

ファジィ理論を用いたエコー情報の分類・分析システム(その2)

5S-2

中村 正規 内山 恵三

東京電力(株) システム研究所

1. はじめに

近年、技術情報や特許情報など文書情報をデータベース化する企業が増えてきている。企業が文書情報をDB化する第1の目的は、膨大な文書を効率的に管理できるようにすることで、そのために必要文書を簡単に取り出すためのいろいろな検索手法が実用化されている。第2の目的は、数値情報を中心とした業務システムでは当たり前となっている分析的活用が文書情報の世界でも扱えるようにすることである。そこで本論文は、現業窓口、電話、葉書などで寄せられたお客様のご意見・ご要望を題材にし、文書の分析方式について述べる。

2. 分析の狙い

お客様からのご意見は、年度別に、契約、広報活動、サービス活動、営業開発など12の視点に分類している。これらの分類に対して、次の3方向の分析を考える。

12の視点

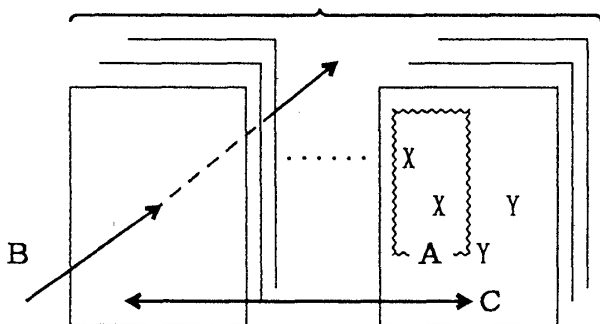


図1 分析の狙い

(1) 潜在課題やお客様意識の抽出

(図中のA)

(2) 潜在課題の傾向把握や改善策の効果分析

(図中のB)

(3) 潜在課題が多視点に与える影響の分析

(図中のC)

3. キーワードによる文書概念の定量化

(1) キーワード対による文書間の定量化

1つの文献に同時に発生する二つのキーワード間(キーワード対)には、統計的な近似関係があることに着目し、12分野(視点)別にキーワード対の依存度を求める。[1][2]

(2) 文書中キーワードの重要度の定量化

我々は、文書の内容に即した重要なキーワードが、文書中の主語と目的語の文節に多く含まれていることに着目し、これら重要文節からキーワード候補を取り出し、さらに、そのキーワードについている格助詞などの情報からキーワード候補を絞り込む方式を開発した。[3]

この方式を用いると、その活用方法(主語、目的語)や文書中での重要度などがキーワードと一緒に取り出すことができる。この重要度という指標はキーワードと文書内容の関連度合いを示すもので、関連性が高いほど1に、関連性が低いほど0に近づく。重要キーワード抽出処理を既存の全文書(Tn)に対して行い、キーワード(KWn)別にまとめると

$$KW1 = \{0.8/T1 \ 0.3/T2 \ 0.6/T3 \ \dots \ d1/Tn\}$$

$$KW2 = \{0.5/T1 \ 1.0/T2 \ 0.2/T3 \ \dots \ d2/Tn\}$$

: : : : :

$$KWn = \{0.5/T1 \ 0.1/T2 \ 0.8/T3 \ \dots \ d3/Tn\}$$

n個のファジィ集合が得られる。

A classification and analysis system for customer's opinion by fuzzy process (part2)

Masaki Nakamura, Keizo Uchiyama

Tokyo Electric Power Company

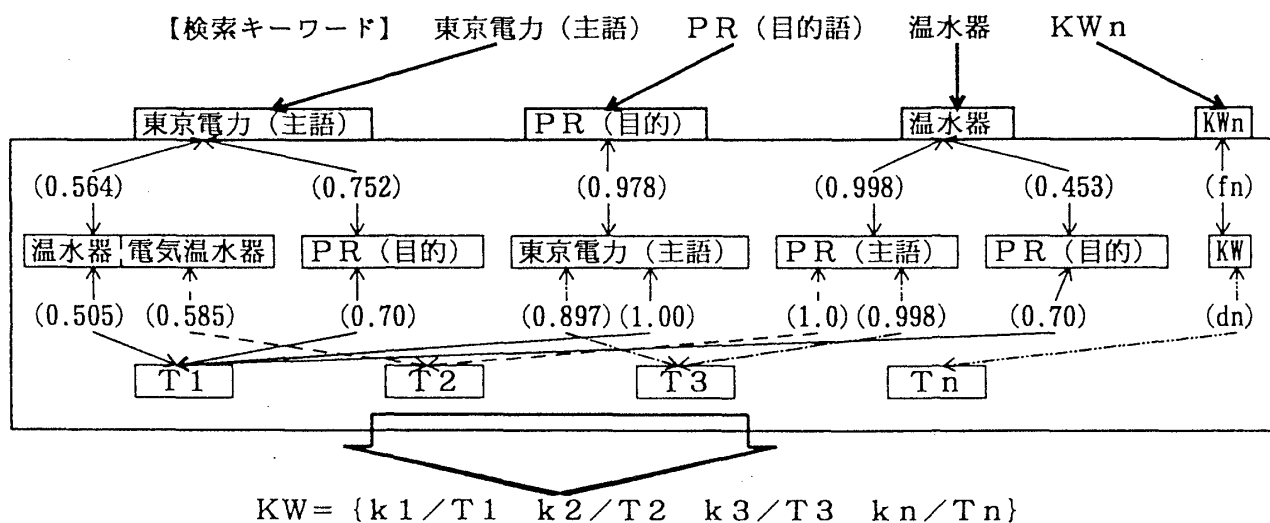


図2 分析用ファジィモデル

4. 分析用ファジィモデル

図2は分析用ファジィモデルの概念図である。温水器の同義語である電気温水器との関係についてこのファジィモデルに検索キーワードを入力すると、キーワードと文書のファジィ関係が得られる。キーワードKWnとKWの概念上の強さはR1 (KWn, KW) = fn、文書nとキーワードKWn (KW) の概念上の強さはR2 (KW, 文書 n) = dn である。この2関係をmax-min演算によって合成した結果がキーワードと文書のファジィ関係である。たとえばキーワードPR (目的) について計算すると、

$$\begin{aligned}
 &PR (目的) \\
 &= \text{MAX} [\text{min} (R1 (東京電力(主語), PR(目的)), \\
 &\quad R2 (PR(目的), 文書1)), \\
 &\quad \text{min} (R1 (温水器, PR(目的)), \\
 &\quad R2 (PR(目的), 文書1))] \\
 &= \text{MAX} [\text{min} (0.752, 0.70), \\
 &\quad (0.453, 0.70)] \\
 &= \text{MAX} [0.70, 0.453] \\
 &= 0.70 \\
 &\text{となる。}
 \end{aligned}$$

他のキーワードについても同様に計算され、
 東京電力(主語) = {0.978/T1 0/T2 0.897/T3}
 PR(目的) = {0.7/T1 0/T2 0/T3}
 温水器 = {0.505/T1 0/T2 0/T3}

$$\text{電気温水器} = \{0/T1 \quad 0.585/T2 \quad 0/T3\}$$

5. 判定

出力結果に対して、キーワード集合間の和集合 (KWa+KWb-KWa×KWb) [4]を計算すると依存度付きの検索結果、T1=0.997, T2=0.585, T3=0.897が得られ、T1が検索条件から推測される概念を一番表している可能性があるかと判断できる。このファジィモデルは視点別年度別に生成することができるので、前述した3方向の分析が自由にできる。

6. 今後の課題

今後は、検索条件が意図する情報をどのくらいの精度で検索できるか実データを用いて評価する計画である。

7. 参考文献

[1]寺野他, 「ファジィシステム入門」, オーム社
 [2]内山, 中村, 「データベースにおける文書の自動分類方式」, 13-117, 平成5年
 [3]内山, 中村, 「重要キーワード抽出方式とその活用法」, 情報処理学会データベースシステム研究会, 1991.7.18, p11
 [4]向殿, 本多, 「ファジィ「あいまい」の科学」, 岩波書店