

2N-3

ファジィ文書検索システム(2)  
 ~キーワードコネクションマトリックスの学習方式~

小川泰嗣 森田哲也 小林清彦  
 リコー中央研究所

1. はじめに

われわれは従来方式の様々な問題点を解決する文書検索方式・システムを考案してきている[1-3]。ここでは、キーワードコネクションマトリックスと呼ぶキーワード間の関連度を記述した一種のシソーラスを導入し、ファジィ理論に基づいた文書のあいまい検索を実現している。キーワードコネクションマトリックスの初期値は、システム内に登録されている多数の文書内におけるキーワードの共出現頻度に基づいて計算される。しかし、この様に統計情報から求められた初期値は必ずしも利用者のキーワード間の関連度に対する感覚と一致しないという問題点がある。この問題点を解決するために、検索結果に対する利用者の判断からキーワードコネクションマトリックスを動的に改善する学習方式を考案してきている。本論文では、キーワードコネクションマトリックスの学習方式を詳細に説明する。以下では、まず2章で文書精度の計算方式を簡単に説明する。3章でキーワードコネクションマトリックスの学習方式の説明を行う。最後の4章で全体をまとめ、結論を述べる。

2. 文書精度の計算方式

文書精度とは、文書データベース内の各文書と利用者が指定する検索条件との関連の強さを表す数値である。検索条件は、複数のキーワードおよびキーワードに否定演算子を付加したものを論理和演算子で結合した副検索式を、論理積演算子で結合した積標準形の検索式で表現されているものとする。この場合、文書精度はつぎのように計算される。まず、h番目の副検索条件に対する文書精度を $r_i(h)$ とすると、

$$r_i(h) = 1 - \frac{\prod_{j \in Q^+(h)} S_{ij} \prod_{j \in Q^-(h)} R_{ij}}{\prod_{j \in Q^+(h)} S_{ij} \prod_{j \in Q^-(h)} R_{ij}} \quad (1)$$

ここで、否定の付かないキーワードの集合を $Q^+(h)$ 、付くキーワードの集合を $Q^-(h)$ 、 $A_i$ をi番目の文書に索引として付けられたキーワードの集合である。また、 $R_{ij} \cdot S_{ij}$ はつぎのように定義される。

$$R_{ij} = \bigoplus_{k \in A_i} W_{jk} = 1 - \prod_{k \in A_i} (1 - W_{jk}) \quad (2)$$

$$S_{ij} = 1 - R_{ij} = \prod_{k \in A_i} (1 - W_{jk}) \quad (3)$$

ここで $W_{jk}$ はj番目とk番目のキーワード間のキーワードコネクションマトリックスにおける関連度である。副検索条件に対する文書精度(副文書精度と呼ぶ)の計算の後、最終的な文書精度を計算する。

$$r_i = \prod_{h=1}^N r_i(h) \quad (4)$$

ここで、Nは検索条件を構成する副検索式の数である。

3. キーワードコネクションマトリックスの学習方式

3.1 学習の原理

キーワードコネクションマトリックスの学習では、検索結果の良し悪しを利用者が文書ごとに判断する。学習対象の文書が自分の欲していたものであれば正解、そうでなければ不正解とする。正解・不正解をそれぞれ"1", "0"に割当て、あいまいな判断は[0, 1]間の実数値tで表現する。評価関数はつぎのように定義される。

$$e(x) = \frac{1}{2} (t - x)^2 \quad (5)$$

一回の学習では、いま注目している文書(ここではi番目の文書とする)に対しその文書に対する文書精度の評価関数の値 $e(r_i)$ を小さくするように、キーワードコネクションのすべての関連度がつぎのように変更される。

$$W_{mn} \leftarrow g(W_{mn} + \alpha \Delta W_{mn}) \quad (6)$$

ここで、 $\alpha$ は正の実数の学習係数であり、 $g$ は変更した値が[0, 1]の範囲となるための正規化関数である。

$$g(x) = \begin{cases} 1 & : 1 < x \\ x & : 0 \leq x \leq 1 \\ 0 & : x < 0 \end{cases} \quad (7)$$

最急降下法を用いて変化分 $\Delta W_{mn}$ を設定すると、(5)式の関係から、

$$\Delta W_{mn} = - \frac{\partial e(r_i)}{\partial W_{mn}} = (t - r_i) \frac{\partial r_i}{\partial W_{mn}} \quad (8)$$

ここで、(4)式より、

$$\frac{\partial r_i}{\partial W_{mn}} = \sum_{h=1}^N \left\{ \frac{\partial r_i(h)}{\partial W_{mn}} \prod_{k=1, k \neq h}^N r_i(k) \right\} \quad (9)$$

実際には、上の計算式の積和の二重ループの計算が $r_i = 0$ か否かによって和あるいは積のみの単一ループの計算で行うことができる。各々の場合の計算式はつぎようになる。

A Fuzzy Document Retrieval System (2)  
 -- A Learning Method of a keyword connection matrix --  
 Yasushi OGAWA, Tetsuya MORITA and Kiyohiko KOBAYASHI  
 Research and Development Center, RICOH Co., Ltd.

### 3.1.1 $r_i \neq 0$ の場合

このとき、全ての  $h (1 \leq h \leq N)$  に対して  $r_i(h) \neq 0$  なので、

$$\begin{aligned} \frac{\partial r_i}{\partial W_{mn}} &= \sum_{h=1}^N \left\{ \frac{\partial r_i(h)}{\partial W_{mn}} \prod_{k=1}^N r_i(k) \right. \\ &\quad \left. \text{s.t. } k \neq h \right\} \\ &= \sum_{h=1}^N \left\{ \frac{\partial r_i(h)}{\partial W_{mn}} \frac{r_i}{r_i(h)} \right\} \\ &= r_i \sum_{h=1}^N \left\{ \frac{\partial r_i(h)}{\partial W_{mn}} \frac{1}{r_i(h)} \right\} \quad (10) \end{aligned}$$

### 3.1.2 $r_i = 0$ の場合

$r_i = 0$  のとき、 $r_i(h) = 0$  である  $h$  の個数により計算式が変わる。 $h$  の個数が1個の場合、そのような  $h$  を  $h^*$  とすると、

$$\begin{aligned} \frac{\partial r_i}{\partial W_{mn}} &= \sum_{h=1}^N \left\{ \frac{\partial r_i(h)}{\partial W_{mn}} \prod_{k=1}^N r_i(k) \right. \\ &\quad \left. \text{s.t. } k \neq h \right\} \\ &= \frac{\partial r_i(h^*)}{\partial W_{mn}} \prod_{k=1}^N r_i(k) \quad (11) \\ &\quad \text{s.t. } k \neq h^* \end{aligned}$$

$h$  の個数が2個以上の場合、

$$\frac{\partial r_i}{\partial W_{mn}} = 0 \quad (12)$$

いずれの場合でも  $r_i(h)$  の  $W_{mn}$  による偏微分が計算できなければならない。その計算式をつぎの節で示す。

## 3.2 副文書確度に対する偏微分の計算式

(1) 式より

$$\frac{\partial r_i}{\partial W_{mn}} = - \prod_{j \in Q^-} R_{ij} \frac{\partial}{\partial W_{mn}} \left[ \prod_{j \in Q^+} S_{ij} \right] - \prod_{j \in Q^+} S_{ij} \frac{\partial}{\partial W_{mn}} \left[ \prod_{j \in Q^-} R_{ij} \right] \quad (13)$$

であり、さらに

$$\frac{\partial}{\partial W_{mn}} \left[ \prod_{j \in Q^+} S_{ij} \right] = \begin{cases} - \prod_{\substack{j \in Q^+ \\ k \in A_i \\ \text{s.t. } (k,j) \neq (m,n)}} (1 - W_{jk}) & : m \in Q^+ \text{ \& } n \in A_i \\ 0 & : \text{otherwise} \end{cases} \quad (14)$$

$$\frac{\partial}{\partial W_{mn}} \left[ \prod_{j \in Q^-} R_{ij} \right] = \begin{cases} \prod_{\substack{k \in A_i \\ \text{s.t. } k \neq n}} (1 - W_{mk}) \prod_{j \in Q^-} R_{ij} & (15) \\ & \text{s.t. } j \neq m \\ 0 & : m \in Q^+ \text{ \& } n \in A_i \\ 0 & : \text{otherwise} \end{cases}$$

このとき、 $Q^+(h)$  と  $Q^-(h)$  に同時に含まれるキーワードはないので、 $r_i(h)$  の  $W_{mn}$  による偏微分はつぎの3つの場合に分けて計算できる。

### 3.2.1 $m \in Q^+(h)$ の場合

(13) ~ (15) 式より

$$\frac{\partial r_i(h)}{\partial W_{mn}} = \begin{cases} \prod_{\substack{j \in Q^+(h) \\ k \in A_i \\ \text{s.t. } (j,k) \neq (m,n)}} (1 - W_{jk}) \prod_{j \in Q^-(h)} R_{ij} & : n \in A_i \\ 0 & : \text{otherwise} \end{cases} \quad (16)$$

であるが、(2) (3) 式等から導出される関係を用くと、

$$\frac{\partial r_i(h)}{\partial W_{mn}} = \begin{cases} \frac{1 - r_i(h)}{1 - W_{mn}} & : n \in A_i \text{ \& } W_{mn} \neq 1 \\ \prod_{\substack{j \in Q^+(h) \\ k \in A_i \\ \text{s.t. } (j,k) \neq (m,n)}} (1 - W_{jk}) \prod_{j \in Q^-(h)} R_{ij} & (17) \\ & : n \in A_i \text{ \& } W_{mn} = 1 \\ 0 & : \text{otherwise} \end{cases}$$

### 3.2.2 $m \in Q^-(h)$ の場合

$m \in Q^-(h)$  の場合と同様にしてつぎの計算式が導出できる。

$$\frac{\partial r_i(h)}{\partial W_{mn}} = \begin{cases} - \frac{1 - R_{im}}{R_{im}} \frac{1 - r_i(h)}{1 - W_{mn}} & (18) \\ & : n \in A_i \text{ \& } W_{mn} \neq 1 \text{ \& } R_{im} \neq 0 \\ - (1 - r_i(h)) \prod_{\substack{k \in A_i \\ \text{s.t. } k \neq n}} (1 - W_{mk}) & \\ & : n \in A_i \text{ \& } W_{mn} = 1 \\ - \prod_{j \in Q^+(h)} S_{ij} \prod_{j \in Q^-(h)} R_{ij} & \\ & \text{s.t. } j \neq m \\ & : n \in A_i \text{ \& } R_{im} = 0 \\ 0 & : \text{otherwise} \end{cases}$$

### 3.2.3 $m \notin Q^+(h)$ または $m \notin Q^-(h)$ の場合

この場合結果は単純であり、計算の必要はない。

$$\frac{\partial r_i(h)}{\partial W_{mn}} = 0 \quad (19)$$

## 4. おわりに

登録文書内の統計情報から計算されたキーワードコネクションマトリックスの初期値を利用者の持つキーワード間の関連度に対する感覚と一致させるためには学習機能が必要である。本論文で述べた方式により学習が実現できる。

### 【参考文献】

- [1] 森田哲也他, "ファジィ文書検索システム(1)~実験システムおよび評価~, 情報処理学会第39回全国大会予稿集
- [2] Y. Ogawa et al, "A fuzzy document retrieval system and learning method based on the dynamic connection method", Pro. of Int. Work. on Fuzzy Sys. Appl., (1988)
- [3] Yasushi Ogawa et al, "A fuzzy document retrieval method and its learning method", Pro. of 3rd Int. Fuzzy Sys. Assoc. Congress, (1989)