

ネットワーク管理者MLを情報源とした 運用ノウハウ抽出手法の提案及び実装と評価

宮本 剛[†] 広瀬 崇宏[†] 市川 本浩[†] 中村 豊[‡] 藤川 和利[‡] 砂原 秀樹[‡]

概要: インターネットの普及や高機能な計算機の低価格化を背景として、利用者の手による利用サービスの共同運用が増加しつつある。これらの運用形態では、各利用者が限られた時間の中で自発的に管理作業に携わる反面、その参加は流動的なものになりがちである。参加者の離脱により運用によって培われた運用技術や経験が損失すると、管理作業に支障をきたす恐れが生じる。よって、運用ノウハウを管理者間で持続的に共有できる環境が望まれている。本研究では、運用ノウハウの損失を防ぐために、利用者参加型運用のメーリングリストから運用技術に関する情報を容易に取りまとめられる環境の構築をを目的とする。そこで、MLの中のメールのやりとりをモデル化することで、「一連の話題である」出来事^①を抽出する手法を提案する。

キーワード: 運用ノウハウ, 情報共有, メーリングリスト

The knowledge extraction method from mailing lists of network management

Tsuyoshi Miyamoto[†] Takahiro Hirofuchi[†] Motohiro Ichikawa[†] Yutaka Nakamura[†] Kazutoshi Fujikawa[‡] Hideki Sunahara[‡]

Abstract: Through the spread of broadband network and high performance computers, an increasing number of advanced users cooperate network management by themselves. Though users participate to operation voluntarily, they often leave off irregularly. Therefore, the management knowledge should be shared continuously between operators to avoid the lost of it. In this paper, we aim to help operators to summarize the management techniques from archive of mailing lists easily, in order to keep the shared knowledge in the community. We propose an extraction method of series of emails in the same event from the archives. This scheme is based on the modeling of the email exchanges and the role of operators.

Keywords: network operation, information sharing, mailing list and operational knowledge

1 はじめに

インターネットの普及や関連サービスの低価格化、高機能な計算機の低価格化などから、利用者は単にサービスを利用するだけでなく、自身の手によってこれらのサービスを運用することが可能な環境が整いつつある。そして、個人や企業によるドメイン取得の増加 [7] から、サービスの利用者による自発的な運用(以下、利用者参加型の運用形態)が増加傾向にあることが推測できる。

利用者参加型の運用形態は、利用者による共同運用であることに加えて以下の特徴を持つ。

- ネットワークの管理者は小人数構成であり、そのメンバーの流動性が高い
- 管理者の時間的制約が大きく、運用に十分な時間を割けない

ネットワークや関連するサービスを運用するには、関連する技術などの知識や経験(以下、運用ノウハウ)が必要である。そして、運用ノウハウの損失は、運用に大きな支障をきたすため、運用ノウハウを組織内で維持することが重要な課題となる。そしてそのためには、運用ノウハウの継続的な蓄積と共有が必要となる。

以下、論文の構成に関して述べる。本節では、研究の目的およびその機能要件に関して述べる。2節では、研究の既存技術に関して述べ、研究の位置付けを示す。3節では、研究の提案に関して述べる。4章では、提案システムを用いた実験に関して述べる。5節では、実験の結果を考察し、今後の課題に関して述べる。最後に、6節で論文をまとめる。

1.1 知識の共有手段

本項では、運用ノウハウの共有に用いられる手段に関して述べる。現在、組織内における知識の共有手段

[†] 奈良先端科学技術大学院大学 情報科学研究科
Graduate School of Information Science, NAIST

[‡] 奈良先端科学技術大学院大学 情報科学センター
Information Technology Center, NAIST

表 1: 情報共有手段の特徴

	扱う情報の特徴	取りまとめの必要性
IM メール	個人間での 情報共有を 目的としている	蓄積された情報は、 断片的かつ、散逸的で、 そのままでは、運用 ノウハウの伝達が困難
BBS ML	グループ内での 情報共有を 目的としている	

として、Instant Messenger (IM) や電子メールなどの個人間での情報交換と、メーリングリスト (ML) や電子掲示板 (BBS) などのグループによる情報共有が挙げられる。しかし、これらの手段を用いて情報の共有を行った場合、情報共有の際の履歴などの断片的な情報しか蓄積しない。このため、蓄積された情報を後から参照しても、その内容を理解することができない。従って、情報共有の際には、扱った情報を意味のある形で整理し、まとめて（以下、情報の集約）から、蓄積する必要がある。なお、利用者参加型の運用形態では、管理者が運用に割くことができる時間に強い制約があるため、情報を容易に集約できなければならない。

1.2 情報の集約

本項では、情報源に蓄積されている、断片的な情報を集約し 1 つの話題として提示すること（以下、情報のとりまとめ）を検討する。まず、情報源に関して述べた後、情報の取りまとめの手順に関して述べる。

情報伝達手段の特徴に関してまとめたものを表 1 に示す。ここで、運用組織内で共有すべき情報は、グループ内での情報共有手段を利用した際に蓄積された情報の中に含まれる可能性が高いという考えから、ML と BBS のアーカイブを情報源とする。そして、これらを情報源として、蓄積された情報を取りまとめることを目指す。

次に、図 1 で示している情報を取りまとめるために必要な手順に関して述べる。まず、情報源から互いに関連する情報を機械的に抽出する。そして、抽出された情報を”出来事”と呼ぶ。次に、抽出された出来事を人手によってまとめ、Wiki [8] や Web などを利用し管理者に提示する。ここまでの作業が、情報を取りまとめるためにかかる手順である。つまり、情報の取りまとめには、人手による取りまとめ作業を支援する仕組みを構築が必要となる。なお、管理者は、提示された情報を閲覧することで、運用ノウハウを取得することを試みる。

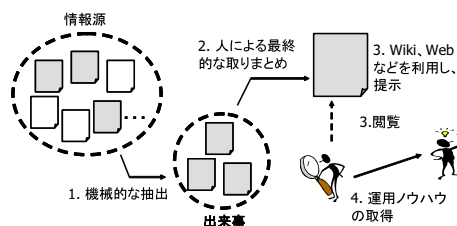


図 1: 情報の取りまとめの手順

1.3 研究の目的

これまで述べた情報源から運用ノウハウを抽出する手法によって、利用参加型の運用組織から運用ノウハウの損失を防ぐことを研究の目的とする。研究目的を実現するためには、情報源から一連の話題を機械的に抽出し、人手による情報の取りまとめ作業を支援する必要がある。そして、これから情報源から一連の話題の抽出に必要な機能の要件は以下の通りである。

- 情報源がそのまま利用できる
情報源として利用する、ML や BBS のアーカイブに蓄積されている情報は、フリーテキストで時系列要素が強いという特徴を持つ。従って、フリーテキストを扱う機能と、持続的に情報を更新する機能が必要である。
- さまざまなタイプの運用ノウハウに対応できる
運用ノウハウには、トラブル対応に関するものや、組織運営に関するものなど、様々なタイプが存在する。従って、運用ノウハウの複数のタイプに柔軟に対応し、抽出する機能が必要である。
- 出来事が提示できる
抽出した出来事を人手によってまとめるため、出来事を管理者に提示する機能が必要である。

2 既存技術

本節では、情報源の蓄積とその利用に関する既存研究として、Expert System [6]、Trouble Tichet System [2]、FISH 関連システム [3] [4]、Instant Case-Based Reasoning System [1] の 4 点を挙げる。そして、これら既存研究と機能要件とを比較する。

2.1 Expert System

事実と規則を組合せ推論を繰り返すことにより結論を得る、Expert System (ES) について述べる。ES は、規則の精度が、得られる結論の精度にそのまま影響する。十分な精度の結果を得るには、システム構築時に十分な規則を準備し規則を適切に更新することが求められる。運用ノウハウのような更新頻度の高い情報を扱う場合、規則の管理作業のオーバーヘッドは大きくなる。利用者参加型の運用形態では、運用に割ける

時間が限られているため、ESを用いて、運用ノウハウの持続的な更新が困難である。また、様々な運用ノウハウにも柔軟に対応できない。

2.2 Trouble Ticket System

各管理者の問題対応状況を把握するためのシステムである Trouble Ticket System (TTS) に関して述べる。TTS では、トラブルシューティングのためのシステムであり、その他のタイプの運用ノウハウの集約及び抽出にはあまり向いていない。

2.3 FISH 関連システム

断片的な情報の管理手法である、FISH システムに関して述べる。FISH システムでは、システムを用いてカードを生成・管理する必要がある。つまり、情報源の情報をそのまま利用できないため、持続的な更新作業への負担が大きい。また、カード間の関係性を維持するためカードに付与されるキーワードは、記述者に依存するため、提示される情報に過不足が生じる可能性がある。

2.4 Instant Case-Based Reasoning System

メールアーカイブを情報源とする Case-Based Reasoning (CBR) System である Instant CBR System に関して述べる。Instant CBR System では、メールアーカイブを情報源とするため、従来の CBR system に比べて、事例の管理が容易である。Instant CBR System は扱う情報が質疑応答形式に特化しており、議論形式やその他の形式の運用ノウハウを扱うことにあまり向いていない。

3 提案

本研究では、ML アーカイブを情報源の例として取り上げる。出来事の抽出に関して、まず、話題の一連のメールの流れを話題の種類ごとにモデル化し、そのモデルに基づき出来事を抽出・提示する。また、モデル化は組織の運用形態によって異なるため、システムの導入時に、モデル化のための作業が必要である。

3.1 出来事のモデル

出来事のモデルの例を、図2に示す。図2では、奈良先端科学技術大学院大学の学生宿者ネットワーク運営委員会(以下、寮ネット)の”トラブルへの対応”モデルを例に挙げている。本モデルの登場人物は、トラブルが発生した利用者、受付を担当する管理者(以下、

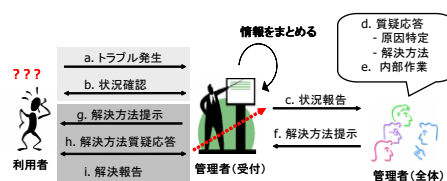


図 2: モデルの例

受付担当), 管理者全体である。次に、出来事対応モデルのメールのやりとりの流れを説明する。トラブル対応モデルでは、利用者が管理者 ML にトラブルの発生を投稿することを起点とする。投稿されたメールに対して、受付担当がその対応に当たる。受付担当が解決できるトラブルであれば、利用者に解決手順を提示する。受付担当が解決できないトラブルである場合、ML を通して管理者全体に相談する。そして、管理者間でトラブルの解決方法を議論する。得られた解決方法は、受付担当を通して利用者に伝達される。

寮ネットでは、今述べたトラブル対応モデルの他に、運用方針、メンテナンス、ミーティング、行時予定に関する出来事のモデル化を行なった。

3.2 情報源に関する予備実験

出来事の抽出に用いる情報源に関して、3つの予備実験を行なった。この予備実験は、情報源から出来事を抽出するため、情報源の特徴を把握することを目的とする。まず1つ目に、出来事が複数のモデルにマッチした際の判別を補助するため、モデル毎の出来事数を分析した。2つ目に、メールの Message-Id などの情報から一連の話題の抽出する作業を補完するため、出来事の起点から終点までの時間である収束時間を分析した。最後に、出来事を構成する各メールの返信時間に関して、各モデル毎に何らかの特徴が得られないかを検証した。

• モデル毎の出来事の数

1つめの予備実験である、モデル毎の出来事数に関して述べる。分析対象は、寮ネットの ML に含まれる 1040 件のメールとした。そして、これらのメールを先に述べた 5 つのモデルに手動で分類した。その結果、表 2 に示す通り、合計 127 個の出来事が得られた。また、出来事の約 6 割が”トラブル対応”モデルであることや、”メンテナンス”モデルの出来事は、出来事を構成するメールの数が少ないことなどの特徴も得られた。

• 出来事の収束時間

出来事の収束時間の分析した結果を図 3 に示す。分析対象は、1つ目の予備実験で手動分類した出来事である。分析の結果、約 9 割以上の出来事が 10 日以内に収束することが分かった。よって、この値を出来事を構成するメールの判別に利用する。

表 2: モデルと出来事

	出来事 の数	出来事 の種類	メール の数
トラブル対応	73	39	483
運用方針	19	14	128
メンテナンス	21	15	84
行事計画	3	2	99
ミーティング	11	5	175
広告/文字化け	-	-	21
その他	-	-	50
合計	127	75	1040

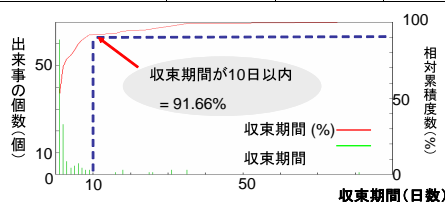


図 3: 出来事の収束機関

- 出来事の平均返信間隔
出来事を構成する各メールの返信時間に関して、表 2 で挙げたモデル間に何らかの特徴を確認することはできなかった。しかし、図 4 に示す通り、各出来事を構成する各メールの平均返信時間が 1 日以内であるということが分かった。よって、この値を出来事を構成するメールの判別に利用する。

以上の予備実験の結果を踏まえて、出来事を抽出するためのアルゴリズムを、以下に述べる。

1. スレッドの生成
ML のアーカイブからメールを取得し、各メールの Message-Id, In-Reply-To, References [5] から出来事の元となるスレッドを生成する。
2. スレッドの補正
予備実験で得られた出来事の収束時間、出来事内の平均返信間隔、単語の出現傾向などから、スレッドの結合、分離などの補正を行なう。
3. 適合するモデルの判別
単語の出現傾向や、メールの送受信者の役割を分析し、適合する出来事のモデルを決定する。

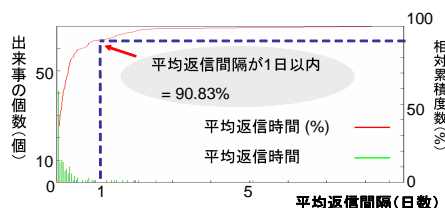


図 4: 出来事の平均返信時間

4. 出来事の提示
最後に、適合した出来事のモデルに基づいて抽出した出来事を提示する。

4 実験

出来事の提示例として、トラブル対応モデルの出来事を示す。まず、図 5 で示す通り、対象となるメール集合を (計 14 件)、スレッド表示技術を用いて表示した。図 5 から、対象となるメール集合の情りの流れを以下に示す。なお、各メールの ML の通し番号 (" [dormnet-staff:00820]help..." であれば "00820" の部分) をメールの ID として扱う。

1. トラブル発生 (ID:00820)
利用者が管理者にトラブルの発生を報告する。
2. 状況確認 (ID:00821, 00822)
利用者と管理者間でトラブル発生の際の状況確認を行う。
3. 議論 1 (ID:00823, 00827, 00830)
管理者間における、トラブルの原因に関する議論。この議論では、原因解明には至らない。
4. 議論 2 (ID:00824)
管理者間における、トラブルの原因に関する 2 つ目の議論。この議論は、原因解明の糸口となる。
5. 原因判明 (ID:00825, 00828, 00829)
トラブルの原因が判明する。
6. 改善作業 (ID:00831, 00838)
判明したトラブルの原因から、原因を改善する内容に関して話し合う。
7. 解決報告 (ID:00831, 00838)
トラブルが解決し、利用者にもその旨を報告する。
8. 御礼 (ID:00831, 00838)
利用者からトラブル解決に関する御礼のメールを受け取る。

以上の情報のやりとりの中で、運用ノウハウを生成するために必要不可欠なメールは、トラブル発生、議論 2, 原因判明, 改善作業に関するメールである。また、今挙げた内容を補足するメールとして、状況確認に関するメールが挙げられる。

次に、図 5 で挙げた 14 件のメールに関して、提案システムを用いて提示される出来事を図 6 に示す。図 6 において、(claim) 若しくは、(reason) の印がついているメールが出来事として抽出されるメールである。(claim) の印がついているメールは、単語の出現度合から、利用者からのトラブル発生報告の内容が含まれることを表す。また、(reason) の印がついているメールは、トラブルの原因に関する内容が含まれることを表す。

次に、提示される出来事と運用ノウハウを生成に関

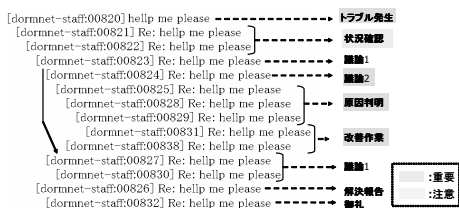


図 5: スレッドの例

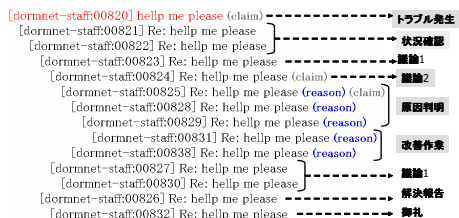


図 6: 出来事の場合

して述べる。図 6 で示した出来事は、運用ノウハウの内容を補足するメールを含んでいないものの、運用ノウハウの生成に必要なメールをすべて含んでいる。よって、この出来事をまとめ挙げることで、運用ノウハウの共有が十分可能であるといえる。

5 考察と今後の課題

本節では、“提案システムを他の情報源に適用するの有効性”、“情報を出来事として提示することの有効性”、“有効な出来事の提示のために必要な事項”の3項目に関して考察し、それぞれの課題に関して述べる。

5.1 他の情報源に適用する際の有効性

本提案手法を他の情報源に適用する際の有効性に関して考察する。提案手法は、導入時に出来事のモデル化を必要とするため、情報源に含まれる情報の把握していることが前提となる。利用者参加型の運用に限れば生成されるモデルに一定の傾向がある可能性が高いと考えられる。また、メール本文の内容だけでは、機械的な出来事のモデルの把握が困難であり、各管理者の役割をシステムが判断できることが重要である。

5.2 情報を出来事として提示することの有向性

情報を出来事として提示することの有向性について考察する。提案システムでは、出来事の有向性の判断をシステムの利用者に委ねるという方針から、最終的な情報の取りまとめを利用者に任せている。しかし、出来事がある程度数になってきた場合には、重要性を判断して提示する必要がある。

5.3 有効な出来事の提示のために

有効な出来事の提示のために考えるべき事柄について述べる。まず、メール本文の語数が少ない場合、提案システムが話題の転換に追従できない場合がある。次に、メールとモデルとの関係に関して、現実のメールのやりとりは多岐にわたるため、モデルを詳細にし過ぎると、機械的な判断ができない場合がある。よって、大局的な状態把握を促すような工夫が必要となる。最後に、機械による自動処理では、出来事の把握にはどうしても限界がある。よって、システム利用者からのフィードバックが動的に反映させることを考えるべきである。

6 まとめ

利用者参加型のネットワーク運用においては、運用ノウハウの継続的な蓄積が求められている。よって、本研究では、情報源から“出来事”を抽出し、運用ノウハウを取りまとめる作業を支援することを研究目的とした。研究目的を達成するには、情報源から一連の話題を機械的に抽出することが必要であり、そのための機能要件は、情報源に蓄積されている情報がそのまま利用可能なこと、様々なタイプの運用のノウハウに対応可能なこと、出来事が提示が可能なことの3項目である。

本研究では、奈良先端科学技術大学院大学の学生宿舎ネットワーク運営委員会のMLアーカイブを情報源とし、蓄積されている情報をモデル化し、それに基づいて出来事を抽出した。情報を出来事として提示することは、単にスレッドを表示するよりも、有用な情報が提示できる可能性を確認できた。また、出来事の抽出手法の改善や更なる有効性の検討が必要である。

参考文献

- [1] Dai Kusui and Hido Shimazu: Transforming Electronic Mail Folders into Case Base, *ECCBR2001, INAI2080, (Poster Presentations) pp690-701* (2002).
- [2] 泉裕, 斎藤彰一, 上原哲太郎, 國枝義敏: “柔軟なネットワーク管理フレームワークを提供するトラブルチケットシステムの構築”, 電子情報通信学会論文誌, B, vol.8 j86-B No.8, pp.1502-1514 (Aug. 2003).
- [3] 関良明, 山上俊彦, 清水宏宏: “ノウハウ蓄積システム FISH の実現とその評価”, 電子情報通信学会論文誌, D2, Vol.J76-D2 No.6, pp.1223-1231 (Jun. 1993).
- [4] 関良明, 山上俊彦, 清水宏宏: “情報連携モジュール Fly-fishing の提案と性能評”, 電子情報通信学

会論文誌, D1, Vol.J82-d1 No.9, pp.1202-1209
(Sep.1999).

- [5] IETF: RFC2822, <http://www.ietf.org/rfc/rfc2822.txt?number=2822>
- [6] ES/KERNEL2/H:ExpertSystem/Kernel2/HostSystem:
<http://www.hitachi.co.jp/Prod/comp/soft1/V0S3FS/v3ga0114.htm>
- [7] 総務省統計局: IT 関連資料集 平成 15 年度
版 表 59. <http://www.stat.go.jp/data/it/zuhyou/i59.xls>
- [8] Wiki: <http://www.wiki.org/>