

保全日報情報を利用した工場トラブル管理支援システムの構築

榊井 文人† 田中 伸一† 河合 敦夫† 井須 尚紀†

† 三重大学大学院工学研究科

{masui, sin, kawai, isu}@ai.info.mie-u.ac.jp

1 はじめに

設備管理は、生産設備を効果的かつ効率的に運用するための手法を追求する分野である。その中で、生産のための材料や労働力、稼働設備の QCDS^{*1}の管理を保全活動または保全管理という [1]。保全管理の方式は、従来 BM(状態基準保全)^{*2}が主流であったが、1950 年代に米国で PM(予防保全)^{*3}が導入されたことがきっかけとなって計画保全の重要性が認識され、その後 TBM^{*4}(定期保全)、CBM^{*5}(予知保全)へと発展してきた。その結果、今日では設備安定稼働率が向上し、トラブル発生に伴う設備停止頻度や事後保全コストは低減した。

一方、設備保全に関する知識やノウハウを習得するためには、日常業務の中で実践経験を積む必要がある。しかしながら、前述した設備安定稼働率の向上は保全管理担当者のトラブル遭遇機会を減少させ、若手や中堅の保全担当者の経験不足や対応能力の低下の要因となっている。さらにこの問題は、昨今の経済状況の影響による経験豊富なベテラン社員減少の事態と相まって、生産現場における突発的トラブルや発生頻度の低いトラブルに対して、保守に要する時間的、人的コストが増大してしまう問題が生じている。コスト低減のために高度なデータベースシステムを導入して保全管理業務日報を有効利用しようとする取り組みも進んでいるが、実際には効果が上がっておらず、問題は解消されていない。

ではなぜデータベースは有効活用されないのだろうか。その主要因として以下の二点が考えられる。

第一に、データベースシステムを駆使するために操作技能が求められる点である。確かに高度なデータベースシステムは、複雑かつ柔軟なデータベース操作が可能である。しかし、システムを適切に利用

するためには、登録しようとするデータの十分な検討とシステム操作に関する豊富な知識と経験が不可欠である。この問題を防止するためには、データベースシステムの操作経験に関係なく手軽に過去の知識を参照できる仕組みの提供が必要となる。

第二に、日報情報の内容が既存のデータベースシステムで上手く扱いきれていない点である。保全日報には、発生したトラブル、発生原因、応急処置、防止対策などの情報が主に自然言語によって記述されており、保全計画や人材育成にとって貴重な知識源となる。しかし、データベースシステムを用いて言語情報の曖昧性を吸収することは案外難しい。例えば「ボード 400 中央部クローズ停止」と「ボード 400 中央部にクローズして停止」は同じトラブルを表すが、データベースシステムの単純な操作ではこれらは異なるデータとして扱われてしまう。日報データを有効な知識として効率良く参照するためには、この問題に対応する必要がある。

我々は、経験の浅い担当者の知識・ノウハウ獲得や保全担当部門の保全計画に対する計算機支援を目指している。本論文では、特に上記二点に対応することで保全管理日報データベースの操作を支援するトラブル管理支援システムについて報告する。以下、2 章で保全管理日報データベースの概要を、3 章でトラブル管理支援システムの概要を説明する。4 章では幾つかの評価および結果の考察を行う。

2 保全管理日報データベース

本章では、本研究で対象とする保全管理日報データベースについて述べる。対象データベースは製品組立工場における 500 以上の対象設備に関する保全管理実績情報が過去 8 年間に渡り蓄積されており、データ総数は 2008 年の時点で約 60,000 件に上る。日報データベースへの登録であるが、操作に慣れた少数の担当者以外は、通常は担当のオペレータを介して行われる。データベースには、トラブルの発生日時や対応時間、発生場所とそれぞれの管理コード、担当者氏名などに加え、自然言語で記述さ

*1 quality・cost・delivery・safety

*2 BM: Breakdown Maintenance

*3 PM: Preventive Maintenance

*4 TBM: Time Based Maintenance

*5 CBM: Condition Based Maintenance

れたトラブル，原因，処置，対策などが登録されている．図 1 に保安全管理日報データベースの自然言語データ例を示す．

トラブル		原因	
...	ボード 100 が中央でクローズして停止	CRI 不良	...
...	ボード 400 中央部でクローズ停止	CRI 不良	...
...	ボード 400 中央部クローズ停止	CRI 不調	...
...	ボード 400 中央部にてクローズ停止	CRI 取付不備	...
...

図 1 保安全管理日報データベースのデータ例

図からもわかるように，記述スペースの制約のために日報情報の記述は新聞などと比べて非常に簡潔である点が特徴である．そのため，省略が多用されたり文法を逸脱するなど，ルーズな記述となっている．さらに，「チョコ停」「取付 R/B」「MS リレー」「SW」など，分野依存，現場依存の専門用語も頻出する．したがって，対象とするトラブル管理日報の自然言語データには，既存の形態素解析や構文解析を用いても十分な性能が得られない可能性が高い．専用の知識辞書を用いて重要概念を判定する手法も提案されている [2] が，人手作業のコストが問題となる．また，本研究から得られる成果は知識財産権化や技術移転が前提となるため，既存の自然言語ツールは導入しにくい面もある．

3 トラブル管理支援システム

本章では，トラブル管理支援システム *minimamo* の概要について説明する．*minimamo* は (1) 索引語抽出と重み付け処理，(2) ベクトル選択処理，(3) 類似度計算と検索処理，(4) フィードバック処理から成る．以下，各処理について説明する．

・索引語抽出と重み付け処理

統計的アプローチによる索引語抽出方法としては *n*-gram を利用した手法が [3, 4] 知られているが，本論文では接尾辞配列と LCPs に基づく Yamamoto らの手法 [5] を応用する．

(1) 対象となる自然言語データについて，接尾辞配列を構築する．ヒューリスティックスとして，数字の連続は同一単語とみなすステミングを行う．

(2) 構築した接尾辞集合を先頭から辞書順に整列する．前後の接尾辞を比較して LCPs を抽出する．

(3) LCPs を抽出する際に頻度情報も記録しておく．頻度があらかじめ設定した閾値 α に満たない場合は索引語候補から削除する．以上の手順を接尾辞全体に適用することにより頻度 α 以上の LCP の集合が候補リスト C_{lcp} となる．

(4) 接頭辞配列を利用して，(1)～(3) と同様の手順で候補リスト C_{lcs} を作成する．

(5) C_{lcp} と C_{lcs} を比較し，重複する候補を索引語として抽出する．

(6) 索引語の重み付けを行う．索引語 w_i の重み S_{w_j} は， w_i の出現頻度 F_j と w_j と同一文中に共起する他の索引語の総異なり数 D_i を用いて式 (1) から得られる．

$$S_{w_j} = \log\left(\frac{F_j}{D_j} + 1\right) \quad (1)$$

・索引語ベクトル選択処理

日報情報データを得られた索引語を要素とするベクトルで表現する．ただし，対象データを構成する索引語の組み合わせを決定する必要がある．この問題は索引語並びの最尤推定探索問題と考えられるので，ビタビアルゴリズムを用いて決定する．

・類似度計算と検索処理

日報データベース検索処理の初期画面はクエリと対象フィールドを指定するだけの極めてシンプルなものである (図 2)．入力クエリを索引語ベクトルに近似し，日報情報データを表す索引語ベクトルとの類似度を計算する．類似度が高いデータは類似データと判断し，因果関係のあるフィールドデータを関連項目として表示する．類似度計算は余弦尺度を用いて行う．

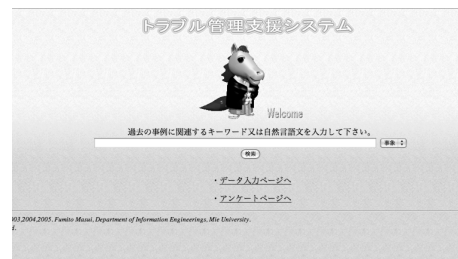


図 2 システムの初期画面

結果は類似度の高いものからランキングして 10 件単位で表示する (図 3)．同時に，対象事項と因果関係のある他の事項も合わせて表示する．例えば，トラブル事象「ステップ 3 でダンマリ停止」をクエリとすると，左フレームに「ステップ 3 にてダンマリ停止」や「ドア卸し部手前にてダンマリ停止」などの過去の類似事例が表示される．同時に右フレー

解析が 0.551 と同程度であった(図 4). 形態素解析の過分割を正解範囲としたことを考慮すればデータに対する実装手法は有効であるといえるだろう.

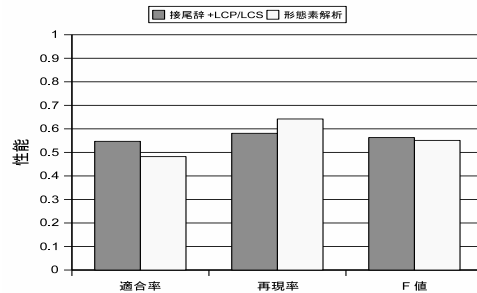


図 4 保全日報データからの索引語抽出性能の比較

次に、検索語の適合フィードバック操作を行った場合の効果を知るための実験を行った。

前述した 130 件のデータを用いて無作為に選んだ 1 件をクエリとして検索を行い、適合フィードバックを行い、索引語集合を再構成した。フィードバック毎に索引語集合の性能を評価して記録した。この操作を 3 回行い、評価値の平均を評価結果とした。

図5は、フィードバック効果を索引語集合の適合率と再現率の推移で示したものである。どちらの数値もフィードバックを20回程度繰り返した時点から性能の向上が鈍りはじめ、適合率は0.7を超えた時点、再現率は0.8を超えた時点で安定した。両数値の変化を比較すると、一貫して再現率の方が高い。この結果は、精度よりも網羅性を重視したいという保全管理グループの要求を満たしている。

4.2 検索性能

システムの検索性能を検証するために擬似的に日報データ検索実験を行った。

まず、日報データベースから任意に1データを取り出し、これをクエリとして日報情報検索を行い、出力されたランキングの最上位データを確認した。適切に検索されていれば最上位にクエリと同一データが得られる。次に、クエリデータの関連項目とクエリデータのランキングを調べた。

例えば、トラブルを示す文字列 T をクエリとした場合、類似度に基づいてトラブルフィールドがランキングされ、同時に関連項目「原因」に文字列 C が表示される。このとき、クエリデータと関連項目は因果関係を持つので、トラブルフィールドの類似度計算が適切であれば、文字列 C の頻度分布はトラ

*6 <http://www.kc.t.u-tokyo.ac.jp/nl-resource/juman.html>

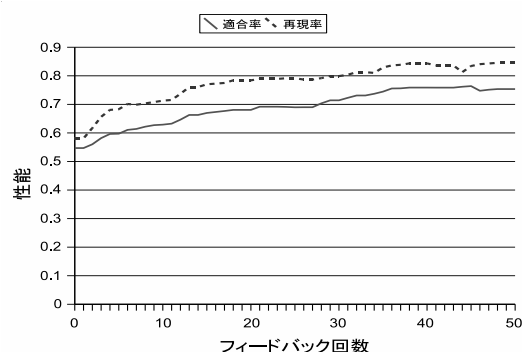


図5 適合フィードバックの効果

ブルフィールドのランキングの上位に集中するはずであり、この分布を分析することでシステムの検索性能を検証できる。関連項目を決める因果関係は表1に示す6通りを用い、検索試行を1,000回行った。

表1 評価対象とした因果関係

関係の種類	例(クエリ... 関連事象)
事象-原因	ボード400 中央部で停止... CR1 不良
事象-処置	ボード400 中央部で停止... CR1 交換
事象-対策	ボード400 中央部で停止... CR 制御点検
原因-事象	CR1 不良... ボード400 中央部で停止
処置-事象	CR1 交換... ボード400 中央部で停止
対策-事象	CR 制御点検... ボード400 中央部で停止

実験結果を図6に示す。クエリデータとの類似度は0.9~1.0に大部分の関連項目が集中しており、本システムの検索処理が有効であったといえる。

その反面、類似度の低い場合にも関連項目が出現している^{*7}。この原因について調べてみると、主に二つの問題によることがわかった。まず一方は、同じ原因によって連鎖的に発生するトラブルが存在する場合である。この問題に対しては、クエリの類似性だけではなく関連項目同士の類似性も考慮することで対応可能であると考えている。

他方の問題は、意味的には同一のものであるが、記述によっては表層上は異なると判断されてしまう場合である。この問題についても関連項目同士の類似性がある程度有効であると思われるが、根本的な対応としては、表層上だけでなくデータの同義性判断が必要であり、今後の課題である。

^{*7} 0.9以下の各類似度間に頻度についてFisherの正確な検定を行った結果、有意水準1%で“0.1~0.2”に他との有意差が認められた。

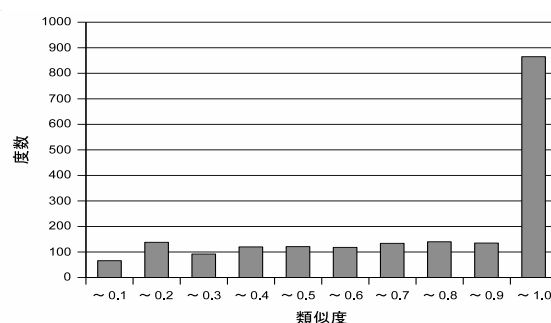


図6 クエリ事象の類似度と因果関係を持つ事象の関係

5 おわりに

本論文では、実社会における自然言語処理技術の応用事例のひとつとして、生産現場の保全管理支援を目的としたトラブル管理支援システムについて報告した。本システムの基本性能については、いくつかの評価実験によってその有効性が確認できた。

現在、より大規模な実験に基づく信頼性の高い評価と、生産現場での試験運用に基づく定性評価を進めている。さらに、本システムを要素技術とした自動演習システムの構築も進行中である。

謝辞

本研究を進めるにあたり、日報管理情報を提供頂いた株式会社データスピリットの中山慎司代表取締役役に感謝致します。

参考文献

- [1] 池田良夫, 萩原正弥, 中村雅章. 設備管理. 日刊工業新聞社, 2004.
- [2] 市村由美, 鈴木優, 酢山明弘, 折原良平, 中山康子. 日報分析システムと分析用知識記述支援ツールの開発. 電子情報通信学会論文誌, Vol. J86-D-II, No. 2, pp. 310-323, 2003.
- [3] 長尾真, 森信介. 大規模日本語テキストのnグラム統計の作り方と語句の自動抽出. 情報処理学会研究報告, 第93巻 of *IPSJ-SIG-NL*, pp. 1-8, 1996.
- [4] 下畑さより, 杉尾俊之. 隣接文字情報を用いたn-gram抽出文字列からの名詞句の自動抽出. 情報処理学会研究報告, 第96巻 of *IPSJ-SIG-NL*, pp. 13-18, 1996.
- [5] Mikio Yamamoto and Kenneth W. Church. Using Suffix Arrays to Compute Term Frequency and Document Frequency for All Substrings in a Corpus. *Computational Linguistics*, Vol. 271, pp. 1-30, 2001.