b) に示す電話番号の例のように単語クラス間の接続性を用いる。

(2) 単語列の矛盾¹⁰⁾

項目を単語列に対応付けると、他の論理オブジェクトのコンテンツは、単語列の組として表すことができる。その性質としては、単語列の妥当な組合せが制限されることがある。例えば、名刺には「部長」と「課長」という二つの肩書が同時に存在することはない。本論文では、このような単語列の関係を単語列の矛盾と呼ぶ。

単語列の矛盾は、単語列がどの論理オブジェクトに属するかに応じて変化する場合がある。例えば、同じ住所項目群に「東京都」、市外局番「06」が存在する場合は矛盾である。しかし本社と支社など、異なる住所項目群にあれば矛盾ではない。そこで本手法では、単語列の矛盾を論理オブジェクトごとに記述する。実際には、論理オブジェクトに含まれる項目の対を対象



(a) 単語と単語クラスの接続性 (住所)

(b) 単語クラス間の接続性 (電話番号)

図 4 単語の接続性

Fig. 4 Connectivity between words.

(a) Connectivity between a word and a class of words.

(b) Connectivity between classes of words.

に、矛盾の場合を記述するか、逆に無矛盾の場合を記述するか、いずれか記述数の少ないほうを選択する。

3. 処理の概要

本手法における処理は、大きくレイアウト処理、コンテンツ処理の2処理に分割される。レイアウト処理では、構造解析、文字切り出し・認識の2処理が順に適用される。また、コンテンツ処理では、単語列生成、単語列検証の2処理が順に適用される。以下、これらの処理について概要を述べる。

構造解析では、レイアウトに関する知識を満たすレイアウトオブジェクトをすべて抽出する。処理はレイアウト構造の根から葉へと部分全体関係の木構造に基づいて進められる。すなわち、親のレイアウトオブジェクトが抽出されると、その領域を分割することにより、子のレイアウトオブジェクトを抽出する。以上の処理により得られたレイアウトオブジェクトは、対応規則の適用により論理オブジェクトに対応付けられる。本手法では、子が一意に定まらない場合、制御知識を用いて選択せず、すべてを仮説として保存・管理し、以後の処理結果を考慮して総合的に正誤を判断する。

構造解析が終了すると、次に、得られた文字列領域の仮説を対象として文字切り出し・認識を行う。文字切り出しでは、文字列領域から文字領域の仮説を生成する。一方、文字認識では、文字領域に対して文字カテゴリの候補（以後、候補文字と呼ぶ）を生成する。以上の処理により、文字に関する仮説が生成される。

単語列生成では、単語の接続性に関する知識を用いて文字の仮説を検証する。具体的には、接続性の記述を満たす単語列を生成することにより、単語列として成り立たない文字の仮説を除去する。この処理により生成された単語列は、項目のコンテンツとなる。

単語列生成により、ある項目に対して複数のコンテンツ（単語列）が生成された場合、適切なものを選択する必要がある。単語列検証では、単語列の矛盾に関する知識を用いて、生成された単語列の無矛盾性を検証し、単語列を選択する。単語列の矛盾を発見することは、同時に、単語列を生成する元となった仮説の検証を意味する。最終的には、名刺全体として無矛盾な仮説と単語列が得られる。仮説は論理オブジェクトに対応しているため、論理構造により構造化されたコンテンツが求められたことになる。

4. 仮説管理

4.1 仮説管理の必要性と概要

以上のような仮説駆動型の処理では、一般に、仮説を用いない場合に比べて、解を得るための探索空間が非常に大きくなる。文書画像理解の解は仮説の無矛盾な組合せとして表現されるが、仮説の依存関係や矛盾を考慮せずに、仮説の組合せを総当たりに試すと、組合せの爆発を起すため、処理効率の観点から望ましくない。また同時に、処理の一貫性維持の観点からも適切ではない。したがって、仮説駆動型の処理では、仮説の依存関係や矛盾の管理が重要な課題となる。

本手法の特徴は、1)対象文書に依存した制御知識を排し、仮説の依存関係と矛盾のみに基づいて、統一的に仮説を管理すること、2)レイアウト処理やコンテンツ処理を実行する処理部とは別に、独立した仮説管理部を設け、仮説を集中的に管理すること、の2点にある。具体的には、仮説の依存関係と矛盾は、依存関係データベースおよび矛盾データベースと呼ぶ2種類のデータベースにより管理される。1)により制御知識を変更する必要がなくなるため、システムの汎用性が向上する。また2)により、処理部は生成した仮説や発見した矛盾を仮説管理部に通知するだけでよく、システム構成が単純化される。

以下では、仮説の生成、検証の観点から、仮説管理の詳細について述べ、最後に他の仮説管理法との比較考察を行う。

4.2 仮説の生成と依存関係

図5に生成された仮説の具体例を示す。この図は、名刺の住所部分に対する構造解析の途中結果を表す。例えば住所項目群の仮説 h_a からは、三つの仮説が生成されている。このうち、 h_c は親仮説 h_a の領域が直接、子の領域に対応する子仮説であり、 h_d , h_e は h_a の領域を分割して生成された子仮説である。

文書画像理解の解は、仮説の無矛盾な組合せとして表現される。上記の三つの仮説についていえば、無矛盾な組合せは図5に示す $c_b=(h_a)$, $c_c=(h_d, h_e)$ の2組に限られる。換言すれば、 c_b , c_c など、一つの親仮説から生成される子仮説の無矛盾な組は、仮説を取捨選択する際の単位と考えられる。本論文では、このような仮説の組を選択肢と呼ぶ。また、上記の c_c のように、複数の仮説からなる選択肢を選択することは、選択肢の要素であるすべての仮説を正解と仮定し

ていることにはかならない。逆にいえば、選択肢が正しいためには、選択肢を構成するすべての仮説が正解でなければならない。すなわち、選択肢は仮説の論理積と同値である。

さらに、 $S_b = \{c_b, c_c\}$ など、一つの親仮説から生成された複数の選択肢は、親仮説の領域に対する異なる領域分割の結果を表すため排他的である。ここでは、このような選択肢の集合を排他集合と呼ぶ。本手法では、排他集合が網羅的 (exhaustive) であること、すなわち親仮説から得られる選択肢が排他集合に属するものに限られることを仮定している。この仮定は、親仮説が正解であるならば、排他集合の中に、必ず正解の選択肢が含まれていることを意味する。

本論文では、以上のような選択肢、排他集合を介した親仮説と子仮説の依存関係を、次のように表す。

$$h_p \rightarrow \{c_1, \dots, c_m\} \quad (2)$$

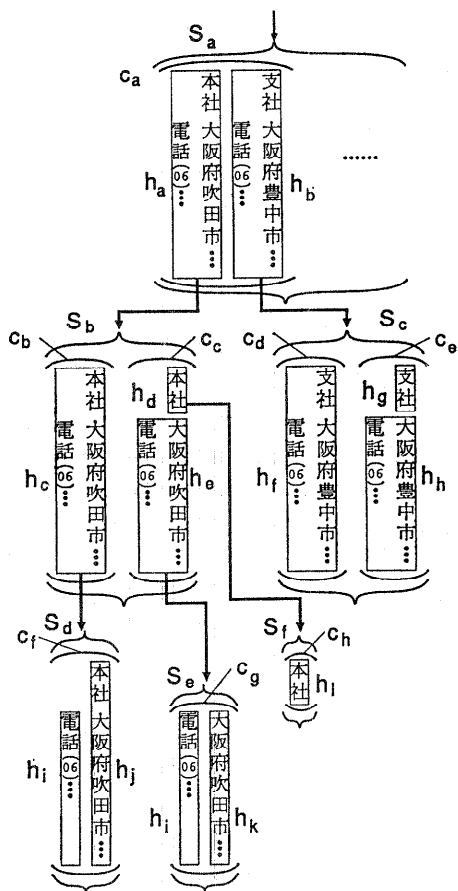


図5 仮説の依存関係

Fig. 5 Dependency between hypotheses.

$$c_i = (h_{i1}, \dots, h_{in}) \quad (3)$$

$$h_{ij} = \langle r, x, \{y_1, \dots, y_k\} \rangle \quad (4)$$

ここで、 h_p, c_i, h_{ij} は、それぞれ親仮説、選択肢、子仮説を表す。また、 r, x, y_i は、レイアウトオブジェクトの領域、レイアウトオブジェクト名、対応する論理オブジェクト名を表す。 $k > 1$ となるのは、2.1 節の (3) に述べた対応規則の右辺に、複数の論理オブジェクト名が記述されている場合である。

4.3 依存関係の記録形式

処理部で得られた仮説生成の結果、すなわち親仮説からどのような子仮説が得られたかは、式 (2) ~ (4) の形式により仮説管理部に通知される。仮説管理部では、通知された処理結果をもとに、仮説の依存関係を、次のように依存関係データベースに記録する。

まず最初に、選択肢、仮説に識別子を与える。選択肢に対しては、図5の c_a などの識別子を一意に割り当てる。一方、仮説に対しては、識別子を割り当てる際に、すでに同一の仮説が依存関係データベースに登録されているかどうかを検査する。すなわち、異なる選択肢に属する二つの仮説 $h = \langle r, x, Y \rangle, h' = \langle r', x', Y' \rangle$ において、 $r=r', x=x', Y=Y'$ ならば、 h と h' に同じ識別子を割り当てる。図5では、 h_i の識別子を持つ仮説がその例であり、二つの選択肢 c_f, c_g の双方に属している。

次に仮説管理部では、依存関係データベースにおいて、式 (2), (4) の形式、および式 (3) を以下のように変換した形式により、結果を記録する。

$$h_{ij}: \{\dots, c_i\} \quad (5)$$

ここで、 c_i は、 h_{ij} を要素として含む選択肢である。この形式は、仮説 h_{ij} に対して、 h_{ij} が属している選択肢の集合を表す。以後は、この集合を仮説 h_{ij} に対する支持集合と呼ぶ。仮説管理部では、式 (3) により仮説 h_{ij} を含む選択肢 c_i が通知されると、 h_{ij} の支持集合に c_i を追加する。例えば、図5の h_e と h_i, h_k の関係は、

$$\begin{aligned} h_e &\rightarrow \{c_g\} \\ h_i &: \{c_f, c_g\}, h_k: \{c_g\} \\ h_i &= \langle r_i, t10, \{住所, \dots\} \rangle, \\ h_k &= \langle r_k, t10, \{住所, \dots\} \rangle \end{aligned}$$

のように記述される。

以上の形式による記録が終了すると、処理部は、まだ排他集合が生成されていない仮説を仮説管理部に問い合わせる。その結果、例えば図5の状態に対しては、 $h_f \sim h_h, h_i \sim h_k$ が仮説管理部から通知され、これ

らを親仮説とした処理が行われる。

4.4 依存関係データベースの特性

仮説の検証と管理の詳細について述べる前に、依存関係データベースにより表現される依存関係、矛盾についてまとめておく。

まず、選択肢について考える。仮説は選択肢を単位として選択されるが、図5からもわかるように、仮説と同様、選択肢にも親子間の依存関係がある。さらに、同一の排他集合に属する選択肢は、互いに他と矛盾する。従って、選択肢を選択する際には、これらの依存関係、矛盾を考慮しなければならない。

さて、二つの選択肢 $c_1, c_2 (c_1 \neq c_2)$ について各々の先祖をたどる経路を考えると、ある世代から前はすべて共通の先祖を持つ。例えば、図5の c_f と c_g は、

$$c_f \leftarrow c_b \leftarrow c_a \leftarrow \dots$$

$$c_g \leftarrow c_c \leftarrow c_a \leftarrow \dots$$

のように、 c_a から前の世代がすべて共通の先祖となる。 c_1, c_2 に対して、以上の経路が唯一に定まる場合、選択肢 c_1, c_2 の矛盾は、次のように判定できる。

[選択肢の矛盾] (経路が唯一に定まる場合)

c_1, c_2 の経路において、最も若い共通の先祖を c とする。また、 c の子のうち、 c_1, c_2 の先祖となるものを、それぞれ c_1', c_2' とする。 c_1', c_2' が同一の排他集合に属するならば、 c_1, c_2 は矛盾する。□

前の例では、 $c_1 = c_f, c_2 = c_g$ としたとき、 $c_1' = c_b, c_2' = c_c$ であり、 c_b と c_c は同一の排他集合 S_b に属するため、互いに矛盾する。

一般には、一つの選択肢が複数の親を持ち得るため、複数の経路を考えなくてはならない。 c_1, c_2 が互いに矛盾するのは、 c_1 から得られる経路、 c_2 から得られる経路のすべての組合せについて、上記の条件を満たす場合である。逆に、上記の条件を満たさない経路の組合せが存在するならば、 c_1, c_2 は互いに矛盾しない。

図5の選択肢について、矛盾を考慮すると、無矛盾な組合せのうち、極大のもの(矛盾することなくさらに他の選択肢を追加できないもの)は、

$$[1] (c_a, c_b, c_f, c_d), \quad [2] (c_a, c_b, c_f, c_e)$$

$$[3] (c_a, c_c, c_g, c_h, c_d), \quad [4] (c_a, c_c, c_g, c_h, c_e)$$

の4組に限られる。文書全体に対する選択肢の極大な組合せは、文書画像理解の解候補を表す。仮説検証では、仮説の矛盾を発見することにより、これらの極大な組合せから、不適当なものを発見・除去することになる。

また、選択肢の矛盾に基づくと、仮説、排他集合の矛盾についても、次のように判定できる。

[仮説の矛盾]

仮説 h_1, h_2 の支持集合を、それぞれ $L_1 = \{c_{11}, \dots, c_{1m}\}, L_2 = \{c_{21}, \dots, c_{2n}\}$ とする。支持集合に属する選択肢の組合せ $(c_{1i}, c_{2j}) (1 \leq i \leq m, 1 \leq j \leq n)$ がすべて矛盾するならば、仮説 h_1, h_2 は互いに矛盾する。□

[排他集合の矛盾]

排他集合 $S_1, S_2 (S_1 \neq S_2)$ を生成するもとなった親仮説を、それぞれ h_1, h_2 とする。 h_1, h_2 が互いに矛盾するならば、 S_1, S_2 も互いに矛盾する。□
同様に、これらの条件を満たさない仮説、排他集合は互いに矛盾しない。

4.5 仮説の検証と管理

仮説検証とは、仮説の矛盾を発見することにより、不適当な選択肢の組合せを除去することである。仮説検証により発見された仮説の矛盾は、以下のいずれかの形式により仮説管理部に通知される。

$$h \rightarrow \perp \quad (6)$$

$$h_1, h_2 \rightarrow \perp \quad (7)$$

ここで、 h, h_1, h_2 は仮説、 \perp は矛盾を表す。式(6)は、構造解析における仮説を生成できないなど、仮説 h に対して処理が続行不能なことを表す。このように単独の仮説からなる矛盾は、仮説が誤りであることを意味する。また式(7)は、二つの仮説 h_1, h_2 が知識記述を満たさないことを表す。この形式の矛盾は、二つの仮説が同時に正解とはならないことを意味するが、どちらが誤りかは判断できない。

仮説管理部では、矛盾の通知を受けると、表1に示す管理規則を用い、可能な選択肢の組合せを制限する。表1の管理規則は、これまでに述べた仮説、選択肢、排他集合の性質に基づいて定められたものであり、概要は以下の通りである。規則1, 2は、通知された仮説の矛盾から、誤った選択肢を発見する方法を表す。規則3は、誤った選択肢を依存関係データベースから削除するために用いられる。規則4, 5は、規則3による選択肢の除去により、さらに誤った仮説を発見できることを示している。以上の規則により、誤った選択肢を発見・除去することは、選択肢の可能な組合せを大幅に削除することにつながる。

規則6は、二つの異なる選択肢が互いに矛盾すると判定された場合の管理方法を示している。選択肢 c_1, c_2 は互いに矛盾するため、同時に正解とはならないが、どちらが誤りかは判断できない。そこで本手法で

は, c_1, c_2 の矛盾を矛盾データベースに記録しておく。このような矛盾の記録は, 選択肢の可能な組合せを制限するという意味を持つ。例えば, 図5において, $nogood(c_b, c_d)$ が記録されているならば, 前節の [1]~[4] の組合せのうち [1] は矛盾を含むため, もはや適切な組合せとはなり得ない。

規則7は, 矛盾データベースの記録に基づいて, 誤った選択肢を発見できることを表している。例えば, 図5で, 前述の $nogood(c_b, c_d)$ に加え, $nogood(c_b, c_e)$ が記録されている場合, 前節の [1]~[4] の組合せのうち, 矛盾を含まないものは, [3], [4] の2組となる。これらの組を見ると, c_b は含まれていない。規則7は, このような選択肢を発見・除去するために用いられる。

以上の動作は, 仮説の矛盾が通知された後, 依存関係データベースが変化しなくなるまで適用される。

4.6 他の仮説管理法との比較

人工知能の分野では, 仮説管理を行う汎用システムとして, すでに, TMS (Truth Maintenance System)¹¹⁾, ATMS (Assumption based TMS)¹²⁾ が提案されている。ここでは, これらの2システムと比較することにより, 本手法の利点を明らかにする。

まず TMS との比較を試みる。TMS は汎用システムであるため, TMS を用いても, 機能的に本手法と同等のものが実現可能である。しかし, TMS では単一の無矛盾な状態しか扱うことができないため, 以下の

表1 仮説の管理規則
Table 1 Management rules for hypotheses.

1	仮説 h が誤りである場合, h を含む選択肢は誤りである。
2	仮説 h_1, h_2 が互いに矛盾する場合, h_1, h_2 を共に含む選択肢は誤りである。
3	誤りと判定された選択肢があるならば, すべての支持集合, 排他集合から除去する。
4	仮説 h の支持集合が空集合ならば, h は誤りである。
5	排他集合 S が空集合ならば, S を生成するもとになった親仮説は誤りである。
6	仮説 h_1, h_2 が互いに矛盾する場合, h_1 を含む選択肢 c_1 と, h_2 を含む選択肢 c_2 ($c_1 \neq c_2$) は互いに矛盾するため, $nogood(c_1, c_2)$ を矛盾データベースに記録する。
7	互いに矛盾しない排他集合を $S_1 = \{c_{11}, \dots, c_{1m}\}$, $S_2 = \{c_{21}, \dots, c_{2n}\}$ ($S_1 \neq S_2$) とする。矛盾データベースに, S_2 に属するすべての c_{2j} ($1 \leq j \leq n$) について, $nogood(c_{1i}, c_{2j})$ が記録されているならば, 選択肢 c_{1i} は誤りである。

ような問題が生じる。本論文で対象としているようなタスクでは, 知識として記述された種々の制約を用いても解が一意に定まる場合は少なく, 複数の解候補が得られるのが普通である。このような場合, 得られた解候補を比較して, 最良のものを選択したい。TMS を用いて可能な解をすべて求めるためには, 一つの解が得られるたびに, 強制的に後戻り (バックトラック) を起こす必要がある。一般に, このような強制的な後戻りは, 処理効率を低下させることが知られている¹²⁾。これに対して本手法では, 互いに矛盾する複数の状態を同時に扱うことができるため, このような問題は生じない。

次に, ATMS と本手法を比較する。ATMS を用いると, 複数の矛盾する状態を保持することができるため, 本手法と同様, 上記の問題点を解決することができる。ATMS を用いて, 本手法と同様の仮説管理を実現する場合, 具体的には, 文献13) に述べられている choose 形式の表現を用いることになる。しかしながら, このためには, 1) 計算コストの高価な hyper-resolution rule を適用しなければならない, 2) 選択肢の排他的な関係を表すために組合せ的に多くの $nogood$ を導入しなければならないなど, 効率上の代償を払う必要がある。

これに対して, 本手法では, ATMS を用いた場合に比べて計算コストのかかる管理動作を必要とせず, また, 排他的な選択肢の関係は依存関係データベースにより表されるため, 組合せ的な $nogood$ を必要としない。この理由は, 本論文で述べた仮説管理部が, 式(2)~(4), (6), (7)の形式を持つ処理結果のみを対象とした専用システムであるのに対して, ATMS は他のさまざまな形式による処理結果の通知* も受理可能な, 汎用システムであるためと考えられる。

次章以降では, 仮説を実際に生成・検証する, レイアウト処理, コンテンツ処理の詳細について述べる。

5. レイアウト処理

5.1 構造解析

構造解析では, まず, 親仮説の領域を部分全体関係に従って分割することにより, 子仮説を生成する。以下に, 図6に示す例に沿って処理手順を説明する。

最初に, レイアウト構造に関する知識記述を検索し, 親仮説 h_p の領域内に存在可能なすべての子

* ATMS では正当化 (justification) と呼ばれている。

める。子に個数変動がある場合には複数の組が得られる。図6では、親仮説が $b6$ に対応し、 $b8$ の個数変動が $[0, 1]$ であるため、 $(b8, b9)$, $(b9)$ の2組が得られる。子の組が得られると、各々を仮定して子仮説を生成する。

次に子の並ぶ方向とは逆の方向に投影処理を施し、領域内の画像を分割する。図6の場合は、子が縦方向に並ぶため、横方向に投影される。投影処理の結果において、黒画素が存在する部分に着目し、親領域を分割すると、いくつかの領域が得られる。さらにこの領域において、すべての黒画素を囲む矩形領域を求めることにより、図6の中段に示す領域を得ることができる。以後は、この領域を分割領域と呼ぶ。

画像に対して十分な傾き補正、ノイズ除去が成されていれば、分割領域は、子領域の一部あるいは全部であり、複数の子を含むことはないかと仮定できる。本手法では、以上の仮定のもとで、知識の記述を満たすように分割領域を統合し、可能な子領域をすべて生成す

る¹⁴⁾。ここで対象となる知識記述は、図6の $p1$, $p2$, $s1$ のように、対象とする親子関係に関連したものであり、 $s2$ のような記述は対象としない。

以上の処理により、一般には知識に記述された一つの子に対して複数の領域が生成される。なお、この段階で領域が生成できない場合には、 $h_p \rightarrow \perp$ を仮説管理部に通知する。

最後に、対応規則に基づいて論理オブジェクト名を付与し、子仮説として生成する。対応規則の右辺に複数の論理オブジェクト名が記述されている場合もそのまま付与されるが、これはレイアウト構造を考慮しただけでは曖昧性が残ることを意味する。生成された仮説は、式(2)~(4)の形式にまとめられ、仮説管理部に通知される。

仮説が生成されると、次に矛盾の発見を行う。対象は、新たに生成された仮説、およびその仮説と矛盾しない仮説である。矛盾の発見には、仮説の生成で対象外とした類似差異関係の記述(図6の $s2$ のように、複数の親にまたがる記述)を用いる。もし、仮説の組が知識の記述を満たさないならば、式(7)の形式により、矛盾を仮説管理部に通知する。矛盾の発見が終了すると、仮説管理部から呈示された仮説を親とし、さらに処理を進める。

5.2 文字切り出し・認識

構造解析が完了すると、文字列領域を対象として文字切り出し・認識を実行し、文字の仮説を求める。文字領域としては、基本的に全角、半角の2種類を考える。図7(a)の“号”のような場合は、双方の解釈が可能となるため、図7(b)の領域2, 3, 4のように、互いに矛盾する文字領域の仮説が生成される。ただし、これだけでは生成される仮説数が膨大になることから、漢数字の“二”, “三”などの典型的な分離文字や、“.”, “|”など文字を囲む矩形領域の形状が特殊なものに関しては、矩形領域の配置や形状に関する知識を用いて、不適当な仮説の生成を回避する。また、電話番号、郵便番号などの項目では、横書き文字列が混在するため、図7(b)の領域8のような領域を文字列領域と見なし、さらに文字切り出しを実行する。

文字領域の仮説は、ブロック領域、文字列領域の場合に比べて非常に多数となる。従って、式(3)のように、仮説の組合せを表す選択肢形式に基づいて、処理結果を記録することは実際的ではない。そこで本手法では、図7(c)に示す文字グラフを用いて、仮説を記録する。文字グラフにおいて、ノードは文字領域、

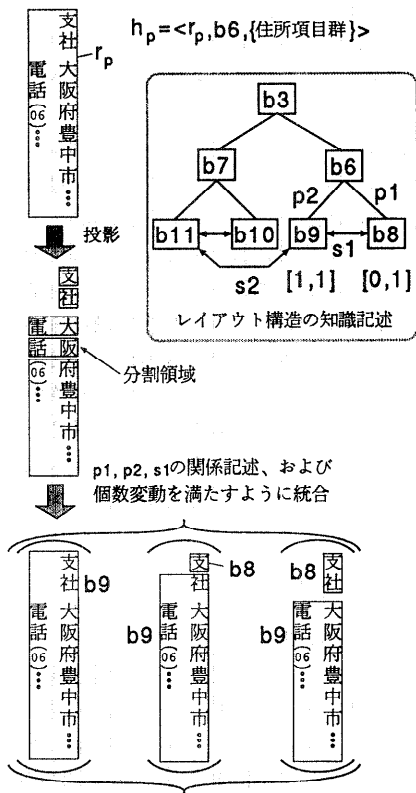


図6 構造解析 Fig. 6 Structure analysis.

アークは文字の読み順を表す。ここで読み順とは、縦書きの場合は文字列領域の上から下に、横書きの場合には左から右に文字領域をたどるものである。ノード1のように、複数のアークが出ることは、互いに矛盾する文字領域(2, 3と4)が存在することを意味する。

次に、各文字領域について文字認識を施す。認識手法としては、複合類似度法を用いる。認識辞書は、漢字、仮名、英数字、記号を合わせた3238カテゴリについて、教科書体、楷書体、ゴシック体、明朝体の4種類のフォントを考慮したものである。認識結果については、類似度をもとに候補文字数を決定する。得られた候補文字は、文字グラフのノードに記録される。

6. コンテンツ処理

6.1 単語列生成⁹⁾

構造解析により得られた項目の仮説を、 $h = \langle r, x, \{y_1, \dots, y_n\} \rangle$ とする。単語列生成では、項目の仮説 h を正しいと仮定して単語列を生成することにより、文字の仮説を検証する。本手法では、単語列生成に適用する知識(単語の接続性)を、仮説に付与された項目名(y_i)に基づいて選択するため、すべての知識について総当たりの処理を施す必要がなく、処理量が軽減される。

文字の仮説は項目の仮説ごとに文字グラフにより表現されるので、単語列生成とは、文字グラフのノードに記録された候補文字、およびアークの分岐から適切なものを選択することといえる。以下、処理手順について述べる。

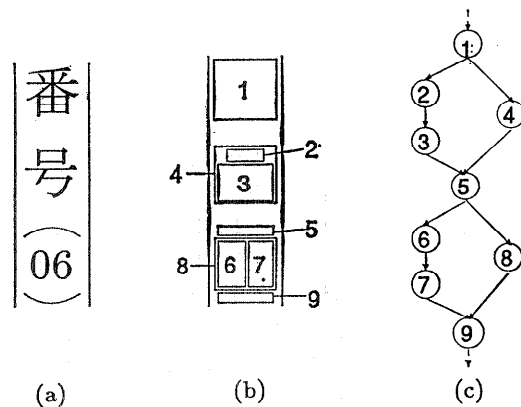


図7 文字グラフ
Fig. 7 Character graph.

a) 項目名が一つだけ付与されている場合

仮説が一つの項目に対応するため、知識に記述されている単語列の始端、終端が、それぞれ文字グラフの始端、終端と一致しなければならない。そこで、文字グラフの始端から順に、アークに沿ってノードを探索しながら単語照合を繰り返す。その結果、知識に記述された単語列の終端と文字グラフの終端が一致すれば単語列として生成する。なお、単語照合の際には、探索空間を削減するため、照合対象となる単語を直前に照合された単語と接続性を持つものに制限する。また、候補文字に正解が含まれない場合を考慮し、単語照合では、辞書と照合できない文字(未照合文字)を、上限を設けて許容する。

b) 複数の項目名が付与されている場合

各項目名 y_i に対して、以下の処理を施す。対象の仮説は複数の項目を含むことがあるため、知識に記述されている単語列の始端が文字グラフの始端と一致する必要はない。そこで、文字グラフの始端からアークに従って単語列の始端を順に仮定し、a)の場合と同様の単語照合を繰り返す。その結果、知識に記述された単語列の終端まで単語照合が成功すれば、単語列として生成する。

次に、仮説の修正を行う。5.1節にも述べたように、ここで対象とする仮説 h は、複数の項目名を持つ曖昧なものである。この曖昧性を除去するためには、仮説 h を、単一の論理オブジェクト名が付与された仮説に分解する必要がある。ここで、単語列を囲む領域を r' 、単語列を生成するときに仮定した論理オブジェクト名を y_i とすると、元の仮説を $\langle r', -, y_i \rangle$ なるいくつかの仮説に分解できる。次に、元の仮説 h を依存関係データベースから除去し、分解して得られた仮説について、領域の重複を考慮することにより、新たに式(2)~(4)の形式にまとめ、仮説管理部に通知する。

以上の処理により、項目の仮説 h に対して単語列が得られないならば、 $h \rightarrow \perp$ を仮説管理部に通知する。一方、単語列が得られた場合には、仮説は正しい可能性がある。ただし、複数の単語列 d_1, \dots, d_n が得られる場合には、どれが妥当であるかを判断できない。そこで本手法では、項目の仮説 h に対して得られた単語列を、

$$h: \{(d_1), \dots, (d_n)\}$$

の形式により記録し、次の単語列検証に選択を委ねる。

6.2 単語列検証

単語列検証では、上記の処理で得られた項目の単語列を対象に、単語列の矛盾に関する知識を用いて矛盾を検査する。各項目に対して複数の仮説が生成され、加えて各仮説に対して複数の単語列が生成されていることを考えると、総当たりの矛盾の検査は組合せの爆発を起こすため実際ではない。ところが本手法では、仮説を依存関係データベースにより管理しているため、すべての組合せを検査する必要はない。仮説が正解となる組合せは選択肢に限られることを考えると、依存関係データベースに記録された選択肢の矛盾を検査すればよい。

具体例を図8に示す。選択肢 c_3 では、仮説 h_{32} と h_{33} が矛盾するため、 $h_{32}, h_{33} \rightarrow \perp$ を仮説管理部に通知する。一方、選択肢 c_1, c_2 では、矛盾しない組合せが得られるため正しい可能性がある。しかしながら、どちらの選択肢が正しいのか、また c_1 のように複数のコンテンツの組合せが得られる場合にどれが正しいかを、この段階では判断できない。ただし、親仮説 h_{p1} のコンテンツは、

$$h_{p1}: \{(d_{11}^{11}, d_{11}^{12}), (d_{21}^{11}, d_{21}^{12}), (d_{11}^{21}, d_{11}^{22})\}$$

に制限される。そこで、親仮説 h_{p1} に対する上記のコンテンツを記録し、 h_{p1} の属する選択肢についてさらに矛盾を検査する。

以上の処理を繰り返すことにより、無矛盾な仮説は次第に減少し、同時に単語列も絞られてくる。その結果、最終的には名刺全体について無矛盾な単語列（名刺のコンテンツ）が得られる。なお、名刺のコンテンツとして複数の可能性が残る場合には、単語列中の文字に着目し、文字認識の際に得られた類似度の平均が最大のものを最終結果とする。

7. 実験結果と検討

7.1 実験概要

本手法の有効性を検討するため、電機関連企業の縦書き名刺100サンプルを対象として実験を行った。画像データは400 dpiのスキナーにより2値画像として得た。また、画像中のノイズは、位置と面積を手掛りに除去した。構造解析に用いる知識は、対象の半数に基づいて記述した。一方、単語の接続性に関する知識は、住所については対象に含まれる9都府県の全地

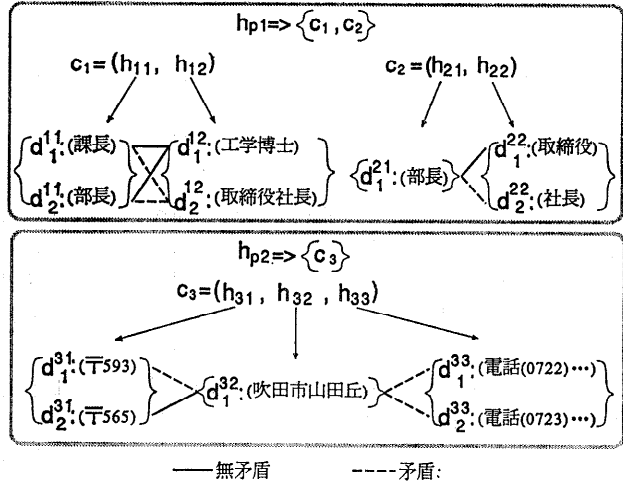


図8 単語列検証

Fig. 8 Verification of word sequences.

名を用いて、その他については対象から単語を抽出して作成した。総単語数は約3万語である。単語列生成処理における未照合文字の許容上限は、単語中の文字数の50%、単語列中の文字数の30%とした。ただし、氏名については、単語列生成処理を施さず、文字認識候補第1位をそのまま採用した。単語列検証処理では、約1500個の矛盾に関する記述を使用した。

7.2 評価基準

本手法では仮説駆動型の処理を行うため、手法の評価には、最終結果だけではなく、処理途中で生成された仮説も評価する必要がある。また、文書画像理解の目的はコンテンツの同定であることから、この点を評価する基準も必要となる。

そこで本論文では、以下の基準を用いて結果を評価する。レイアウト処理の評価基準としては、平均仮説数、信頼度を考える。平均仮説数とはレイアウトオブジェクト1個あたりに生成された仮説数の平均である。また、信頼度とは生成された仮説の中に正解が含まれている割合を表す。平均仮説数が1に近く、かつ信頼度が高いとき、レイアウト処理の有効性が高いといえる。一方、コンテンツ処理の評価基準としては、抽出率、同定率を考える。抽出率は仮説検証の役割を評価する基準であり、具体的には、コンテンツ処理後に選択された論理オブジェクトの仮説が正解である割合を表す。対象が文字の場合には、文字の領域、カテゴリがともに正解である割合を表す。同定率はコンテンツの同定という役割を評価する基準であり、項目の

コンテンツが正しく同定されている割合を表す。ここで、項目のコンテンツが正しく同定されるとは、項目が正しく抽出され、かつ項目内のすべての文字が正しく同定されることである。

7.3 レイアウト処理の結果と検討

レイアウト処理に要求される事項は、生成した仮説に正解が含まれること、および誤った仮説の数を少なく抑えることである。まず、構造解析の結果を表2に示す。全体としては、98.6%の信頼度をレイアウトオブジェクト数の約2倍の仮説により得ている。これは、名刺1枚に対して、抽出に失敗したレイアウトオブジェクトが平均0.23個であり、平均して二つの仮説から適切なほうを選べばよいことを表すため、ほぼ満足のいく結果であると考えられる。

ブロック領域について信頼度が100%に満たない原因は、画像の傾き、ノイズ除去の失敗のほか、知識に記述されていないレイアウト構造が未知サンプルに存在したことである。具体的には、図1のb8に横書き文字列が含まれていたことであり、知識の追加記述が必要となる。また、文字列領域に対する信頼度が100%に満たないのは、ブロック領域が正しく抽出されていないためである。

文字切り出し・認識の結果を表3に示す。切り出しに関しては少数の仮説により高い信頼度を得ているが、認識に関しては必ずしも十分とはいえない。このとき、文字認識率は78.7%であった。原因としては、画像の解像度が不十分なために住所等の極端に小さい文字に潰れ・かすれが頻出すること、社名ロゴなどの特殊文字に認識辞書が未対応なことが挙げられる。

表2 構造解析の結果
Table 2 Results of structure analysis.

レイアウト オブジェクト	個数	仮説あり		仮説なし
		平均 候補数	信頼度 (%)	信頼度 (%)
ブロック領域	780	1.90	98.7	78.7
文字列領域	845	2.07	98.5	80.8
合計	1625	1.99	98.6	79.8

表3 文字に対する結果
Table 3 Results for characters.

	個数	平均 仮説数	信頼度 (%)	抽出率 (%)
切り出し	8811	1.26	98.8	93.0
認識		8.29	94.2	

以上から、文字認識を除けば、ほぼ満足の結果が得られており、レイアウト処理の有効性が確認された。文字認識に対しては、解像度を上げる、認識辞書に特殊文字を登録するなどにより対処可能と考えられる。

7.4 比較実験

本手法の有効性を評価するために、以下の2種類の比較実験を行った。まず、仮説駆動型処理の有効性を示すため、仮説を用いない処理との比較実験を行った。仮説を用いない処理とは、5.1節に述べた構造解析において、親仮説から複数の選択肢が生成された場合、その段階で選択肢を一意に選択するものを指す。

具体的な選択基準を以下に示す。いま、式(2)~(4)の形式により選択肢が生成されたとする。ここで、排他集合*S*中の選択肢*c_i*に対して、距離*d_{ci}*を、

$$d_{ci} = \sum_j d_{ij}/n \quad (8)$$

$$d_{ij} = ||\mathbf{f}_x - \mathbf{f}_{x'}|| \quad (9)$$

$$\mathbf{f}_x = (f_1/\sigma_1, \dots, f_k/\sigma_k, \dots, f_m/\sigma_m) \quad (10)$$

$$\mathbf{f}_{x'} = (\bar{f}_1/\sigma_1, \dots, \bar{f}_k/\sigma_k, \dots, \bar{f}_m/\sigma_m) \quad (11)$$

とする。仮説を用いない処理では、親から複数の選択肢が生成された場合、距離*d_{ci}*の値が最小となる選択肢*c_i*を選択する。ここで、*n*, *d_{ij}*は、それぞれ選択肢に含まれる仮説の個数、式(4)の仮説*h_{ij}*に対する距離(特徴ベクトル*f_x*, *f_{x'}*間のユークリッド距離)である。また、*f_x*, *f_{x'}*は、それぞれ仮説*h_{ij}*の矩形領域*r*から求めた特徴ベクトル、既知サンプルに含まれる同種のレイアウトオブジェクト(レイアウトオブジェクト名が同じもの)から求めた特徴ベクトルである。特徴ベクトルを構成する特徴量としては、2.1の(2)に述べた部分全体関係を規定する特徴量と同じものを用いた。ここで、*f_k*, *σ_k*は、それぞれ、既知サンプルから得た特徴量の平均、標準偏差である。

結果を表2に示す。この結果から、仮説を用いると、信頼度が大幅に改善されることが分かる。この理由は、対象とした名刺が、多様なレイアウト構造を持っていたためと考えられる。

次に、仮説管理部の有効性を示すため、横型探索を用いるものと比較した。ここで横型探索とは、4.2に述べた仮説の管理を行わず、生成された仮説をすべて異なるものとして横型に処理を実行する場合を指す。横型探索を用いると、例えば、図5の選択肢*c_f*について処理を行った後、選択肢*c_g*について処理を行う場合、仮説*h_i*に対して再度、同一の処理を施してしまう。この場合は、仮説*h_i*が文字列なので、文字切り出し・認識、単語列生成、単語列検証が重複して適

用されることとなり、処理効率の低下を招く。

横型探索を用いた実験の結果、ブロック領域に対する平均仮説数は本手法と同じであったが、文字列領域に対する平均仮説数は 2.54 に増加した。増加分を文字列数に換算すると約 400 となる。本手法では、この増加分だけ、後続の処理を重複して行わなかったことになる。このことから、仮説管理部が処理効率の向上に大きく寄与していることがわかる。

7.5 コンテント処理の結果と検討

文字の抽出率を表 3 に示す。ここで、文字の抽出率とは、文字の領域、カテゴリの両方が同定された割合である。文字認識率が 78.7% であったことを考えると、大幅な改善効果が認められる。コンテント処理の結果、文字が正しく同定された例を図 9 (a) に示す。本手法では、この例のように、複雑なレイアウトを持ち、潰れ等の画像劣化が著しい文字列に対しても、文字を正しく読み取ることができる。

約 7% の抽出誤りの主な原因は、数字や特殊文字の誤認識を単語列生成により救済できなかったことである。単語列生成では、一般の単語に対しては単語照合時に未照合文字を許しているが、数字部分に対しては意味を持たないため許していない。したがって、電話

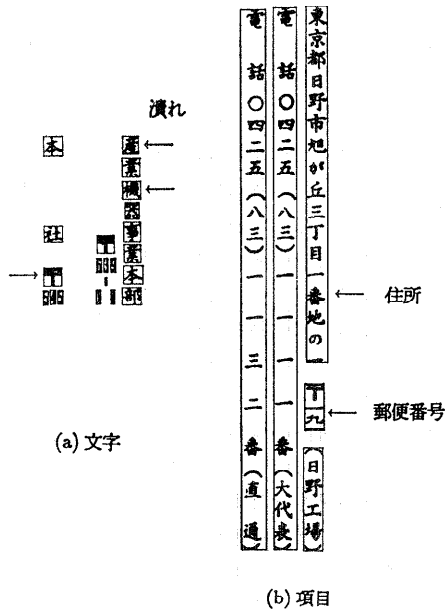


Fig. 9 Examples of processing results.
(a) Characters.
(b) Items.

表 4 コンテント処理の結果
Table 4 Results of contents processing.

種類	論理オブジェクト	個数	抽出率 (%)	同定率 (%)
項目	社名および数字が主体の項目*	401	80.5	76.3
	その他の項目**	571	96.7	93.9
	小計	972	90.0	86.6
その他	住所項目群	119	95.8	—
	住所群, 氏名群, 社名群	300	99.7	
	小計	419	98.6	
合計		1391	92.6	

*) 社名, 電話番号, FAX, TELEX, 郵便番号, 内線

**) 部署名, 肩書, 氏名, 見出し, 住所, ビル名

番号などの数字部分の候補文字中に正解が存在しない場合には、単語列を生成できない。また、社名については、特殊文字が連続する場合、未照合文字の許容数を越えてしまうため、救済できなかった。

論理オブジェクトの抽出率を表 4 に示す。抽出に失敗した項目数は 97 個であった。このうち約 8 割は社名および数字が主体の項目である。これらの仮説は、6.1 の b) に述べたように、単語列生成により得られるものである。失敗の原因については前述と同様である。その他の論理オブジェクトについては、表 2 の信頼度に比べて抽出率があまり低下していないことから、正しい仮説の選択にほぼ成功しているといえる。項目の抽出成功例を図 9 (b) に示す。本手法では、図 9 (b) の住所、郵便番号などのように、レイアウト構造だけでは曖昧性が残る項目に対しても抽出に成功している。

同定率についても結果を表 4 に示す。社名および数字が主体の項目に対しては、前述の原因から同定率が低下しているが、その他の項目に対しては、高い同定率を得ている。項目全体で見れば、抽出率 90.0% に対して同定率が 86.6% となり、正しく抽出された項目の大半に対して、コンテントが正しく同定されている。

8. おわりに

文書画像理解の能力向上には、文書のレイアウトに基づく処理とコンテントに基づく処理の両者を融合する必要があるとの観点から、本論文では仮説駆動型の文書画像理解法を提案した。本手法の特徴は、仮説の

管理において汎用性を考慮している点にある。具体的には、仮説の依存関係と矛盾を、依存関係データベースと矛盾データベースを用いて管理し、仮説駆動型処理を制御することである。

縦書き名刺を対象とした実験の結果、文字の抽出率93.0%、論理オブジェクトの抽出率92.6%、項目のコンテンツの同定率86.6%を得、本手法の有効性を確認した。今後の課題には、文字認識の精度を向上させることなどがある。

謝辞 筆者の一人黄瀬が日頃御激励いただき、大阪府立大学の日下浩次教授、福永邦雄教授、高松忍助教授に感謝する。また、本研究を遂行する上で種々御協力いただいたシャープ株式会社の関係各位に感謝する。本研究の一部は文部省科研費の補助による。

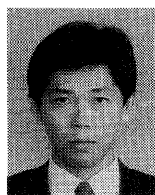
参 考 文 献

- 1) 野口健一郎, 大谷 真: OSIの実現とその課題, 情報処理, Vol. 31, No. 9, pp. 1235-1244 (1990).
- 2) Niyogi, D. and Srihari, S. N.: A Rule-based System for Document Understanding, *Proc. AAAI-86*, pp. 789-793 (1986).
- 3) Dengel, A. and Barth, G.: ANASTASIL: A Hybrid Knowledge-based System for Document Layout Analysis, *Proc. IJCAI-89*, pp. 1249-1254 (1989).
- 4) 中野康明, 藤沢浩道: 自動ファイリングのための文書画像理解の方式, 信学論(D), Vol. J71-D, No. 10, pp. 2050-2058 (1988).
- 5) 黄瀬浩一, 杉山淳一, 馬場口登, 手塚慶一: レイアウトモデルに基づく文書構造解析, 信学論(D-II), Vol. J72-D-II, No. 7, pp. 1029-1039 (1989).
- 6) 黄瀬浩一, 馬場口登, 手塚慶一: 文書画像理解における推論と制御の一提案, 信学技報, PRU 89-76 (1989).
- 7) 駱 琴, 渡邊豊英, 吉田雄二, 稲垣康善, 齊藤隆夫: 知識ベースに基づいた図書目録カードの理解, 情報処理学会論文誌, Vol. 31, No. 12, pp. 1755-1767 (1990).
- 8) 黄瀬浩一, 山岡正輝, 馬場口登, 手塚慶一: 文書画像構造解析のための知識ベースの一構成法, 情報処理学会論文誌, Vol. 34, No. 1, pp. 75-87 (1993).
- 9) 百田賢一, 黄瀬浩一, 馬場口登, 手塚慶一: 単語間の接続関係に着目した文字認識後処理, 1990年信学秋季全大, D-363 (1990).
- 10) 杉山淳一, 百田賢一, 黄瀬浩一, 田中直樹, 馬場口登, 手塚慶一: 文書画像理解における単語情報と論理構造の援用法, 信学技報, PRU 89-90 (1990).

- 11) Doyle, J.: A Truth Maintenance System, *Artif. Intell.*, Vol. 12, pp. 231-272 (1979).
- 12) de Kleer, J.: An Assumption-based TMS, *Artif. Intell.*, Vol. 28, pp. 127-162 (1986).
- 13) de Kleer, J.: Extending the ATMS, *Artif. Intell.*, Vol. 28, pp. 163-196 (1986).
- 14) 山岡正輝, 黄瀬浩一, 馬場口登, 手塚慶一: 知識ベース型文書構造解析システムの汎用性に関する一考察, 信学技報, PRU 90-119 (1991).

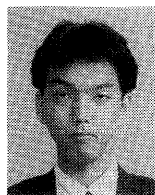
(平成4年8月31日受付)

(平成5年6月17日採録)



黄瀬 浩一 (正会員)

昭和38年生。昭和61年大阪大学工学部通信工学科卒業。昭和63年同大学院前期課程修了。平成2年より大阪府立大学工学部助手。パターン認識、画像理解の研究に従事。工学博士。電子情報通信学会、人工知能学会各会員。



百田 賢一

昭和41年生。平成2年大阪大学工学部通信工学科卒業。平成4年同大学院修士課程修了。現在、日本電信電話(株)勤務。在学中、パターン認識に関する研究に従事。



杉山 淳一

昭和39年生。昭和63年大阪大学工学部通信工学科卒業。平成2年同大学院修士課程修了。現在、富士通(株)勤務。在学中、文書画像理解に関する研究に従事。電子情報通信学会会員。



馬場口 登 (正会員)

昭和32年生。昭和54年大阪大学工学部通信工学科卒業。昭和56年同大学院前期課程修了。昭和57年愛媛大学助手、大阪大学工学部助手、講師を経て、現在、大阪大学産業科学研究所助教授。工学博士。人工知能、パターン認識、画像処理の研究に従事。IEEE、電子情報通信学会、人工知能学会各会員。

**手塚 慶一 (正会員)**

昭和3年生。昭和26年大阪大学工学部通信工学科卒業。同大学院特別研究生、愛媛大学、山口大学、大阪大学助教授を経て、大阪大学工学部教授。平成4年大阪大学名誉教授。情報通信工学に関する広汎な分野の研究に従事。工学博士。著者「電子計算機基礎論」、「電子計算機システム論」、「デジタル画像処理工学」など多数。平成5年1月22日逝去。
