需要家地図システムのための大規模 データベース間推論連係方式

正員 近田 伸行 (東京電力)
非会員 山口 剛 (東京電力)
非会員 嶋田 茂 (日立)
非会員 山本 直樹 (日立)

An Inferential Matching Method for Customer Service Mapping System

Nobuyuki Chikada, Member, Tuiyoshi Yamaguchi, Non-member (Tokyo Electric Power Co.),
Shigeru Shimada, Non-member, Naoki Yamamoto, Non-member (Hitachi, Ltd.)

A mapping system using a commercialized residential map is getting attention as an effective management tool in public service sectors. Utilization of a residential map by linking with an existing customer database allows efficient round inspection and quick response for customer's claims.

A conventional matching method for linking between an customer database and a residential map is simple comparison of customer's names and resident's names. This simple method is applicable only for a limited number of data, but it is not feasible in linkage of a large number of data because of extremely long computational time and low matching ratio.

In this paper, we propose an inferential matching method which enables automatic matching for a large volume database. This method rationalizes matching process by adaptive focusing of candidate-data. These data are selected in corporate with calculated matching certitude. We applied the method for 200,000 customer data and the result shows the significant improvement of matching ratio from 30% to 60% in comparison with the conventional method.

キーワード：推論連係，地図システム，データベース連係

1. はじめに

市販の地図データの上に，各企業固有の属性情報をマッピングする汎用地図システムが注目を集めている。従来より，各企業が独自の地図をデジタル化し属性情報とともに計算機に入力する専用地図システムは多数発表されてきた。これらのシステムでは，地図や図面の計算機への入力とデータベース化に膨大な費用が必要となるので，なかなか本格的な運用にまで至らなかった⁽¹⁾⁽²⁾。

しかし近年，日本の主要都市の住宅地図や道路経路図といった汎用性の高い地図データがデータベース化

され，CD-ROMなどの媒体を介して販売されるようになった。そこで，これらの市販の地図データベースを有効に活用し，低コストで地図システムを構築する汎用地図システムが注目を集めている。

これらの市販の地図データベースを各種の業務において活用しようとした場合，それぞれの業務で使用していた既存のデータベースと融合しなければならない。すなわち，既存データベースとこの地図データベース間の自動連係技術が必要となる。この自動連係技術に関しては，既に電話帳データベースと住宅地図データベース間の連係を図った報告例がある。そこでは自動連係率を向上させる技術として，名義を単語分割

しそれらの属性別に部分的な比較を行う方式が提案され、その有効性が報告されている⁽³⁾。しかし方式確認の地域が特定の1地域に限られ、確認件数が約2万程度と少ないこと、および実際の地図システムとしての運用にまで至っていないなどの問題があった。

そこで本論文の目的は、実用レベルの地図システムで使用可能なDB構築のための連係方式を提案し、その評価を述べることにある。

2. データベース構築

〈2・1〉 システムのねらい 電力会社においては、需要家からの多様な問合せに対して、迅速かつ的確に対応する必要がある。そのためには、地図上で需要家データが視覚的に把握でき、需要家の周辺地図と訪問作業に必要な関連情報が即座に検索できるような「需要家地図システム」が有効である。このシステムでは、1,000万件に近い大規模な需要家データベースと東京都区内で3 Gbyteを越す大容量の住宅地図データベースとを高精度に連係させることにより、次のような基本的な機能が容易に実現できる。

(1) 関連検索機能 需要家番号や住所・氏名などから関連する家屋周辺の地図を表示したり、逆に地図上の家屋を指示することにより、関連する需要家データを検索する機能。

(2) 統計分布表示機能 設備機器の保有家屋の分布など、各需要家の属性情報を統計的に処理した結果を、色別の分布状況として住宅地図上へ表示する機能。

(3) 情報の一時的付与機能 需要家へのサービス結果を、手書きのメモなどのアノテーション（一時的付与）情報にして需要家の住居に関係づけ、時系列順や項目別などの各種の条件による検索を可能にする機能。

これらの基本的な機能を実現するための情報源として、マルチメディアデータベース（以降、MMDBと略記）が位置づけられる。このMMDBには、需要家の名義をなすテキストデータや、住所・家屋番号などのコードデータのほか、住宅地図を構成する座標データ、建物の外観を示す画像データなどのマルチメディア情報が記憶される。そして、これらのマルチメディア情報は互いに関連づけられ、高速に検索できるように管理される。以下、このMMDBを自動的に構築する処理方式を述べる。

〈2・2〉 データベース構築の流れ 本地図システムにおけるMMDBは、住宅地図データのように長期間にわたり変化の少ないスタティクな部分と、手書

図1 データベース構築の流れ
Fig.1. Processing flow of database construction.

きメモのような一時的に付与されるダイナミックな部分とで構成される。このうちデータベースが自動的に構築されるのはスタティクな部分であり、ダイナミックな部分はマッピングシステムとして動作させる過程で、ユーザが必要に応じて付与することにより構築される。以下、このMMDBのスタティク部の構築方法について述べる。

このMMDBのスタティク部は、それぞれ独立に存在する需要家属性DBと住宅地図DBを、住所と需要家名称をキーに自動的に連係させて構築する。需要家属性DBの需要家名称はカナであるので、カナ漢字変換を行った後、住宅地図DBの居住者名と漢字での推論連係処理を行う。この推論連係とは、需要家属性DBの需要家名称と、住宅地図DBの居住者名との間の表記上のずれを、エキスパートシステムの前向き推論を用いて補完し、両者を連係することである。なお、この推論連係処理においては、単純に両DBのデータを比較して連係するのではなく、まず確実性の高い情報を用いて連係処理を行い処理範囲を限定した後、より不確実な情報を用いて推論連係を行う階層型の連係処理を行っている。この階層型の推論連係処理の詳細は後述する。この推論連係処理を経ても連係が一意に決まらなかったデータに関しては、対話的なメンテナンス処理により連係させ、DB構築を完了する。図1に、このMMDB構築の処理の流れを示す。

〈2・3〉 各ファイルとデータベースの構造 まず連係処理の対象となる需要家属性マスタファイルと住宅地図要素ファイルの基本的な構造、および連係処理後に作成されるMMDBの基本的な構造について述べる。

(1) 需要家属性DBのマスタファイル このファイルは、マッピングシステムとは別の、公共企業体における需要家の契約状況や住所など大規模な需要家

アドレスキー部					名称キー部	
住所コード	街区番号	地番	住居ID	棟名称	階	部屋番号
					需要家名称	

図 2 需要家属性 DB のレコード形式
Fig. 2. Record format of customer DB.

インデックスキー部			レコード本体制御部		
レコードID	住所コード	地番	住居ID	属性	ガイナター

(a) 制御部の構造

インデックスキー部			レコード本体図形部		
レコードID	地番	...	構成点表示属性	第1座標(X,Y)	第2座標(X,Y) ...

(b) 図形部の構造

インデックスキー部			レコード本体文字部		
レコードID	地番	...	文字数	表示属性	第1文字コード 第2文字コード ...

(c) 文字部の構造

図 3 住宅地図 DB のレコード形式
Fig. 3. Record format of residential map DB.

属性 DB の一部をなすもので、図 2 に示すような構成をなす固定長のキャラクタベースのファイルである。

(2) 住宅地図 DB の要素ファイル 市販の地図データベースの一つに、住宅家枠の一軒一軒に世帯主名称を明示可能な住宅地図データベースがある。このデータベースは、住宅家枠や道路経路などの形状要素ファイルと、地名・建物名など名称要素ファイルとで構成される。それらの構造は、図 3 に示すような図形構成点数やキャラクタ数などの指定による可変長マルチメディアのファイルである。

(3) MMDB の構造 MMDB のスタティック部の構造は、上記住宅地図要素ファイルの構造を基本にして、これを住宅地図要素ファイルの居住者名称と連係した需要家属性マスタファイルの属性レコードの部分追加された形式となっている。すなわち、上記(1)、(2)のファイル構造と同様に、キー部と本体部とで構成される。

一方 MMDB のダイナミック部は、マルチメディアの管理特性とインタラクション特性に優れたオブジェクト指向による管理方式を導入したが、その詳細は、他の文献にゆずる⁽⁴⁾。

3. データベース間の推論係方式

(3・1) データベース間の係処理上の問題点

本 MMDB 構築における係処理は、係元となる需要家属性ファイルの各レコードと、係先となる住宅地図要素ファイルの各レコードの一致性を判定し、係することである。

この MMDB を自動的に構築する係処理で問題となるのは、次のような項目があげられる。

(1) カナ漢字変換による不確定性の増加 係元の需要家属性マスタファイルの需要家名称やビル名称などがカナになっているのに対して、係先の住宅地図要素ファイルにおける居住者名やビル名などは漢字になっている。これより本研究の係処理は、さきに文献で示した電話帳 DB と地図 DB との係に比べ、カナ漢字変換係処理が余分にかかることになり、係係処理上の不確定要因が増加している。

(2) DB 作成時期のずれによる不確定性 需要家属性 DB と住宅地図 DB とでは作成時期が異なるため、本来同一の対象であっても需要家名称と居住者名や住所などにずれが存在する場合がある。すなわち需要家属性 DB は、東京電力の契約者を対象に長期間蓄積し日々更新した結果であるのに対して、住宅地図は 1 年単位で調査更新されるものとなっている。従って、例えば比較的新しい時期に住所などの表示変更が起きた地域では、係係処理上の不確定要因が増加することになる。

(3) 集合住宅居住者の係係上の問題 都市の地価高騰のため、都市中心部での集合住宅の数が増加している。電力会社の契約者もこのような集合住宅に住む居住者の割合が増加し、同一の住所をもった契約者が多量に需要家属性 DB に登録される件数が多くなっている。ところが集合住宅の場合、棟名称や部屋番号の表記方法に関する規則は標準化されていない。従って、住所に関しては一意的に決まらないような不確定要因が増える。

(3・2) 階層型の推論係方式

(1) 基本方針 以上説明したように、需要家属性 DB と住宅地図 DB 間の係係を行うには、不確定な要因が多いことがわかった。従って、単純な需要家名称と居住者名称間での直接係係係係では、係係が一意的に収束しない懸念がある。また本研究では、係係係係の前に需要家属性 DB の各レコードに漢字変換を施す必要があるため、名義レベルの係係係係が組合せ的に増大する恐れがある。

そこで、著者は、まずマクロに住所レベルでの限定を行ったのち、名義レベルの係係係係対象の数を減らし、更に必要に応じて同一の集合住宅のレベルにまで限定し、再び係係係係数を減らすといった階層型の係係係係方式を用いた。この階層型の係係係係方式で用いたキーは次のようになっている。

(i) 需要家住所と居住者住所との一致性 (住所キー)

- (ii) 需要家および居住者の住んでいる集合住宅名称またはビル名（名義キー）
- (iii) 需要家および居住者の住んでいる集合住宅の部屋番号（住所キー）
- (iv) 需要家名義と属性名称との一致性（名義キー）

また、これらのキーは名義キー部と住所キー部とに分け、住所キーはコード情報になっており、名義キーはカナと漢字テキストになっている。

（2）処理の流れ 次に、この階層型の連係方式による処理の流れを図4に示す。連係処理の基本方針は既に述べたように、よりマクロな観点から連係候補の数を限定し、処理負荷の重い名義キー間の連係処理は候補が少なくなった状態で、処理量を減らすことにある。以下、具体的な処理内容を段階別に説明する。

〔第1段階〕 住所コードを用いた連係候補の限定
連係対象を同一の住所コードの範囲に限定し、連係候補の数を減少させる。この限定処理により一意的に対応対象が決まった場合にも、第4段階の需要家名義により一致性の判定は行う。一方、全く候補がない場合には、需要家属性DBの住所が更新されている可能

性があるので対話的なメンテナンス処理に移る。

〔第2段階〕 住居名による連係候補の限定

上記限定された連係候補のなかで集合住宅の場合には、名義テキストを用いた住居名称（マンション名など）の連係チェックを行い、更に連係候補を限定する。この場合の連係チェックは推論連係を行い、カナ漢字変換を含む高い精度の連係性を判定する。

〔第3段階〕 部屋番号による連係候補の限定

前ステップで限定されたように、同じ名称を有する集合住宅に住んでいる需要家を更に限定するため、部屋番号による連係候補の限定を行う。

〔第4段階〕 需要家名義による連係の一致性判定

最終的に需要家名称と居住者名との一致性を判定して、限定された候補が最終的に一意に決定できるかどうかのチェックを行う。前ステップまでに連係対象が一意的に限定されて、この段階で名義が異なる場合には、契約者の名義変更を忘れているか、または住宅地図の名義が変更されていない可能性があるため、対話的なメンテナンス処理に移り、その是非を判定する。

以上のステップにおいて、第1・第3段階では単純なコード間の完全一致をとるのに対して、第2・第4段階ではカナ漢字変換を含む省略名称や類語名称に対する推論連係を採用し、連係精度を向上させている。

（3）連係レベル 階層型の推論連係処理では、処理途中で処理候補が一意に選択される場合もある。従って、最終的に連係された結果の信頼性を把握できるような「連係レベル」を導入して、連係結果の信頼性の高いものから順に区分けできるようにする。

その連係レベルの具体的な内容を表1に示す。この表より明かなように、連係が有効とされるためには、住所コードと街区番号とが最低限度一意に合致する必要があり（レベル4以上）、それ以外の場合には自動的な連係は不可能で、人間による判断が必要と考

図4 階層型連係処理の概要

Fig. 4. Flow of the adaptive focusing in the inferential matching method.

表1 連係レベルの内容

Table 1. Classes of matching certitude.

レベル	ケース	一致項目	信頼性
1	一戸建 集合住宅	住所コード, 街区番号, 地番, 姓名 住所コード, 街区番号, 地番, 集合住宅名, 室番号, 姓名	高
2	一戸建 集合住宅	住所コード, 街区番号, 地番, 姓名 住所コード, 街区番号, 集合住宅名, 室番号, 姓名	
3	一戸建	住所コード, 街区番号, 姓名	
4	集合住宅	住所コード, 街区番号, 集合住宅名, 室番号, 姓名	
5	—	一意に決定できない場合	低

える（レベル5以下）。

〈3・3〉 推論連係の内容

（1） 推論の実現方針 推論連係処理の流れは、連係元の需要家属性ファイルの内容から連係先の住宅地図ファイルへ対応をつけるような方向で考える。この場合、既に述べたように需要家属性ファイルの名義部分はカナコードになっているので、カナ漢字変換処理を加えながら連係をとる必要がある。従ってカナ漢字変換用の辞書を用意する必要がある。

一方、この推論連係処理の対象となる名義部分には、省略名称や類語名称が含まれる。例えば、まず省略名称には「東京電力」を「東電」、「日本航空」を「日航」といった省略形をなすものがあり、類語名称には「本社」を「本店」や「本部」「センター」などといった類似した呼称で言い換えるものがある。このうち省略名称の連係の場合には、従来から確立されたテキスト部分列の照合アルゴリズムを用いることにより解決できる見通しがある。しかし、類語名称の場合には、類語を正当な呼び方に変換するような何らかの辞書を備える必要がある。

ところで、本研究で推論連係の対象とする名義部分は、住宅地図 DB における居住者名称や住宅名称などである。この名義部分は、居住者名称・住宅名称いずれの場合も、固有名詞部と普通名詞部とに分かれることが多い。例えば、居住者名称が「東電本店」、住宅名称が「日本ビル」であったとすると、普通名詞部は「本店」、「ビル」で、固有名詞部は「東電」、「日本」ということになる。しかも上記類語は、これらの普通名詞部と固有名詞部それぞれ独立に考えられ、複合した類語を形成する場合は極めて少ないと予想される。例えば、普通名詞「本店」の類語として「本社」、「本部」などが考えられ、一方、固有名詞「日本航空」の類語として「JAL」、「日航」などが考えられる。

以上の理由から、既に必要性を示した「カナ漢字変換辞書」や「何らかの類語変換用辞書」は、この普通名詞部と固有名詞部とに分けて考えることにした。しかも後者の類語変換のための辞書には、単なる辞書の形式をとるのではなく、連鎖的に展開可能なプロダクションシステムのルールベースを採用した。また、この場合の推論方式としては、ある名詞をキーとしてその類語を展開するような推論方式で十分なので、エキスパートシステムの前向き推論部⁹⁾を活用した。

これによって、類語変換のための記憶容量増加を抑える効果があるほかに、新たに発見された変換規則を追加する場合のメンテナンス特性を向上させるなどの効果が期待できる。

図 5 推論連係の流れ
Fig. 5. Flow of the inferential matching method.

（2） 推論連係の具体的な内容 ここで、推論連係処理の具体的な処理の内容を図5に従い説明する。説明をより具体化するため、連係処理の連係元となる需要家属性マスタファイルの名義部が「トウデンホンテン」で、これを連係先となる要素ファイルの名義部が「東京電力本社」であるような場合を想定し、各処理ステップ別に説明する。

〔ステップ1-1〕 固有名詞部と普通名詞部との分離 まず、需要家属性マスタファイルのカナ部分を普通名詞辞書を使って固有名詞部と普通名詞部とに分ける。普通名詞辞書を用いて分離する理由は次のとおりである。すなわち、普通名詞の数は固有名詞の数に比べ圧倒的に少ないため、文字列の部分参照回数が少なくすむ。また、普通名詞のカナ漢字変換は固有名詞の場合より一意に決まりやすいことから、普通名詞の分離能力は固有名詞より高い。この分離処理に使用する普通名詞辞書の構造は、「カイシャ」、「カブシキ」などのカナをキーとして、それに対応した漢字を配置したカナ漢字変換辞書の形式である。

〔ステップ1-2〕 普通名詞部の類語展開 「普通名詞類語ルール」を用いて、前ステップで分離した普通名詞を出発とした類語の展開を行う。この場合、

表 2 普通名詞類語ルールの例
Table 2. Rules of synonyms for common nouns.

カテゴリ	IF 部 入力 キーワード	THEN 部 連想 キーワード	確信度	連想 キーワード	確信度
1	会社	株式会社	0.8	(株)	0.5
		K. K.	0.6	合資会社	0.3
	本社	本店	0.5	本部	0.3
		センター	0.5		
2	屋	店	0.6	商店	0.4
	商店	商事	0.5	商会	0.4
3	建設	工務	0.5	工務店	0.4
		建築	0.6	組	0.2
		土木	0.3	土建	0.2
		工事	0.2	興業	0.1

総計 450 ルール

使用する普通名詞類語ルールの構造は、表 2 に示すように、IF 部に展開対象とするキーを置き、THEN 部には展開可能な類語を確信度付きで配置している。更にこのルールは、会社・建設などのカテゴリ別にフレーム構造化して、不必要に推論が進みすぎないように工夫した。

〔ステップ 2-1〕 固有名詞部のカナ漢字変換

ステップ 1 で普通名詞部と分離された需要家属性の名義部を対象にして、「熟語辞書」を参照した固有名詞部のカナ漢字変換を行う。この場合、使用する熟語辞書は、約 11 万語の姓名辞書部と約 9 万語の一般名称部とで構成される。従って、固有名詞として漢字変換される候補は多数考えられるが、優先度順に抽出可能である。

〔ステップ 2-2〕 固有名詞部の類語展開 この処理は、既に説明した普通名詞の部の類語展開処理とはほぼ同一の内容をもつ。この場合使用する「固有名詞類語ルール」の構造は、さきに示した普通名詞類語ルールと同一であるが、ルールグループの単位はカテゴリ別ではなく単純な企業別の小さな単位となっている。例えば表 3 に示すように、企業名「東電グループ」に対するルールとしては、IF 部に類語展開のキーとなる「東電」を配置し、THEN 部には、「東京電力」、「東京電機」などの候補が確信度付きで展開されるように配置される。

〔ステップ 3〕 普通名詞部類語と固有名詞部類語との組合せ推論 上記〔ステップ 1, 2〕でそれぞれ独立に推論された普通名詞部と固有名詞部の類語を、組み合わせることにより連係候補を作成する。この場合使用する「組合せルール」には、表 4 に示すよ

表 3 固有名詞類語ルールの例
Table 3. Rules of synonyms for proper nouns.

カテゴリ	IF 部 入力キーワード	THEN 部 連想キーワード	確信度
1	東電	東京電力	0.8
		東電設計	0.4
		東電設備	0.4
		東電テレコム	0.3
		東京電気	0.2
		東京電機	0.2
2	日立	日立製作所	0.8
		日立電線	0.5
		日立マクセル	0.4
		日立家電	0.4
		日立造船	0.2
		日立建機	0.2

総計約 1,000 ルール

表 4 組合せルールの例

Table 4. Rules of linkage.

No.	IF 部	THEN 部
1	固有名詞部(PN)の確信度が THU 以上 AND 普通名詞部(CN)の確信度が THU 以上	連係処理用リストに 組合せキー PNCN と CNPN を追加する
2	固有名詞部(PN)の確信度が THU 以上 AND 普通名詞部(CN)の確信度が THU 以下	連係処理用リストに 組合せキー PNCN を追加する

総計 9 ルール

うな形式で、普通名詞と固有名詞の並びの順序を規定するルールを記述する。しかも連係の可能性が高い順に候補が出現するように、普通名詞の類語と固有名詞の類語の推論時に発生したそれぞれ個別の確信度を用いて連係候補の組合せを推論する。

〔ステップ 4〕 多項目並列連係 前ステップで推論された連係候補の数は、組合せ的に多数存在する。これの一つずつ順に連係先の住宅地図要素ファイルの居住者名称と、テキスト列レベルで単純に比較する方式では処理速度が上がらない。そこで本研究では、テキストの部分列単位に並列的な比較が可能な Aho & Corasick によるテキストサーチ方式⁽⁶⁾を変形して用いている。

4. 適用結果とその評価

〈4・1〉 MMDB 構築実験への適用結果 以上提案してきた階層型の推論連係方式を用いたプログラムを試作し、電力需要家地図システムのための MMDB を構築する実験を行った。構築の規模は、東京都の中

表 5 適用結果
Table 5. Table of matching results.

区名	住所体系	総件数	一致件数	無効件数	連係率
中央区	新住所 旧住所	52,950 0	32,610 0	2,518 0	64.66
千代田区	新住所 旧住所	17,343 15,610	10,990 5,810	899 190	66.83 37.68
港区	新住所 旧住所	119,825 271	68,891 109	6,011 0	60.53 39.78
合計	新住所 旧住所	190,118 15,884	112,491 5,919	9,428 190	62.26 37.72
総計		206,002	118,410	9,618	60.30

中央区・千代田区・港区の3区内の需要家206,200件を対象とし、その推論連係処理の詳細な結果を表5に示す。

ただしこの表5において、連係率の算出には次のような計算式を用いている。

$$\text{連係率} = \text{一致件数} / (\text{総件数} - \text{無効件数})$$

この式において無効件数とは、既に〈3・3〉節で述べた連係レベルの値が5以下のものを対象にしており、同じ住所に一致する需要家名義が存在しないものや、カナ漢字変換が不可能なものなど、自動的に連係処理が適用できない場合を対象にしている。また、表5において住所体系が新住所と旧住所とが存在することは、住居表示の変更や地番整理などの施行が進行中で、新住所への切替わりの過渡期にある地域であることを示している。

以上、表5より本提案の階層型の推論連係方式を適用した試作プログラムによる処理の結果は、需要家総数：206,002件に対して、平均60.3%の連係率が得られた。そしてこの206,002件のデータ全体を推論連係するために要した処理時間は、汎用大型計算機システムHITAC M-680を用いた場合、約50時間を要した。

〈4・2〉適用結果の解析と評価 本推論連係方式を適用することの効果を評価するために、同一のDB間を対象にして、住所と個人名義間の完全一致を用いた連係処理の実験も行った。その結果、完全一致による平均連係率は31%と低く、今回提案した階層型の推論連係方式を適用することにより約30%の連係率向上が期待できることがわかり、その有効性を確認できた。しかし今回試作したプログラムでは、平均連係率が60.3%とまだ低く、後述の対話による連係処理の負荷が大きいため、まだ改良の余地が残っていると判断される。

そこで推論連係をはばむ要因を解析し、その対策方

表 6 自動連係できない要因とその対策
Table 6. Unmatching cases and solutions.

区分	要 因	対 策
一戸建	一つの住宅に複数のお客さま契約が存在する	複数の需要家契約を1住宅に対応させる
	同じ名義の住宅が複数存在する	判断不可能なため連係不可能とする
集合住宅	需要家データと集合住宅データの部屋番号の管理が一部異なる	需要家データと住宅地図間の部屋番号管理変換表を作成
	需要家データの中には集合住宅全体を1個人とするものがある	需要家の集合住宅全体を管理するデータを集合住宅に対応させる
雑居ビル	需要家データの中にはビル全体を1個人として管理するものがある	需要家のビル全体を管理するデータをビルに対応させる
	一つの部屋で複数のお客さま契約が存在する	複数の需要家契約を集合住宅の1部屋に対応させる
	1個人または企業が一つのビルの複数階を使用している	ビル内の複数の部屋を同一の需要家契約に対応させる

針をまとめると表6のようになった。すなわち、今回試作した推論連係のためのプログラムでは、集合住宅や雑居ビルに居住する需要家と住宅地図データとの対応関係が1対1にならない場合に対して、十分対処できていないことがわかった。このような場合には、通常の一戸建ての需要家を処理する場合とは独立に実行できるような改良を加える必要がある。この集合住宅や雑居ビルの対応を行う一方式を、後述の〈4・3〉節で説明する。

一方、今回連係の処理対象とした需要家属性DBと住宅地図DB間との間には、調査時期に関する時間的なずれがあることがわかった。すなわち、表5にも示したように、旧住所から新住所への住所変更過渡期にある地域が含まれるうえに、需要家属性DBの調査時期と住宅地図DBの調査時期とが2年以上ずれている場所が多いことがわかった。推論連係の詳細な地区別にまとめた結果を調査すると、連係率が80%に達する地区が随所にみられた。

今回提案した推論連係方式に基づくプログラムを実際の業務レベルで使用する場合には、このような整備時期のずれをなくすることが可能である。従って今回試作したプログラムによっても、連係処理の対象とするデータベース間の整備時期を合わせることで、この80%程度の連係率に達する見通しが得られたことになる。

一方、今回提案した推論連係方式では、連係の方向が省略や類語の多い需要家名称から、正式の漢字名称

図 6 半自動対話システムの画面例

Fig. 6. A screen from the semiautomatic matching system shows candidate-data in order of matching certitude.

で記述されることの多い住宅居住者名称を推論する方向のアルゴリズムとなっている。しかし、本提案方式をより汎用化するためには、上記と逆の正式名称から省略や類語を含む名称を推論する方向のアルゴリズムにもなっている必要がある。このためには、〈3・3〉節で述べた前向き推論のほかに、後ろ向き推論能力を備える必要がある。そして、この後ろ向き推論を活用することにより、更に高精度の推論連係が得られる可能性があるが、その効果は確認できなかった。

〈4・3〉 MMDB 構築実用化のための検討 需要家地図システムのための MMDB を実用レベルで構築するためには、推論連係を用いても連係できずに残ったデータをどのように効率良く対処して、連係を完全にするかといった課題を解決する必要がある。このための基本方針は、データベースの視覚化と効率の良い対話操作環境を作成することである。この場合、需要家属性 DB と住宅地図 DB とを全く独立に処理するのではなく、MMDB の構築過程のなかで、連係情報のない状態でもキー指定から該当候補が相互に検索可能な対話システムが有効と考えられる。

この対話システムの構成例として、需要家属性データと住宅地図データとを住所コード・街区で結びつける住所ファイルを設け、両データが 1 対 1 に結びつい

図 7 集合住宅・雑居ビルのための対話システム画面例

Fig. 7. A screen from the semiautomatic matching system for apartment houses and office buildings.

ていない場合でも住所データを経由することにより、相互に候補が検索可能な半自動の対話システムを試作した。この半自動対話システムでは、図 6 のような地図表示画面の状態から、地図上の家屋を指定するだけで該当する需要家の候補リストが検索できたり、逆

に需要家データを指定するだけで関係候補となる家屋が地図上にマーク表示される。この半自動対話型システムは、階層型の推論関係処理と連動させることが可能なため、推論関係結果の確認と並列させて実行可能となり、対話処理効率は大幅に向上する。

更に、〈4・2〉節で述べた関係をはばむ要因の一つである「集合住宅や雑居ビルに居住する需要家と、住宅地図データとの対応づけ」の対策として、図7に示すような専用の対話処理を試作した。現在関係先に用いている住宅地図データは、二次元の地形図データが基本になっており、集合住宅の階数や部屋番号といった三次元的なデータは、家屋図形にリンクさせた別記情報となっている。

すなわち、地形図の段階で家枠図形と居住者とは1対Nの関係になっている。そこで集合住宅の家枠とは別に三次元的な仮想的家枠を発生させ、その仮想的な家枠と集合住宅居住者を1対1に対応づけることにより、一戸建の住宅の関係処理と同様な、統一的な関係処理を実現した。

5. む す び

電力会社における「需要家地図システム」で使用する大規模なマルチメディアデータベースを自動的に構築する方法として、大容量の需要家データベースと住宅地図データベースとを高精度に関係させる方式を検討した。その関係方式として、関係時の手がかりとなる需要家のカナ名称から関係先の住宅居住者の漢字名称を推論する新しい方式を提案した。この推論関係方式を、実際の東京都心部の20万件のデータを用いたデータベースの構築実験に適用しその効果を測定した。

その結果、直接的な関係処理では関係率が30%程度の質の悪いデータベースでも、本推論関係方式により60%にまで向上させることが可能となること、更に関係対象のデータベースの調査時期が合致した質の良い場合には、80%程度の関係率まで向上可能な見通しがあることを確認した。そして、自動的に関係できなかったデータを効率良く対話的に関係するための半自動の対話処理を試作し、需要家地図システムのための大規模なマルチメディアデータベースを実用レベルで構築可能とする見通しを得た。

今後推論部分に後ろ向き推論能力を与えることにより、推論関係精度の向上を図るほか、関係処理のための計算機処理時間が現行の1/4以下となるようなプログラム改良を図る予定である。

(平成4年6月26日受付)

文 献

- (1) 三枝, 他: 「地理情報システム WING」, 昭58 情報処理学会大会, 6M-2
- (2) 東明: 「地下埋設物管理のためのコンピュータマッピングシステム」, PIXEL, No. 39, p. 131
- (3) 武崎: 「電話帳 DB と住宅地図 DB の結合方法の検討」, 昭62 通信学会大会, 1413
- (4) S. Simada, K. Suzuki, Y. Yosizaki & N. Chikada: "Object-oriented Multimedia Annotation for Large-Scale Mapping System", Proc. The second International Symposium on Database Systems for Advanced Applications (DCSFA '91), p. 219
- (5) 金森, 他: 「ES/KERNEL での知識表現方法と高速推論方式」, 日立評論, 63, No. 3, 219
- (6) A. O. Aho & M. J. Corasick: "Efficient String Matching; An Aid to Bibliographic Search", *Comm. ACM*, 18, No. 6, 333 (1975)

近 田 伸 行 (正員)

昭和23年7月6日生。47年3月電気通信大学電子工学科卒業。同年4月東京電力(株)入社。同社システム研究所情報研究室に勤務。主として、地図システム、AI 応用などの研究に従事。

山 口 剛 (非会員)

昭和33年7月7日生。56年3月東海大学電気工学科卒業。同年4月東京電力(株)入社。同社システム研究所情報研究室に勤務。主として、地図システムなどの研究に従事。

嶋 田 茂 (非会員)

昭和50年3月名古屋工業大学工学部生産機械工学科修士課程修了。同年4月(株)日立製作所入社。中央研究所にて画像処理・図面認識・地図情報処理システム・知識ベースなどの研究に従事。現在、同所研究員。

山 本 直 樹 (非会員)

昭和34年11月12日生。60年3月慶応義塾大学電気工学科卒業。同年4月(株)日立製作所入社。同社システム事業部に勤務。主として、公益企業における大規模システムの開発に従事。平成3年6月マサチューセッツ工科大学経営学部修士課程修了。