

辞書ベースの連想記憶に基づく日常生活場面の同定

角田 達彦¹、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)

状況の解釈に基づく文脈依存の意味の曖昧性解消の問題の解決に際し、大規模化に向けて求められる要素技術の一つとして場面同定があげられる。脳が言語を受理する過程では、語義の曖昧な言葉は過去の場面ごとの記憶による感覚的情報を補い、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。この論文では小数の語による早期の最尤選択と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に基づく脳の連想機能に、視覚情報を基に記述した辞書を知識源として日常生活の場面同定システムを構築した結果を報告する。解析評価の結果4語の文脈情報保持により90%の同定率が得られている。

Identification of Daily Life Scenes by Associative Memory with Pictorial Dictionary

Tatsuhiko TSUNODA², 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

Scene identification is one of fundamental technologies to disambiguate semantical meaning of words depending on contexts. In communicating processes, our brains seem to be disambiguating words by recalling sensory data associated with words at each situation. In this paper, we report on the scene identification ability based on our original architecture and pictorial english dictionary as a data resource. The architecture consists of a associative memory aimed at picking up a maximum-likelihood solution with least number of input keys in earlier phase, and a logical objective function that gives negative feedback to the associative memory. As a result of it, the system has an identification ability of 90% ratio by preserving four previously appeared words as context.

¹日本学術振興会特別研究員

²Supported by the Fellowships of the Japan Society for the Promotion of Science for Japanese Junior Scientists

1 はじめに

機械翻訳などの自然言語処理の意味処理の問題は文脈依存の曖昧性解消問題に帰着する。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が求められる。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、

1. 非常に不安定な動作をする、
2. 一般辞書は知識の常識的側面が欠ける
3. 新たな知識獲得の方針がとれない

などの問題がある。問題の原因としては、適切な知識およびそれに応じた適切な処理を区分けして実装を行っていない点があげられる。特に、言語は脳の連想機能を最大限に利用する表現をとるという仮定に立てば、言語をデータとして扱った研究は脳機能の研究の助力となり、また脳機能を想定した認知科学的システムが自然言語処理などでの推論の基本となるという考え方もできる。

この論文では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連想記憶場面同定システムとして初めてモジュール化 [2] し、評価を行なった結果を示す。それに基づき、提案を行ってきた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) [3] 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実用性について検討する。

2 文脈依存の曖昧性解消問題

自然言語に限らず、曖昧性解消問題は全ての情報処理の基本問題である。後述の解析で用いるように、曖昧性を持つ状態は乱雑でありエントロピーが高いと意味づけられる。その曖昧性を解消する何らかの機構は情報量を持っているという。表現の上で不完全情報、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。

従来の論理記号処理のみを用いてこれらの問題を解決しようとする、可能性のある全事例を列挙する記述量、およびそれに伴う処理効率の問題が生じる。脳は少ない事例からの推量を行っていると考えられ、それは次章のような連想推論になっている。連想を概念的に分類すると [6]、

1. 空間的連想
2. 時間的連想
3. 類似
4. 反対

のようになり、さらに背景(文脈)も必要となっているが、ここでは文脈もさらに連想によって表現されるという仮定でシステムを実装する。ここでの研究では特に空間的連想に絞っている。空間的連想は五感の感覚情報とより抽象的シンボルの連合、感覚情報どうし、そしてシンボルどうしの連合と考えられる。後述の辞書を用いた場面同定では、その中でも一場面に出てくる視覚情報の個々の対象(事物)を代表するシンボルと、より抽象的な概念であるカテゴリ名(場面)との連合を扱って

いるが、将来の画像認識技術との言語情報との統合の際の近似的インタフェースになっている。例えば上の例では“One day, John went to a forest.”という文が前にあれば、森の中という場面が同定され、さらに“John shot some bucks.”という文が入力された時点で森の中で狩猟している場面の尤度が高くなり、その部分空間が選択される(後述の辞書で可能)。この過程は本来は画像などの感覚情報によるエピソード記憶および意味記憶による連合によって獲得されるものであるが、それを近似したシステムであると位置付けられる。

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) すなわち特徴ベクトル空間 $X = \{x \in R^N\}$
- 出力:
 - 多義性解消: 各 x_i に対する意味素 y_{ij} からの選択 ($j = 1, 2, \dots, m$)
 - カテゴリ同定: カテゴリ集合 C_j からの選択 ($j = 1, 2, \dots, m'$)

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 $\Omega = \{C_j\}_{j=1}^k$
- $P(C_j)$: 各類 C_j の事前確率 ($\sum P(C_j) = 1$)
- $p(x | C_j)$: 条件付確率密度分布 ($\int p(x | C_j) dx = 1$)
- $p(C_j | x) = \frac{P(C_j)p(x|C_j)}{\sum_j P(C_j)p(x|C_j)}$

- 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

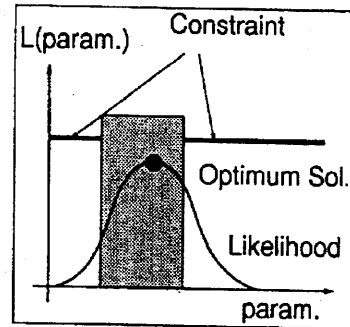


図 1: 最尤解選択と目的関数(制約)評価

として、

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

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 $r(C_i | C_j)$ は目的が定まるまで決まらず(学習時に未知)、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 [7] に帰着する。

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

- 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。例えば未知語や、学習時になかった単語群の組合せがあげられる。そのような場合には確率は0になり、推論が停止する問題が生じる。また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。

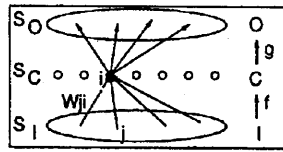


図 2: 簡易化モデル

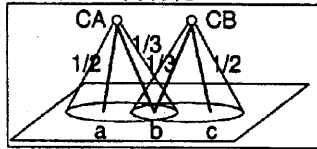


図 3: リンクの重みづけとカテゴリ選択

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
4. カテゴリの事前確率を等確率と仮定 ($\forall i, j. P(C_j) = P(C_i)$)

尤度識別関数を以下のように設定する。

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

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

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

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

$$= P(x_i | C_j) \quad (5)$$

$$L(C_{j^*} | \mathbf{x}) = \max_j [L(C_j | \mathbf{x})] \quad (6)$$

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。

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 [8] 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 が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 、ボタン数 (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。二つのボタンが 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 ボタンの場合に拡張するとき、一つ以上のボタンが s 個の要素を共有し、他のボタンは $0 \sim s-1$ 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 ($P(n, r, 0, k) = 0$)。

$$P(n, r, s, k) \quad (7)$$

$$= \left(\sum_{t=1}^s V(n, r, t)\right)^{k-1} - P(n, r, s-1, k) \quad (8)$$

$$= (p_1 - p_2) \left(\sum_{q=0}^{k-2} p_1^q p_2^{k-2-q}\right) \quad (9)$$

$$= V(n, r, s) \left(\sum_{q=0}^{k-2} p_1^q p_2^{k-2-q}\right) \quad (10)$$

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

4 辞書による場面同定

大規模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

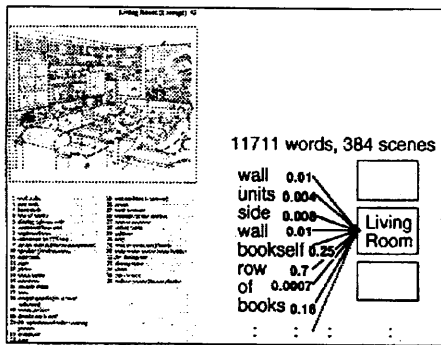


図 4: 居間のシーンと連想記憶 WAVE での重みづけ

場面の分類 :			
学問	27		模式図
行為	78		道具
職業	43		情景
物	240		単体説明
記号	5		種類
合計 393 (重複あり)			

図 5: カテゴリ名の種類と数、および概念分類

を行なった。上述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標としている。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 4(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。登録単語数は 27、500 語である。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。それを独自に分類した結果、図 5 のようになった。カテゴリ名は例えば “Living room(Lounge)” などの名詞などであるが、一般の自然言語文中では前にとる前置詞によって「情景」であったり、部屋そのものを表す「単体説明」であったり、また視覚的に分類された「種類」が列挙が求められる場合もある (例えば家のさまざまな形

など)。そのため図 5 のカテゴリ名と絵の使われ方の関係が多対多になっている。また一つのカテゴリ名に “Roof and Roofer”(屋根と屋根職人) のように、複数項目が入っている場合があるため、辞書のカテゴリ名 384 と図 5 の合計数が一致していない。

ここでは 384 カテゴリから列挙されている語により同定することを目標にする。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 4) ている。場面名自体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 lead = 1. リード線 (コード)、2. 鉛) が複数の場面 (例えば台所での器具のコードと素粒子物理学実験での鉛の放射線シールド) に現れ意味が異なる場合もある。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 情報量による理論的評価方法

OPED の視覚的場面情報を連想記憶とは独立に評価する方法として、シャノンの情報理論 [9] に基づく相互情報量を考えてみる。

会話を図 6 のように情報伝達としてとらえてみる。曖昧性が少なからず存在する単語列が話者から聴者へと渡されても、両者が共通に持つ何らかの知識によって聴者はそれらの曖昧性を解消することができる。すなわちエントロピーの高い状態を何らかの知識を用いて低い状態に変化させる知

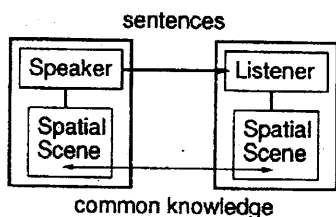


図 6: 伝達文の意味の曖昧性を解消するための話者と聴者の共有知識

識の情報量を相互情報量として求めれば良いことになる。例えばここでは単語が入ってきた場合の場面同定の曖昧性を考える。まず単語が入力される前は各カテゴリの選択は任意である(等確率)ため、

$$SI_0 = -\sum_j P(C_j) \log_2 P(C_j) = \log_2 384 = 8.59 \text{ bits}$$

である。場面が決まれば語の曖昧性も解消されると仮定すれば、語の曖昧性解消問題は場面同定問題に等しい。1語のみの単語入力後による場面同定によるエントロピーは

$$SI_1 = CE(C|X) \tag{11}$$

$$= \langle -\sum_{j_i} P_{j_i} \log P_{j_i} \rangle \tag{12}$$

$$P_{j_i} = P(C_j | x_i) = P(x_i | C_j) \tag{13}$$

と書くことが出来る(<>はアンサンブル平均)。

相互情報量は、ベイズ更新に従い他の語 (y_k, \dots) が現れたときに、前段階に比較してどの程度場面同定の曖昧性を減らせるかという問題に帰着する。 θ を前段階までの入力単語群、 θ' を単語入力後とすれば

$$MIE(\theta; \theta') = CE(C|\theta) - CE(C|\theta')$$

である。解析結果を表 1 にまとめる。全ての場合を計算すると計算量が爆発するので、Monte-Carlo で 100,000 回の試行の平均とした。相互情報量は 1 語入力の場合に非常に大きい、実際の会話で意味を持つのは 2 語入力以後の更新時である。これを用いて近似的に推測すれば、3 語入力での実用的な同定が出来ると見積もることができる。

表 1: 辞書の相互情報量

	場面同定エントロピー	相互情報量
無入力	8.59 bits	—
1 語入力	0.80 bits	7.79 bits
2 語入力	0.32 bits	0.48 bits

4.3 連想記憶による場面同定

前章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て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。結果を次章に示す。

5 実験結果と考察

- 図 7(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 8 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 2 のように、OPE D 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 9 に示す。
- 図 8 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10(a)に示す。OPED-2というものは(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10(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2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。

表 2: 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

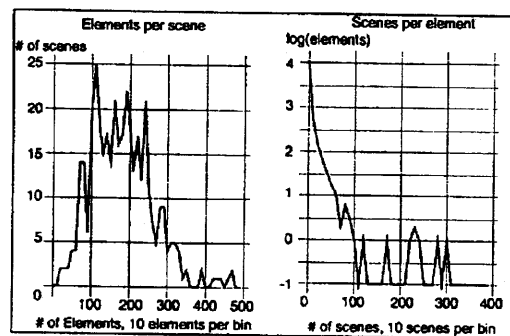


図 7: (a) シーンあたりの要素数の分布 (b) 一要素に関係するシーン数の分布

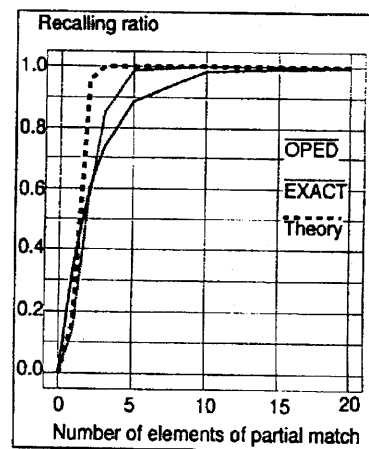


図 8: 部分要素数に対する想起確率

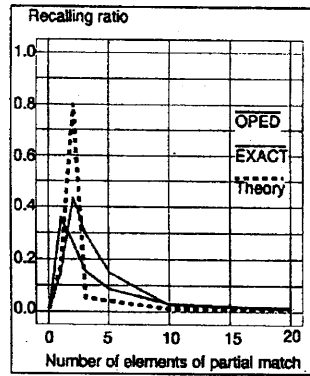


図 9: シーン同定に必要な要素数の分布

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、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

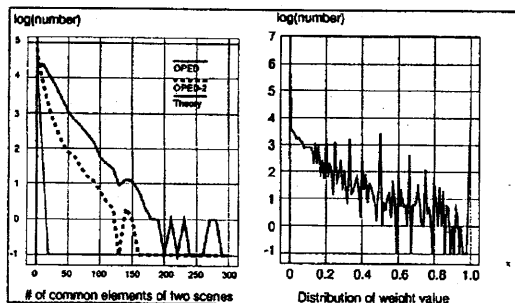


図 10: (a) 二つのシーンに共通に含まれる要素数の分布 (b) 各重み値の分布

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] 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.
- [3] 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.
- [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] T.Kohonen. *Self-Organization and Associative Memory*. Springer-Verlag, 1984.
- [7] C.K.Chow. An optimum character recognition system using decision functions. *IRE Trans.*, Vol. EC-6, pp. 247-254, 1957.
- [8] 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.
- [9] C.E.Shannon. A mathematical theory of communication. *Bell System Tech. J.*, Vol. 27, pp. 373-423, 623-656, 1948.