

社団法人 電子情報通信学会  
THE INSTITUTE OF ELECTRONICS,  
INFORMATION AND COMMUNICATION ENGINEERS

信学技報  
TECHNICAL REPORT OF IEICE,  
DE 94-2 (1994-05)

## 辞書ベースの連想記憶による文脈理解のための場面同定機構

角田 達彦\*, 田中 英彦

{tsunoda,tanaka}@mtl.t.u-tokyo.ac.jp

東京大学工学部

〒113 東京都文京区本郷7-3-1 東京大学工学部電気工学科 田中英彦研究室  
03-3812-2111(7413)

\*日本学術振興会特別研究員

あらまし 脳が言語を受理する過程では、語義の曖昧な言葉は前後の文脈から推測し決定しつつ推論を進めるものと予測される。この機能は意味および状況を把握し最尤解を求める知的データベース検索の決定過程にも必要とされる。連想と呼ばれるこの脳機能は計算論的レベルでは目的関数の評価値を最大にする統計的決定理論として捉えられるが、学習時には目的関数は未知であり一元的学習はできない。ここでは小数の語による早期の最尤選択をすることを目的にベイズ識別の線形近似関数を用い、論理による目的関数評価の結果からフィードバックを行なう方式を選択し評価した。さらに自然言語処理やデータベース検索を補助する実用的応用例として視覚情報を基に記述した辞書による日常生活の場面同定システムを構築し、解析評価の結果、4語の文脈情報保持により90%の同定率が得られた。

和文キーワード 場面、辞書、連想記憶、文脈、多義性解消、統計

## Scene Identification Mechanism for Understanding Context by Associative Memory with Pictorial Dictionary

Tatsuhiko TSUNODA \*, Hidehiko TANAKA

Tanaka Hidehiko Laboratory, Department of Electrical Engineering,

Faculty of Engineering, University of Tokyo

7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan

{tsunoda,tanaka}@mtl.t.u-tokyo.ac.jp

\*Supported by the Fellowships of the Japan Society for the Promotion of Science for Japanese Junior Scientists

Abstract Our brains seem to be disambiguating words by deciding each meaning with accompanying words in communicating processes. This brain function is also necessary to intelligent data-base retrieval systems with context understanding ability. Associative inference, proposed for this aim, is interpreted as one of statistical decision processes to maximize a evaluation value of objective function in the level of computational theory. However, the objective function is unknown to the system in learning phase. We use a linear approximating function of the Bayes discrimination for the purpose of taking a maximum-likelihood solution with least number of input keys in earlier phase. The logical objective function checks it and give negative feedback to the associative memory. We also constructed an scene identification system based on a pictorial dictionary as an module of natural language processing systems and data-base retrival systems. It has an identification ability of 90% ratio by preserving four previously appeared words as context.

英文 key words scene, dictionary, associative memory, context, disambiguation, statistics

## 1 はじめに

脳を理想的情報処理装置としモデル化する場合、最も解決が困難で複雑な問題の一つは文脈依存の曖昧性解消問題である。機械翻訳などの自然言語処理や柔軟な知的データベースに要求される処理は、曖昧性を含んだ不完全な情報から曖昧性のない完全な情報へ何らかの知識を利用して変換することである。近年、一般の辞書を利用した大規模な曖昧性解消の研究例はあるが [1]、処理方法が必ずしも脳機能を反映せず非常に不安定な動作をする、辞書の知識の常識的側面が欠ける、新たな知識獲得の方針がとれないなどの問題がある。

この論文では状況依存の曖昧性解消の決定推論をする脳機能のモデル Parallel Distributed Associative Inference and Contradiction Detection (PDAI&CD)[2] の統計的決定論からみた計算論的意味付けを行ない、論理的な目的関数の評価によるフィードバックによる最尤解選択方法の重要性について検討する。さらに自然言語処理および知的データベース検索における曖昧性解消問題や話題依存の知識源(辞書)選択への応用を目的に、常識を扱う試みとし日常生活の一般場面に対して視覚情報と対応する記号を用いて表現した辞書を基に上の方法を用いて場面同定システムとしてモジュール化 [3] し、評価を行なった結果を示す。

## 2 文脈依存の曖昧性解消問題

通信で既知のごとく、情報伝達において表現の効率化に対立するものは情報の曖昧性である。不完全情報、曖昧性を含んだ情報の伝達は効率上有利であるものの欠損した情報量を補完する知識や前後の文脈を必要とする。自然言語理解で良く知られた次の例題 [4] をとりあげる。

1. John shot some bucks.
2. The astronomer married the star.

前者は shoot という単語、buck という単語には各々十数通りの意味があるので、文全体の解釈としては数百通りあるという例題である。だが native speaker は文脈に応じた解釈ができ、

1. 狩猟：ジョンは鹿を撃った。

2. ギャンブル：ジョンは何ドルかすってしまった。

という数通りに絞り込むことができる。また後者は「天文学者は星と結婚した」と言いそうになって「いやいや、映画スターと結婚した」と言い直す例題である。天文学者という単語から天文的な情景を思い浮かべ、「星」にアトラクトしてしまう semantic garden path 問題 [4] [5] である。

またデータベース検索でも自然言語と同様に常識を扱う問題が存在する。

- キーワードの多義性
- 解を絞り込みのキーワード数最小化
- 話題に応じた知識源の選択 (情報処理能力の制約)

いずれの場合も

- 文脈、意味の定義方法
- 効率の良い処理方法
- 知識獲得の方法

が共通の課題となる。

## 3 PDAI&CD

上述の曖昧性解消問題は一般化すると推論の決定問題に帰着する。

- 入力：  $x_i (i = 1, 2, \dots, n)$  (単語群とする) → 特徴ベクトル空間  $X = \{x \in R^N\}$
- 出力：
  - 多義性解消：各  $x_i$  に対する意味素  $y_{ij} (j = 1, 2, \dots, m)$  からの選択
  - データベース検索：検索集合  $C_j (j = 1, 2, \dots, m')$  からの選択

であるが、両者は曖昧性のある情報を用いて 1 つの解を決定する点で統一的に論じることができる。ここでは簡単のため後者を論じる。

- 類  $C_j (j = 1, 2, \dots, k)$
- 概念集合  $\Omega = \{C_j\}_{j=1}^k$

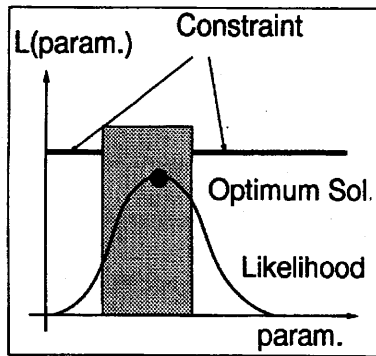


図 1: 最尤解選択と目的関数 (制約) 評価

- $P(C_j)$ : 各類  $C_j$  の先験確率 ( $\sum P(C_j) = 1$ )
- $p(x | C_j)$ : 条件付確率密度分布 ( $\int p(x | C_j) dx = 1$ )
- $p(C_j | x) = \frac{P(C_j)p(x|C_j)}{p(x)}$  ( $p(x) = \sum P(C_j)p(x | C_j)$ )

● 決定: 統計的決定論に従う

- $a = d(x)$ : 入力 (単語群) に対して行動  $a$  と決定する決定関数
- $r(C_i | C_j)$ :  $C_j \rightarrow C_i$  と誤った場合の損失関数

として、

$$R[d] = \sum_j \int r(d(x) | C_j) P(C_j | x) p(x) dx$$

の最小化をする決定関数  $d(x)$  を求める。これを実装・実用化する場合、以下のようないくつかの問題が発生する。

- 問題点 1: 損失関数  $r(C_i | C_j)$  は目的が決まるまで決まらず、予測の必要性があったり、非線形性がある場合がある
- 問題点 1 の解決方法:

1. 0-1 損失関数を用いた連想記憶部
2. 離散的目的関数の論理関数による実現

の 2 段階に分け、前者で求めた最尤解を評価する方法をとる。脳は学習時は分布の最尤推定を行なっていると仮定し、0-1 損失関数

$r(C_i | C_j) = 1 - \delta_{ij}$  を採用する。すると  $d(x)$  はベイズ識別 [6] に帰着する。

$$x \in C_i \text{ if } p(C_i | x) \geq p(C_j | x), \forall j = 1, 2, \dots, k.$$

ここで、入力ベクトルの各要素が相互の独立分布であり、かつ完全情報が得られた場合を想定すると、

$$p(C_j | x) = \prod_i p(C_j | x_i)$$

となる。

- 問題点 2: 全ての場合 ( $x$  が不完全、ノイズ混入時を含めて) について  $p(C_j | x)$  を予め求めておくことは事実上不可能である。一般に情報の統合は ill-defined problem である。そこで次の要請を満たすものを考える。

1. 何らかの近似によって表現する
2. 少ない語数での識別を目標にする

● 問題点 2 の解決方法:

1. 識別関数を線形関数にして近似
2. 各要素に対する確率に指数分布関数族の独立分布を仮定
3. 1 要素に対する尤度に基づく尤度関数の設定

尤度識別関数を以下のように設定する。

$$L(C_j | x) = f\left(\sum_i W_{ji} x_i\right) \quad (1)$$

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

$$W_{ji} = P(C_j | x_i) \quad (3)$$

$$= \frac{p(x_i | C_j) P(C_j)}{p(x_i)} \quad (4)$$

$$L(C_{j^*} | x) = \max_j [L(C_j | x)] \quad (5)$$

最後の式が最尤解選択で、ハードウェアレベルでは Winner-take-all ネットワークで選択され、上述の論理処理部の目的関数で評価される。

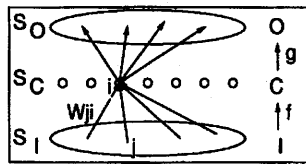


図 2: 簡易化モデル

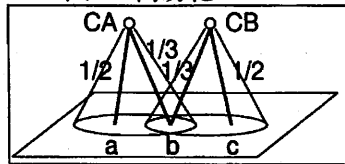


図 3: リンクの重みづけとカテゴリ選択

### 3.1 連想記憶部の理論的解析

上のように近似した連想記憶の一般的な理論的解析は困難なので、ここでは特殊例として次の場合について解析 [7] を行なう (具体的な大規模実験に関しては次章に記述)。

$$p(C_j | x_i) = \begin{cases} p(C_l | x_i), (l = 1, 2, \dots, k) \text{ or} \\ 0 \end{cases}$$

すなわちある要素から各カテゴリへの重みの分散が生じない場合である。この場合、一つのカテゴリを構成する要素の集合の中から、組合せとして唯一性のある部分集合を取り出せば、他のカテゴリから有意に識別することが可能である。逆に、組合せとして唯一性のある部分集合を取り出せるかという組合せを求める問題に帰着する。例えば図 3 の要素 b のみで  $C_A$  か  $C_B$  か決めることはできない。だが、a のみ、あるいは a と b が両方とも入力されると選択できる。この分類性を定量化する。要素数  $n$ 、カテゴリの関係概念要素数 (リンク数)  $r$ 、パタン数 (カテゴリ数)  $k$  とし、完全情報は  $r$  要素すべてを入力した場合に相当する。不完全情報は要素数  $s (0 < s < r)$  からなる部分ベクトルである。二つのパタンが  $s$  個の要素を共有する確率は、

$$V(n, r, s) = \frac{r C_{s-1} n-r C_{r-s-1}}{n C_r}$$

となる。  $m$  パタンの場合に拡張するとき、一つ以上のパタンが  $s$  個の要素を共有し、他のパタンは  $0 \sim s-1$  個までの要素の共有を許すとし、必要半径  $s$  に対応した確率を求める ( $P(n, r, 0, k) = 0$ )。

$$P(n, r, s, k) \tag{6}$$

$$= \left( \sum_{t=1}^s V(n, r, t) \right)^{k-1} - P(n, r, s-1, k) \tag{7}$$

$$= (p_1 - p_2) \left( \sum_{q=0}^{k-2} p_1^q p_2^{k-2-q} \right) \tag{8}$$

$$= V(n, r, s) \left( \sum_{q=0}^{k-2} p_1^q p_2^{k-2-q} \right) \tag{9}$$

$$\left( p_1 = \sum_{t=1}^s V(n, r, t), p_2 = \sum_{t=1}^{s-1} V(n, r, t) \right)$$

後述の UNIX コマンドでの実験の結果に合わせ、全要素数  $n=191$ 、カテゴリ毎の関係概念数  $r=3$ 、パタン数  $k=144$  を中心にし、必要半径、関係概念数、パタン数をパラメータとして変化させたときのシミュレーション結果を図 4、5、6、7 に示す。

- 図 4 は  $n=191$ 、 $r=3$  のもと、記銘パタン数  $k$  を 2 から 200 まで変化させた場合の部分パタンの要素数と想起される確率を示す。要素数が 1 のみの場合には記銘パタン数が多くなると想起されにくいだが、2 要素が入力されると、概して想起される確率が非常に高くなる。
- 図 5 は  $n=191$ 、 $r=3$  での、各必要半径の割合のパタン数 ( $k$ ) 依存の推移を示す。必要半径 3 のグラフはパタン数とともにだだらかに上昇し、必要半径 4 (即ち全要素呈示で想起不可) はほぼ 0 である。
- 図 6 は  $n=191$ 、 $r=3$  の必要半径の期待値の記銘パタン数 ( $k$ ) 依存性を示す。数パタンの記銘では要素数は 1 個示すとかかり絞れるが、50 パタンを超えると、2 個以上の要素を示さない絞れる確率がほとんどないことがわかる。だが記銘パタン数の増加に対する必要半径の期待値は 2.1 付近に飽和していく。
- 図 7 は  $n=191$  での  $r$  を 1 から 7 まで変化させたときの必要半径の期待値を、記銘パタン数毎に示している。 $r$  依存性は非常に小さく、このことはカテゴリ毎の関係概念要素数は多い方が必要な部分情報数の割合が小さくなる点で有利であることを示す。

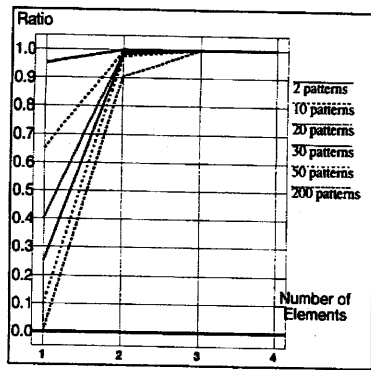


図 4:  $n=191, r=3$  の部分情報要素数と選択確率

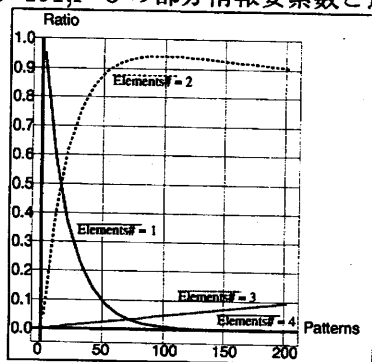


図 5:  $n=191, r=3$  の確率分布のパターン数依存性

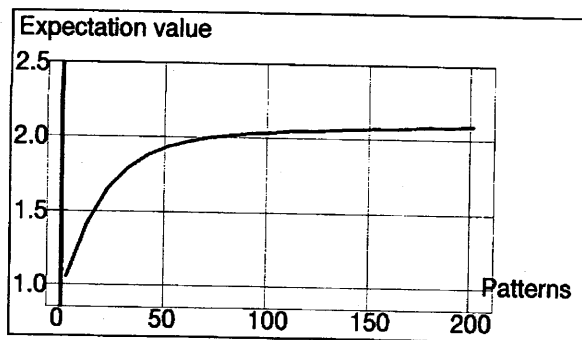


図 6:  $n=191, r=3$  の必要半径のパターン数依存性

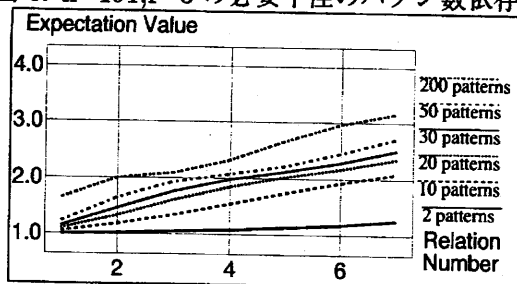


図 7:  $n=191$  の必要半径の関係概念数依存性

### 3.2 連想記憶部の小実験と評価

具体的実験例として、UNIXのコマンド連想を評価し、比較する。UNIXのコマンド連想は、Workstation上の/usr/bin中の144コマンドに対し、コマンド毎に特徴的キーワードを連想した結果に基づく。例えば、pwdに対してshow、current、working、directoryが対応している。これらのキーワードは多くのコマンドに共通に現れる単語で、一つだけではコマンドを特定できないという多義語の例である。それらのキーワード(要素)の組から逆にコマンド(カテゴリ)を想起することを目的とし、部分的キーワードを与えて想起される確率を割出す。その結果を表に示す。部分的キーワードセットの全組合せ事例数62,634での確率と必要半径の期待値を示す。

必要半径	1	2	3	4	5	6-
確率	0.40	0.41	0.18	0.02	0.00	0

必要半径の期待値 1.82

全要素数は191、カテゴリ数(記銘ボタン数)は144である。カテゴリ毎の関係概念要素数のばらつきは以下の通りであり、平均要素数は3.0である。前章のシミュレーションではこの値を用いた。

関係概念要素数	1	2	3	4	5	6	7
カテゴリ数	10	42	41	37	12	1	1

実験上の必要半径の期待値1.82と予想値(2.06)の違いは、以下の表のように、コマンド特有の概念と特有でない概念の分散が激しいためと思われる。

関係カテゴリ数	1	2	3	4	5	6
要素数	116	30	15	8	8	4
関係カテゴリ数	7	8	12	13	19	29
要素数	4	4	2	1	1	1

このように必要半径は1.82であるので、2語のキーワードでかなり絞り込めることがわかる。逐次に情報が呈示される場合には、文脈として1語分以上の情報を保持すれば良いことがわかる。

### 4 辞書による場面同定

より大規模な実験と実用的な応用を兼ね、視覚情報に基づく辞書を用いて日常生活のほとんどの

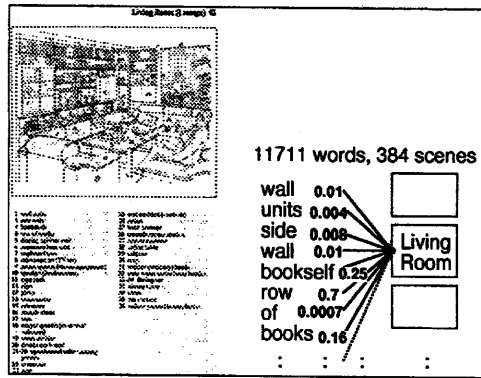


図 8: 居間のシーンと連想記憶 WAVE での重みづけ

場面を、入力された少量の自然言語文やキーワードを基に同定するシステムを上記の方法を用いて構築し、評価を行なった。脳機能としての連想を Kohonen の分類に従い、(a) 空間的連想 (b) 時間的連想 (c) 類似 (d) 反対の 4 種類に分け、ここでは (a) の空間的連想を外界 (Real world) からセンサを通じて獲得することを目指している。視覚情報に基づく辞書を用いることは、このような将来のセンサによる関連づけ (association) の第 1 近似とみなし常識を扱う知識となる。

#### 4.1 視覚情報を基にした辞書

OXFORD-DUDEN Pictorial English Dictionary (OPED) は、あらゆる年齢層を対象に、欧米の日常生活の場面が図 8(左) のように絵とそれに対応する英単語 (特に事物名) によって説明されている辞書である。11 クラスの大分類の下に 384 のカテゴリがあり、それぞれの絵の事物に対応する語が列挙されている。さらに絵でセグメント分けをすると数千シーンが含まれている。ここでは 384 カテゴリを列挙されている語により同定することを目指す。この辞書を使用するにあたり、以下のような仮定、実装を行なった。

- OPED にある日常生活の場面のみを扱い、組合せによりほとんどの場面を近似できると仮定する。
- 構文情報は使わず、OPED の単語のみ扱う。
- 形態素解析は (株) 日本電子化辞書研究所の EDR を使用。

辞書中の 1 場面に複数の単語のセットが対応づけられ (図 8) ている。場面に多義性のある場合 (例

えば game = 1. ゲーム 2. 狩猟動物) もあれば、同一の形をとる単語 (例えば cushion) が複数の場面 (例えば居間のソファの上のクッション、ピリヤードのテーブルの周りのクッション) に現れる場合もある。また明示的な多義性を持つ単語のみならず、例えば wall(壁) でも周りに chair や dining table がある文脈 (居間) と、street や car がある文脈 (屋外) では、違う意味合いを持つ。このような場合、カテゴリとして場面を決められれば、情景に応じた多義性解消が可能になる。場面の単語セットの部分集合が与えられた場合に、その場面が他の場面と区別され同定されるか否かが問題となる。これを次章で解析する。

#### 4.2 連想記憶による場面同定

上述の識別型連想記憶を用い、視覚情報の辞書の内容をその結合値に反映させて実装を行なった。重みは上述のように、各単語が各場面に現れる頻度を正規化した値を用いている。例えばある単語が一つの場面にのみ現れた場合は重みは 1 であるが、3つの場面に現れた場合は  $1/3$  ずつになる。偏りが生じた場合はその偏りに比例し正規化することになる。解析時は、各場面に記述されている全単語集合の部分集合をランダムに取り出し、それを用いて同じ場面を同定できるかによって評価を行なった。

### 5 実験結果と考察

- 図 9(a) は 1 場面あたりの要素数の分布で、平均 184.2 語。(b) は単語の意味素数、つまり多義性の度合いを示す。100 場面以上に現れる単語は 'a', 'the' など (10 種類) で、場面同定には不要だが、公平さのため用いている。
- 場面の一部を入力したとき、その単語数に対する場面同定確率を図 10 に示す。理論的シミュレーション (前章参照、辞書なしのランダム分布) も表示してある。その基本的パラメータ値に関しては、表 1 のように、OPED を解析した結果を用いている。また辞書は用いているが論理積 (目的関数を固定) を使用した結果は 'EXACT' というグラフとして示してある。いずれも全解探索は事実上無理

のため、Monte-Carlo による 1,000 回の試行の結果である。また場面同定に必要な単語数の分布を図 11 に示す。

- 図 10 により、完全セットに近ければ近いほど、想起確率は上昇することがわかる。約 90% の認識率を達成するには、5 語示だけでよい。すなわち文脈として 4 語示だけで十分であることがわかる。これは場面あたりの平均単語数約 184 に比べて極めて小さい。またグラフの途中で理論的予想 (辞書なし)、論理積、線形識別型連想記憶 (OPED) のグラフが交差している事実が見られる。語数が少ない場合は、尤度による識別が効果を表し、他者よりも同定率が良い。逆に語数が多い場合は論理積を用いた方が同定率が良いことがわかる。実際のシステムでは、尤度による識別と論理積 (目的関数) を組み合わせるので、両者の包絡線が同定率になる。
- 3 者の解析結果の違いの原因は、(1) 'a', 'the' などの無意味語 (2) 語毎の重みの分散が大きい (3) 最尤解と論理積との性質の違いがあげられる。これを裏付けるものとして、2 つの場面間で重なる単語数の分布を図 12(a) に示す。OPED-2 というのは (1) を削除した結果であり、OPED との違いは顕著に現れるものの、Theory との違いを埋めるにはほど遠い。つまり、どちらの原因も本質的であるが、自然言語の連想性の偏りの存在により、理論的予測をするだけでは不十分で、実際の解析が重要であることが帰結される。語毎の重みの分散に関しては、図 12(b) に示してある。論理積を目的にした場合は、分散がなく、全ての重みが同じ値である場合が理想的である (前章参照)。しかし実際に辞書を解析すると、非常に分散が大きく、各要素の重要性の偏りのため、部分集合全てを含む場面を唯一に同定するというタスクには不利な点が生じるのがわかる。
- 辞書と連想記憶の解析結果と理論的予測を表 1 にまとめる。辞書の登録語数 27,500 と解析時の全単語数 11,711 の違いは、形態素解析の違いによる。前者では 'research laboratory' などの複合語は一つの独立した単語として与

表 1: 連想記憶上の辞書の解析結果

Total # of scenes	384 scenes
Registered # of words	27,500 words
Total # of words	11,711 words
Average # of words / scene	184.2 words
Max # of words in one scene	478 words
Required # of words to identify scenes at 90% ratio	5 words
Required # of words to identify scenes at 90% ratio by exact match algorithm	4 words
Theoretical estimation of required # of words to identify scenes at 90% ratio	2 words

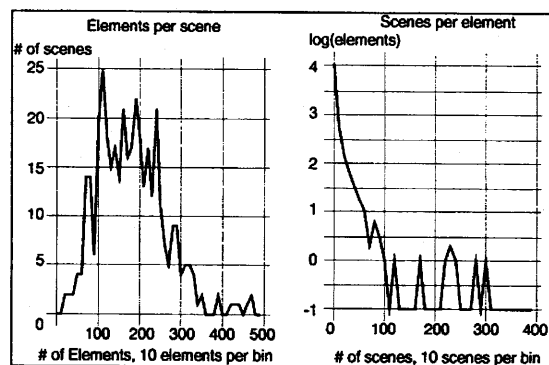


図 9: (a) シーンあたりの要素数の分布 (b) 一要素に関するシーン数の分布

えているが、後者では複数の単語として分けるためである。通常の文では分かれて出現することが多いため、後者の方法を用いている。

## 6 おわりに

脳の基本的機能の一つとしての連想推論は計算理論レベルでの統計学の決定理論と意味付けられる。学習時には目的関数は未知であるという事実から、最尤解選択と目的関数の評価の 2 段階処理および帰還を行なうことを提案し、特に最尤解選択の能力に着目して解析を行なった。

また自然言語の多義性解消のための一要素技術として日常生活のほとんどの場面を扱った辞書 OPED を連想記憶に実装し、空間的連想記憶による場面同定を実時間で行なうモジュールを初めて構築し、

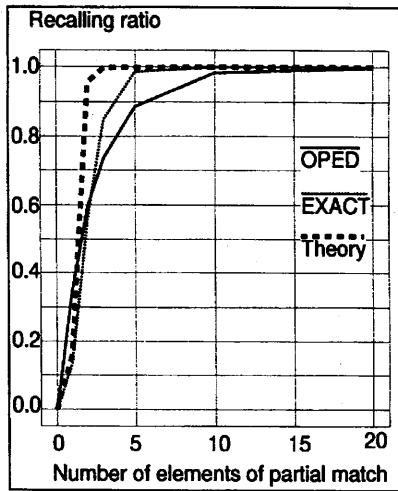


図 10: 部分要素数に対する想起確率

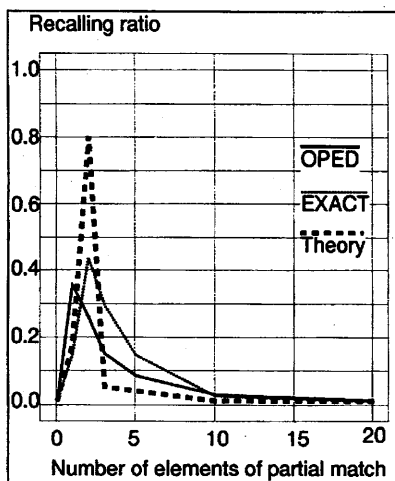


図 11: シーン同定に必要な要素数の分布

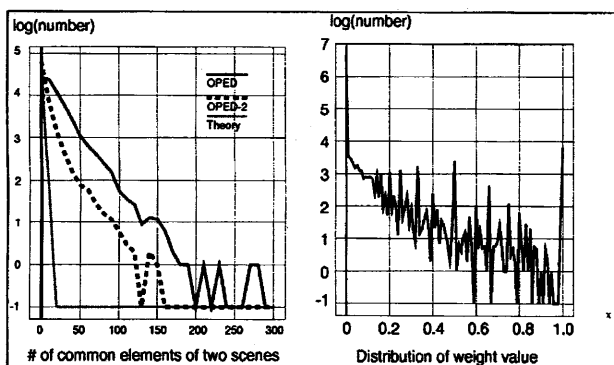


図 12: (a) 二つのシーンに共通に含まれる要素数の分布 (b) 各重み値の分布

解析を行なった。約4語を文脈として保持するだけで90%の同定率が得られている。このモジュールは将来の画像理解技術と自然言語処理の統合のための接点として用いることが想定されている。今後は、新たな場面獲得の具体的方法、そして上記の残りの3種類の連想関係の取得方法が課題である。

この研究の一部は、文部省科学研究費の助成による。またEDR電子化辞書の使用許可に対して(株)日本電子化辞書研究所に、そして日頃討論させて頂いている電子技術総合研究所の新情報計画室の方々に感謝の意を表す。

### 参考文献

- [1] N.M.Ide and J.Veronis. Extracting knowledge bases from machine-readable dictionaries : Have we wasted our time? In *KB & KS 93*, pp. 257-266, 12 1993.
- [2] T.Tsunoda and H.Tanaka. Semantic ambiguity resolution by parallel distributed associative inference and contradiction detection. In *Proceedings of IJCNN, Nagoya-93, vol.1*, pp. 163-166, 10 1993.
- [3] 角田達彦, 田中英彦. 辞書ベース連想による場面同定に必要な文脈情報量の推定. 情報処理学会第48回全国大会, Vol. 3, pp. 171-172, 1994.
- [4] D.L.Waltz and J.B.Pollack. Massively parallel parsing: A strongly interactive model of natural language interpretation. *COGNITIVE SCIENCE*, Vol. 9, pp. 51-74, 1985.
- [5] Graeme Hirst. *Semantic interpretation and the resolution of ambiguity*. Cambridge Univ. Press, 1987.
- [6] C.K.Chow. An optimum character recognition system using decision functions. *IRE Trans.*, Vol. EC-6, pp. 247-254, 1957.
- [7] 角田達彦, 田中英彦. 連想推論における逐次学習方式の定式化とその評価 - 曖昧性解消に必要な文脈情報の定量化. 情報処理学会第47回全国大会, Vol. 2, pp. 35-36, 1993.