

k-NN推定法に基づいた名詞句の意味関係の推定

田中, 省作

九州大学大学院システム情報科学研究科知能システム学専攻 : 博士後期課程

柳瀬, 康雄

九州大学大学院システム情報科学研究科知能システム学専攻 : 修士課程

富浦, 洋一

九州大学大学院システム情報科学研究科知能システム学専攻

日高, 達

九州大学大学院システム情報科学研究科知能システム学専攻

<https://doi.org/10.15017/1515753>

出版情報 : 九州大学大学院システム情報科学紀要. 4 (2), pp.159-164, 1999-09-24. 九州大学大学院システム情報科学研究所

バージョン :

権利関係 :

k-NN 推定法に基づいた名詞句の意味関係の推定

田中省作*・柳瀬康雄**・富浦洋一***・日高 達***

Estimating Semantic Relation of Japanese Noun Phrases based on *k*-Nearest Neighbor Estimation

Shosaku TANAKA, Yasuo YANASE, Yoichi TOMIURA and Toru HITAKA

(Received June 21, 1999)

Abstract: The noun phrase “*NP* ‘*no*’ *NP*”, that consists of two noun phrases *NPs* connected by an adnominal particles ‘*no*’, is frequently used in Japanese sentences. The surface structure of this pattern is simple, but it has various semantic structures. For such noun phrases, there has been a grammar proposed, where their semantic structures can be gotten systematically from their syntactic structures. This grammar fractionates noun phrases into four syntactic categories(*CN, T, RN, EN*). As a result, the syntactic structures in this grammar can be mapped into each semantic structure. But in case that syntactic structure is “*T* ‘*no*’ *CN*”, it is necessary to infer semantic relations between *T* and *CN*, which don’t appear in the surface structure. In this paper, one method of pattern recognition (Bayes decision rule with *k*-Nearest Neighbor estimation, which is nonparametric estimation of a probability density) is applied to estimating semantic relations of a noun phrase whose syntactic category is “*T* ‘*no*’ *CN*”. As a result of the experiment, the accuracy is about 78%.

Keywords: Japanese noun phrase “*NP* ‘*no*’ *NP*”, Semantic relation, *k*-nearest neighbor estimation

1. はじめに

日本語文では、二つの名詞句を助詞「の」で結合した名詞句「*NP*₁ の *NP*₂」が頻繁に現れる。この名詞句は、多様な意味構造をもつことが知られており、しかも *NP*₁ と *NP*₂ の間の意味関係は表層的には明示されていない。そのため、名詞句「*NP*₁ の *NP*₂」の意味構造（すなわち、*NP*₁ と *NP*₂ の意味関係）を求めることは、自然言語処理の基本的な問題の一つと言える。例えば、日本語から英語などの他言語への機械翻訳を考えた場合では、名詞句「*NP*₁ の *NP*₂」における *NP*₁ と *NP*₂ の間の意味関係を推定した上で翻訳する必要がある。「私の車」という場合、「私」と「車」の間に『所有関係』が成り立っていると考えられるため、“my car”と翻訳され、「駐車場の車」という場合、『位置関係』が成り立っていると考えられるため、“a car in the parking lot”と翻訳される。また、データベースの自然言語インタフェースなどでは、名詞句「*NP*₁ の *NP*₂」における *NP*₁ と *NP*₂ の意味関係を推定し、「*NP*₁ の *NP*₂」が何を指示しているのか求める必要がある。

このような多様な意味関係をもつ名詞句「*NP* の *NP*」に対して、Montague の形式化⁵⁾に基づいて形式的に型論

理式の形で意味構造を与えるような文法体系が提案されている⁸⁾。この文法体系では、名詞句を意味的観点から更に四つの統語範疇（普通名詞句 *CN*、項句 *T*、関係名詞句 *RN*、事象名詞句 *EN*）に細分化することによって、名詞句の統語構造と意味構造が対応付けられている。名詞句「*NP* の *NP*」の意味構造には、一般に「私の車」の『所有関係』や「駐車場の車」の『位置関係』といった表層表現中には明示されない意味関係が含まれる。提案されている文法体系では名詞句の統語構造が「*T* の *CN*」である場合のみ、意味関係を別に推定する必要がある。そこで、本論文では統語構造が「*T* の *CN*」である名詞句の意味関係を、ノンパラメトリックな確率密度推定法である *k*-Nearest Neighbor 推定法（以後、*k*-NN 推定法と記す）を用いた Bayes 決定法に基づいて推定することを試みる。

まず、2章で文献 8) の文法体系を概括し、統語構造が「*T* の *CN*」である名詞句の意味関係について述べる。このような意味関係としては、文献 4) に従い、『所有関係』や『位置関係』といった基本的な関係（一般的な意味関係）と、意味関係を動詞で記述される関係（名詞固有の意味関係）を設定する。ここで、名詞固有の意味関係とは、名詞句中の名詞に固有の動詞で関係付けられる意味関係で、例えば「琵琶湖の写真」では、動詞「写す」が「琵琶湖」と「写真」を関係付けるというものである。名詞固有の意味関係を推定するには、名詞毎に意味関係に関する知識が必要となるため、統計的に推定するための用例を収集することが難しい。そこで、本論文では名詞固有の意味関

平成11年6月21日受付

* 知能システム学専攻博士後期課程

** 知能システム学専攻修士課程

*** 知能システム学専攻

係は、一般的な意味関係とは別に推定するとし、統計的に推定する意味関係は一般的な意味関係に関するものに絞る。3章では、文献9)で提案されているノンパラメトリックな確率密度推定法 k -NN 推定法を用いた Bayes 決定法を示し、名詞句の一般的な意味関係の推定への適用を考える。4章では、3章で示した手法による実験及びその結果について述べる。

2. 名詞句「NPのNP」の意味構造

2.1 名詞句の統語範疇

文献8)で提案されている文法体系では、名詞句「NPのNP」に対して、統語規則と意味構造への翻訳規則とを対応づけるために、従来、単一の統語範疇として扱われていた名詞句を意味的観点から四つの統語範疇に細分化する。本節では、この文法体系について概観する。この文法体系では名詞句の意味構造を型論理で記述しているが、本論文では、簡単化のために名詞句の意味構造を一階述語論理で記述することにする。統語規則、翻訳規則および正確な意味構造の記述等、詳細は文献8)を参照して頂きたい。

2.1.1 普通名詞句 (CN)

「色白の美人」における「色白」や「美人」は、それぞれ“色白である”、“美人である”という性質を表している。このような名詞句の統語範疇は、普通名詞句 (common noun phrase: CN) である。統語的には、「ある」「その」「すべての」などの英語の冠詞に相当する連体詞 (本論文では、これらを限定詞と呼ぶ) が結合し、項句になるような句が CN である。

この統語範疇に属す名詞句 n の意味構造は、 n に対応する一変数述語 ' n^* ' によって、

$$n^*(x) \quad (1)$$

と表される。例えば、「色白」であれば、その意味構造は、「色白 $^*(x)$ 」となり、“個体 x が色白である”ことを表す。

2.1.2 項句 (T)

ある特定の個体や事象を指示しているような名詞句の統語範疇は、項句 (term phrase: T) である。統語的には、「太郎」や「彼」や CN に限定詞が結合した「その人」や「すべての男性」というような句が T である。この統語範疇の名詞句は、動詞の格要素や後述する関係名詞句の補項になり得る句である。

この統語範疇に属す名詞 (固有名詞や代名詞等) n の意味構造は、一階述語論理では、個体記号

$$n^* \quad (2)$$

として表される。例えば、「太郎」は、「太郎 * 」という個体記号で表す。

2.1.3 関係名詞句 (RN)

個体や事象間の関係を表す名詞句、すなわち、「 T のNP」が指示する個体と T が指示する個体の関係を表す名詞句 NP の統語範疇は、関係名詞句 (relation noun phrase: RN) である。

例えば、「兄」という名詞句は、RN である。このとき、 T である「二郎」を「の」で結合して「二郎の兄」というような名詞句を構成することによって、“二郎と兄弟関係にある個体で、かつ年上の個体”を指示することになる。ここで、「 T のRN」で T が指示する個体を、その RN の補項と呼び、「 T のRN」で指示される個体を RN の主項と呼ぶことにする。先の例では、個体“二郎”が「兄」の補項であり、「二郎の兄」で指示される個体が「兄」の主項である。

この統語範疇に属す名詞句 n は、ある個体や事象間の関係を表していることから、その意味構造は二変数述語 ' n^* ' によって、

$$n^*(x, y) \quad (3)$$

と表される。例えば、「兄」は、「兄 $^*(x, y)$ 」と表され、兄 $^*(x, y)$ は“個体 x は個体 y の兄である”ことを表す。

2.1.4 事象名詞句 (EN)

事象を表している名詞句の統語範疇は、事象名詞句 (event noun phrase: EN) である。「勉強」や「考え」といったサ変動詞の語幹や、動詞が名詞に転化した名詞などは EN である。これらの EN は、「太郎の英語の勉強」「太郎の考え」のように、「の」を介して元の動詞のように格要素を取り、事象を指示する。

2.2 名詞句の統語構造と意味構造

名詞句「 NP_1 の NP_2 」の意味構造は、細分化した名詞句の統語範疇と密接に対応しており、その意味構造は NP_1 が T か CN で大別される。

2.2.1 「CNのNP」の意味構造

「CNの」は形容詞的に機能し、NP で指示する対象を制限することになる。「CNのNP」の統語範疇は NP の統語範疇と同一である。統語構造が「CNのCN」である名詞句「 n_1 の n_2 」の意味構造は、

$$n_1^*(x) \wedge n_2^*(x) \quad (4)$$

となる。例えば、「大型の犬」という名詞句では、「大型の」は「犬」で指示する“犬である”という性質を持つ個体を、“大型である”という性質をもつ個体に制限し、その意味構造は、「大型 $^*(x) \wedge$ 犬 $^*(x)$ 」という論理式で表される。このとき、「大型の犬」の統語範疇は CN である。

2.2.2 「 T の NP 」の意味構造

「 T の」は、「 T が」や「 T を」のような格要素的に機能し、 NP の翻訳された論理式の引数に代入された形になる。「 T の NP 」の統語範疇は T である。

NP が CN および RN の場合について、それぞれ説明する。

「 T の CN 」の場合

この統語構造の名詞句の意味構造は、 T で指示する個体と CN という性質をもつ個体の集合の間の意味関係を推定し、その意味関係を表す述語 \mathcal{P} を補った上で構成される。統語構造が「 T の CN 」である名詞句「 n_1 の n_2 」の意味構造は、

$$n_2^*(x) \wedge \mathcal{P}(x_1, x_2, \dots, n_1^*, \dots, x_{k_p}) \quad (5)$$

となる。このような \mathcal{P} は、表層表現には明示されていない。例えば、「私の車」であれば、「私」「車」の名詞句間には『所有関係』が一般に成立するので、「私の車」の意味構造は、二項述語‘所有’を補い、‘車*(x) \wedge 所有(x , 私*)’と表される。但し、‘所有(x , y)’は、“個体 y が個体 x を所有している”ことを意味する。このような意味関係 \mathcal{P} の種類については、2.3節で述べる。

「 T の RN 」の場合

統語構造が「 T の RN 」の名詞句では、「 T の」が(i) RN の補項に割り当てられる用法と、(ii) RN の主項を制限する用法とがある。「 T の」が RN の補項に割り当てられる用法の名詞句「 n_1 の n_2 」の意味構造は、 n_2 自身が意味関係となり、

$$n_2^*(x, n_1^*) \quad (6)$$

となる。例えば、「太郎の母」の意味構造は、‘母*(x , 太郎*)’となる。但し、‘母(x , y)’は“個体 x は個体 y の母親である”ことを意味する。

一方、「 T の」が RN の主項を制限するような用法の名詞句は、例えば「福岡の母」などである。この場合は、ちょうど「 T の CN 」と同じように、意味関係 \mathcal{P} を導出して、その意味構造は、

$$\exists y n_2^*(x, y) \wedge \mathcal{P}(x_1, x_2, \dots, n_1^*, \dots, x_{k_p}) \quad (7)$$

となる。「福岡の母」の場合の意味構造は、‘ $\exists y$ 母*(x , y) \wedge 位置(福岡*, x)’となる。但し、‘位置(x , y)’は“個体 y が地点 x に位置する”ことを表す。

2.3 「 T の CN 」の意味関係

本節以降で用いる略記について述べておく。統語構造が「 T の CN 」である名詞句「 NP の NP 」を単に「 T の

CN 」と記し、 T, CN の主辞をそれぞれ n_1, n_2 と記すことにする。

2.3.1 一般的な意味関係

一般的な意味関係は、「 T の CN 」における意味関係のうち、『所有関係』や『位置関係』等は、多くの名詞句で成立する基本的な関係である。本論文では、一般的な意味関係として、次の8つの意味関係を設定した。

1. 所有関係

NP_1 が NP_2 の所有者であることを表す。

「私の車」、「彼の計算機」

2. 所属関係

NP_2 の所属が NP_1 であることを表す。

「九大の職員」、「日本の大使」

3. 位置関係

NP_2 の存在する場所が NP_1 であることを表す。

「机の上のパソコン」、「道端の若木」

4. 生産関係

NP_1 の作成者が NP_2 であることを表す。

「彼のプログラム」、「我が社のパソコン」

5. 時間関係

NP_1 が時間を表している。

「1時の電車」、「明日の会議」

6. 数量関係

NP_1 が NP_2 に関する数量を表している。

「10kmの道のり」、「30kgのみかん」

7. 全体-部分関係

NP_1 が NP_2 を部分としてもつことを表す。

「木の枝」、「車のタイヤ」

8. 部分-全体関係

NP_2 は NP_1 を部分としてもつことを表す。

「赤い屋根の建物」、「髭の男」

統語構造が「 T の CN 」である名詞句の意味関係を調査すると、9割以上がこの一般的な意味関係となり、残る1割弱の名詞句は、次節で述べる名詞固有の意味関係になる⁷⁾。

2.3.2 名詞固有の意味関係

名詞固有の意味関係は、名詞句中の名詞に固有の動詞で記述される関係である。例えば、名詞句「写真の人物」では、「写す」という動詞が「写真」と「人物」を意味的に関係付けている。但し、「写真」「人物」と「写す」の格関係は、それぞれ「ニ格」と「ヲ格」である。このように名詞固有の意味関係は、動詞 v と、 T と v の間の格関係 c_1 、 CN と v の間の格関係 c_2 の3つ組 (v, c_1, c_2) で記述される。

また、名詞固有の意味関係は、たとえ意味的に近くても T, CN の主辞の名詞が変わると、意味関係の動詞は一般に異なる。例えば、先の「写真の人物」の例で「写真」と意味的に近い「絵」に置き換えた名詞句「絵の人物」では、

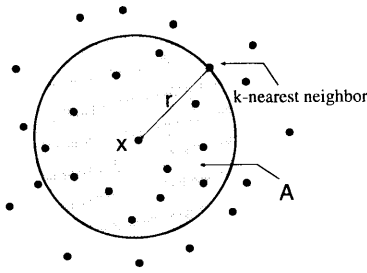


Fig.1 k-Nearest Neighbor estimation.

意味関係の動詞は「描く」となる。よって、名詞固有の意味関係については、大量の用例を収集することは難しく、統計的に推定することができない。本論文では、このような名詞固有の意味関係は、一般的な意味関係とは別に推定するものとして取り扱わない。

3. 一般的な意味関係の推定

名詞句「 T の CM 」で成立する意味関係が、一般的な意味関係のどの意味関係になるかを推定する方法として、ノンパラメトリックな確率密度推定法である k -NN 推定法²⁾を用いた Bayes 決定法を適用する。この手法は、文献 9) で提案され、自然言語の統語的曖昧さの解消（英語の前置詞句の係り先の曖昧さ）に適用した結果、高い正解率が得られている。まず、この手法を概観する。次に、名詞句の意味関係の推定にどのように適用するかを述べる。

3.1 k -NN 推定法

まず、ノンパラメトリックな確率密度の推定法である k -NN 推定法を示す。

定義 1(k -NN 推定法)

大きさ N のサンプル S^N における x の確率密度の推定値は、

$$\hat{p}(x; S^N) = \frac{k-1}{N} \frac{1}{A(k, S^N, x)} \quad (8)$$

である。但し、 $A(k, S^N, x)$ は、 x と S^N における x の k -nearest neighbor との距離 $r(k, S^N, x)$ を半径とする超球の体積である (Fig.1)。□

定理 1(k -NN 推定法の一致性)

k が、次の条件を満たす N の関数 $k(N)$ であるとき、 k -NN 推定法は一致推定である。

$$\begin{cases} \lim_{N \rightarrow \infty} k(N) = \infty, \\ \lim_{N \rightarrow \infty} \frac{k(N)}{N} = 0. \end{cases} \quad (9)$$

□

一致推定であるとは、任意の x と任意の $\epsilon > 0$ に対して、

$$\lim_{N \rightarrow \infty} P\{S^N : |\hat{p}(x; S^N) - p(x)| < \epsilon\} = 1$$

が成立することである。一致性は、統計的推定法に求められる最も重要な要請の一つである。

3.2 k -NN 推定法を用いた Bayes 決定法

パターン認識は、未知の入力 x がいずれのカテゴリに属するかを決定する手法である。本論文の一般的な意味関係を推定する問題は、未知の入力を名詞句、意味関係をカテゴリと捉え、パターン認識手法を適用する。まず、文献 9) で示されている k -NN 推定法を用いた Bayes 決定法を示す。

未知の入力 x に対するカテゴリは、 x が発生したときにカテゴリ $C_i (i = 1, \dots, m)$ に属す確率 $P(C_i|x)$ を計算し、 $P(C_i|x)$ を最大にするカテゴリ C_i を x の属すカテゴリとして決定する (Bayes 決定法)。ここで、 $P(C_i|x)$ は、Bayes の定理より、

$$P(C_i|x) = \frac{P(C_i)P(x|C_i)}{P(x)} \quad (10)$$

となる。分母の $P(x)$ は C_i に依存しないので、 $P(C_i|x)$ を最大にする C_i を求めるには、

$$P(C_i)P(x|C_i) \quad (11)$$

を最大とする C_i を求めればよい。ここで、カテゴリ C_i のサンプルを $S_i (|S_i| = N_i)$ 、サンプル全体の大きさを $N_1 + N_2 + \dots + N_m = N$ とする。すると、(11)式の第1項の $P(C_i)$ は最尤推定より、

$$P(C_i) = \frac{N_i}{N} \quad (i = 1, \dots, m)$$

と推定される。次に、(11)式の第2項の $P(x|C_i)$ は k -NN 推定法を用いると、

$$P(x|C_i) \simeq \frac{k-1}{N_i} \frac{1}{A(k, S_i, x)}$$

と推定される。よって、(10)式は、

$$P(C_i|x) \simeq \frac{k-1}{N} \frac{1}{A(k, S_i, x)}$$

となる。ここで、 k, N は C_i に依存しないので、(10)式を最大にする C_i は、 $A(k, S_i, x)$ を最小、つまり $r(k, S_i, x)$ を最小にするカテゴリである。よって、 k -NN 推定法を用いた Bayes 決定法では、各カテゴリ C_i のサンプル S_i における入力 x の k -nearest neighbor までの距離

$r(k, S_i, x)$ を求め、この値が最小になるカテゴリ C_i を x の属すカテゴリと決定する。

3.3 k-NN 推定法に基づいた一般的な意味関係の推定

前節で示した k -NN 推定法を用いた Bayes 決定法を、名詞句「NP の NP」の一般的な意味関係の推定に適用する。パターン認識の問題として、未知の入力 x を名詞句「NP の NP」、 x の属すカテゴリ C_i を「NP の NP」の意味関係と考える。また、「 NP_1 の NP_2 」の意味関係は、 NP_1, NP_2 のそれぞれの主辞 n_1, n_2 より決定できるものと仮定し、入力 x は $\langle n_1, n_2 \rangle$ とする。

次に、サンプル S は各カテゴリ（意味関係） C_i 毎に、意味関係 C_i が成立している名詞句「 NP_1 の NP_2 」の NP_1, NP_2 の主辞 n_1, n_2 の用例 $\langle n_1, n_2 \rangle$ を収集、分類しておく。 k -NN 推定法を用いた Bayes 決定法は、各カテゴリ C_i のサンプル S_i における入力 $\langle n_1, n_2 \rangle$ の k -nearest neighbor までの距離 $r(k, S_i, x)$ を求めなければならない。そこで、用例 $\langle n_1, n_2 \rangle, \langle n'_1, n'_2 \rangle$ 間の距離を定義する。

α, β, γ が用例であるとき、距離を与える関数 Δ は、次の4つの式を満たす必要がある。

$$\begin{cases} \Delta(\alpha, \beta) \geq 0, \\ \Delta(\alpha, \beta) = 0 \iff \alpha = \beta, \\ \Delta(\alpha, \beta) = \Delta(\beta, \alpha), \\ \Delta(\alpha, \beta) + \Delta(\beta, \gamma) \geq \Delta(\alpha, \gamma). \end{cases} \quad (12)$$

ここでは、用例 $\langle n_1, n_2 \rangle, \langle n'_1, n'_2 \rangle$ 間の距離を、単語 n, n' 間の距離を与える関数 δ を用いて、

$$\Delta(\langle n_1, n_2 \rangle, \langle n'_1, n'_2 \rangle) = w_1 \delta(n_1, n'_1) + w_2 \delta(n_2, n'_2) \quad (13)$$

と与える。但し、 w_1, w_2 は0以上の定数である。関数 δ が、距離を与える関数であるための条件である(12)式を満たせば、(13)式で与えた関数 Δ も(12)式を満たす。

4. 実 験

4.1 用例間の距離

本実験では、用例間の距離を与える関数 Δ を、(13)式で $w_1 = w_2 = 1$ とした、

$$\Delta(\langle n_1, n_2 \rangle, \langle n'_1, n'_2 \rangle) = \sum_{i=1}^2 \delta(n_i, n'_i) \quad (14)$$

と与えた。

次に、単語 n, n' 間の距離 $\delta(n, n')$ はEDRシソーラスを用いて与える。シソーラス上で概念 c と c' が親子関係にあるとき、 c と c' の間に距離1のパスが存在すると考え、 $\delta(n, n')$ はEDRシソーラス上での n と n' の概念間の最短のパス長とした。

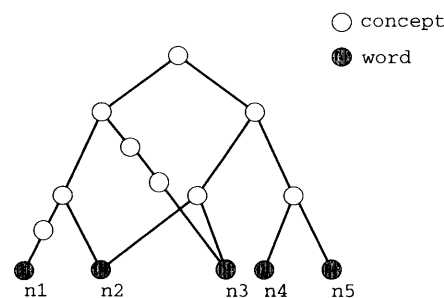


Fig.2 The distance between words.

Table-1 The number of examples at every semantic relation.

Semantic Relation	Number	Rate
Possessive	168	9.2%
'Belong to'	220	12.1%
Positional	420	23.1%
'Make'	84	4.6%
Time	650	35.7%
Quantity	221	12.1%
Has-part	50	2.7%
Part-of	8	0.4%

例えば、Fig.2のような構造のシソーラスでは、 $\delta(n_1, n_2) = 3, \delta(n_1, n_3) = 5, \delta(n_2, n_4) = 4$ である。

4.2 実験データと実験方法

本実験で用いた実験データについて述べる。EDRコーパスより抽出した名詞句「NP の NP」に対して、EDRコーパスで予め付与されている意味関係を、我々が設定している意味関係に変換し、一般的な意味関係に関する実験データを作成した⁴⁾。これらのデータ1,821個の用例をサンプル S とした。但し、部分-全体関係は、用例数が非常に少ないため全体-部分関係と一緒にしている。本実験で用いた各意味関係の用例数および割合をTable-1に示す。また、EDRコーパスでは出現している単語全てにEDRシソーラス上で対応する概念が予め一意に付与されており、単語間の距離 $\delta(n, n')$ は曖昧さなく計算される。

実験は、次のような手順で行った。まず、EDRコーパスより作成したサンプル S をランダムに10分割する。

$$S = S^1 \cup S^2 \cup \dots \cup S^{10}$$

但し、 $S^i \cap S^j = \emptyset (i \neq j)$ である。このうちの1つをテストデータ T とし、残りを改めてサンプル $S' = S \setminus T$ とし、実験を行う。但し、 $X \setminus Y$ は X の Y による差集合を表す。そして、テストデータ T を S^1 から S^{10} に変える度に、サンプル S' を $S \setminus S^1$ から $S \setminus S^{10}$ と構成し直し、 T 中のデータに対する意味関係の推定を行う。

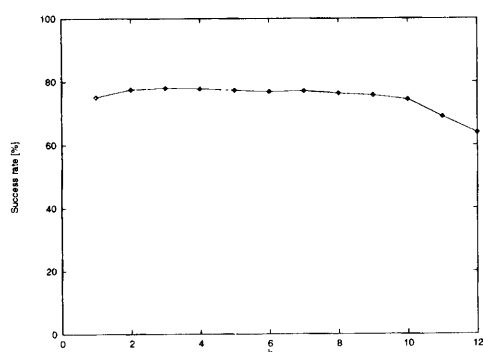


Fig. 3 Relation between k and success rate.

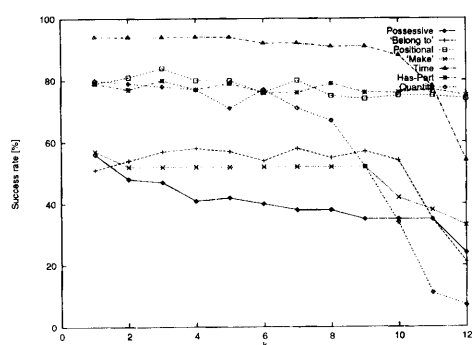


Fig. 4 Relation between k and success rate at every semantic relation.

4.3 実験結果

本手法で推定された意味関係が、予め付与されている意味関係と一致している場合を正解とし、評価を行った。正解率は、次式で計算される。

$$\text{正解率} = \frac{\text{正解のデータ数}}{T \text{ のデータ数}} \quad (15)$$

本実験では、テストデータを T を S^1, \dots, S^{10} , サンプルを $S \setminus S^1, \dots, S \setminus S^{10}$ と変え意味関係の推定を行い、その都度正解率を計算し、最終的に、これらの正解率の平均値を、本実験の正解率と考えた。

Fig. 3に、 k と全体の正解率の関係を、Fig. 4に、 k と意味関係毎の正解率の関係を示す。

4.4 考察

本実験では、 k -NN 推定法に基づいた意味関係の推定法のそのものの有効性を調べることを目的に、(13)式の用例間の距離関数 Δ 中の定数 w_1, w_2 を単純に $w_1 = w_2 = 1$ とした。実験の結果、最も良い場合で 78.4% ($k = 3$) と

り、比較的高い精度で意味関係が決定できたとと言える。また、実際には、この w_1, w_2 も大量の用例から統計的に、これらの最適な値を決定すれば、更に本手法の正解率は上昇することが期待できる。サンプルとなる用例も、一般的な意味関係の名詞句については、将来的に多くのコーパスが整備されるに従い、容易に増やすことができると予想される。

5. ま と め

文献 9) で提案された k -NN 推定法を用いた Bayes 決定法を、名詞句「NP の NP」の一般的な意味関係の推定に適用した。実際に、EDRコーパスより作成した名詞句「NP₁ の NP₂」の実験データを用いて実験を行い、その有効性を示した。

また、現在、我々は、文献 8) に基づいた名詞句「NP の NP」の意味構造システムの構築を試みている。このシステムでは、CFGの生成規則に意味関係および係り受け関係を取り込み、更に確率化することによって、名詞句の意味構造を推定する。 k -NN 推定法に基づいた意味関係の推定は、この確率化の際のパラメタ推定へも応用できると考えている。

参 考 文 献

- 1) 日本電子化辞書研究所: EDR 電子化辞書仕様説明書 (1995).
- 2) Fukunaga, K.: Introduction to Statistical Pattern Recognition, Academic Press Inc., Orlando (1972).
- 3) 平井 誠, 北橋忠宏: 日本語文における「の」と連体修飾の分類と解析, 情報処理学会研究会報告, NL58-1, pp. 1-8 (1986).
- 4) 飯田健二, 田中省作, 富浦洋一, 日高 達: 名詞句「NP の NP」の意味関係について, 電気関係学会九州支部連合大会第 50 回連合大会 講演論文集, pp. 285 (1997).
- 5) Montague, R.: The Proper Treatment of Quantification in Ordinary English, Hintikka, J., Moravcsik, J. and Suppes, R. (eds.) *Approaches to Natural Language*, Reidel, Dordrecht, pp. 221-242 (1974).
- 6) 島津 明, 内藤昭三, 野村浩郷: 日本語文意味構造の分類—名詞句構造を中心に—, 情報処理学会研究会報告 NL47-4, pp. 25-32 (1985).
- 7) 田中省作, 富浦洋一, 日高 達: 統計的手法を用いた名詞句「NP の NP」の意味関係の抽出法, 電子情報通信学会技術研究報告 NLC98-4, pp. 23-29 (1998).
- 8) 富浦洋一, 中村貞吾, 日高 達: 名詞句「NP の NP」の意味構造, 情報処理学会論文誌, Vol. 36, No. 6, pp. 1441-1448 (1995).
- 9) 富浦洋一, 日高 達: k -NN 推定法に基づく統語的あいまいさの解消法, 電子情報通信学会論文誌, D-II, Vol. J80, No. 9, pp. 2475-2481 (1997).