

# メールアドレスを活用した社内情報の推薦システム

谷口 友紀                      上田 良寛                      谷口 元樹  
三沢 翔太郎                  三浦 康秀                      大熊 智子

富士ゼロックス株式会社

{Taniguchi.Tomoki, Ueda.Yoshihiro, motoki.taniguchi, misawa.shotaro,  
Yasuhide.Miura, Ohkuma.Tomoko}@fujixerox.co.jp

## 1 はじめに

個々のワーカースが有する業務知識の利活用に対する需要が高まっている。従来より業務に必要な知識はマニュアルやFAQで文書化され、会社や部門などの組織で共有されてきた。しかし、近年では業務内容が多様化しており、文書化で対応しきれない事例が増えてきている。

業務における重要なコミュニケーション手段のひとつにメールがある。文書化でカバーしきれない知識はワーカース間のコミュニケーションによって伝えられるため、メールアドレスにはワーカース個別のノウハウが蓄積されている。本研究では、メールアドレスの利活用に焦点をあて、メールアドレスを利用者間で共有することにより、ワーカースが持つ個別の業務知識を活用するシステムを提案する。そして構築したシステムを展開し、被験者によるアンケートを実施することでメールアドレス共有の有効性を検証する。

本稿では事務手続き業務を対象に選定した。事務手続きとは「旅費清算」などの諸手続きや「委託契約」「共同研究」などの契約手続きを指す。事務手続きは繰り返し行われる業務であるため、過去の知識が再利用できる一方で、部門や組織ごとにローカルルールが存在し、文書により明文化されないことが多い。個人にとってはじめての取り組みであったり、過去に実施してから時間が経過しているために詳細を記憶していないこともある。そのため、経験者のノウハウを求めて、直近に手続きを実施した社員に依頼し、関係するメールアドレスを提供してもらうことが頻繁に行われている。提案システムによりこれらの情報共有作業を自動化する。

## 2 関連研究

コミュニティー型質問応答サイト(CQA)を対象として、個々の利用者が持つ知識を活用する研究がある。CQAに投稿された質問や回答の質は様々であるため、これらの研究ではCQAからの知識抽出に主眼を置いている。[1]はソーシャルネットワーキングサービス

に投稿された質問に回答するために、CQAから知識を抽出する手法を提案している。また、[2]は類似質問を探し出すタスクを提案している。しかし、これらの研究はメールアドレスを対象としたものではない。

メールアドレスを活用して、メールによるコミュニケーションを支援する研究が盛んに行われている。[3]はメールの送受信履歴に基づいて、宛先やCCに含めるべき人物を推定する手法を考案している。[4]は添付すべきファイルを受信ボックスから検索して、提示するシステムを提案している。さらに[5]は、本文候補の提示により返信文作成を支援するシステムを構築している。しかしながら、これらの研究では、メールアドレスから抽出した知識の共有を目的としていない。

## 3 メール推薦システム

### 3.1 システムの概要

メールアドレスは機密性の高い情報である。ゆえに所有者本人による共有可否の判断が不可欠である。しかし、判断対象であるメールアドレスは大量に存在するうえ、判断結果は開示する相手に依存するため、事前に共有の可否判断を行うことはできない。そこで、各利用者間における情報共有の要否および共有する情報をシステムが自動で決定し、他者に情報を提供するときのみ開示可否の問い合わせを行う推薦システムを考えた。メール推薦システムの構成を図1に、メールの推薦処理手順を以下に示す。

1. 「メール収集」では、メールサーバーから定期的にメールアドレスを収集してシステムのデータベースに登録する。
2. 「推薦要否判断」では、新着のメールアドレスを監視し、その内容に基づいてメール受信者への推薦の要否を決定する。推薦のトリガーとなるメールアドレスをトリガーメールと呼称する。
3. 「検索」では、トリガーメールからクエリを生成する。クエリを利用して過去のメールアドレスを検索し、推薦するメールアドレスを抽出する。
4. 「許諾問い合わせ」では、検索で得たメールアドレスの所有者に開示可否の判断を依頼する。

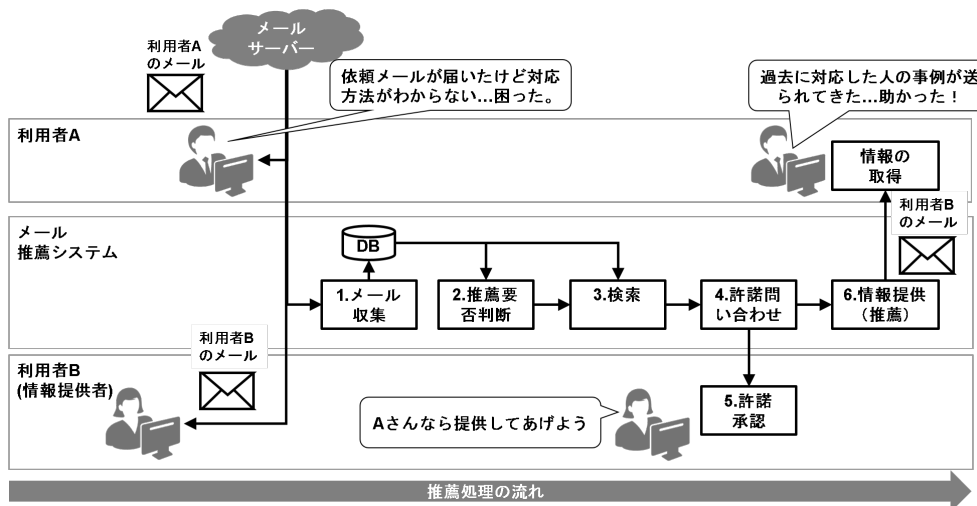


図 1: メール推薦システムの構成

5. 「許諾承認」では、メールアドレスの所有者が開示可否を判断する。
6. 「情報提供」では、開示可と判断されたメールアドレスをトリガメールの受信者に推薦し、情報共有が完了する。推薦結果はポップアップで通知される。実際に推薦されたメールを推薦メールと呼称する。

次節では「推薦要否判断」「検索」「許諾承認」処理を取り上げ、詳細を説明する。

### 3.2 推薦要否判断

推薦の要否は「依頼」「質問」「要望」を意味する検索語を件名に含むか否かで判断される。検索語を含むメールの受信者は業務の当事者として何らかの対応を求められていると推測できるため、メールアドレスの推薦タイミングとして適切であると考えた。ただし、単なる日程調整など明らかに推薦のトリガーとして不適なもの除去するためにストップワードを使用する。

推薦要否判断の精度はシステムの性能に大きく影響する。そこで、検索語による要否判断に加え、Conditional Random Field (CRF) を使った固有名抽出技術による要否判断を導入する。事務手続きに関するメールアドレスには固有の手続き名や文書名が頻繁に出現する。これらの表現が件名または添付ファイル名に含まれることを確認して、トリガーが発火する対象を事務手続きに絞る。CRF モデルを構築するために、実際の業務でやり取りされたメールアドレス 1715 件から独自のコーパスを作成している。表 1 に作成したコーパスに含まれる表現の例を示す。事前実験から件名における表現抽出の F 値は手続き名で 0.735、文書名で 0.808 であった。添付ファイル名では手続き名で 0.270、文書名で 0.446 であった。

表 1: 事務手続きに関する表現の例

手続き名	文書名
社外発表許可申請	研究実施計画書
海外出張申請	情報資産社外持ち出し申請書
共同実験契約	機密保持・著作権契約書

### 3.3 検索

検索には、オープンソースソフトウェアの Apache Solr<sup>1</sup> を利用する。Apache Solr では文書に検索対象となるフィールドを複数設置することが可能である。文書の検索スコアはフィールドごとの検索スコアの重み付き和であらわされ、フィールドごとの検索スコアは検索キーワードの TFIDF をもとに計算される。

提案システムでは、メールアドレスを文書として、インデックス作成と検索を実行する。メールサーバーから定期的に未取得のメールアドレスを取得し、検索に利用するフィールド（件名、本文、添付ファイル名）を Apache Solr に登録する。検索には CRF で抽出した「手続き名」「文書名」および「件名に含まれる語」をクエリとして利用する。件名と本文フィールドを検索対象とし、Apache Solr で検索した結果中で最もスコアの高いメールアドレスを推薦候補として選択する。なお、利用者が過去に実施した手続きであっても、やり方を忘れていている可能性がある。そのため、他者と利用者本人の送受信ボックスを検索対象に加える。

### 3.4 許諾承認

許諾承認では、メールアドレスの所有者に開示可否の判断を行うための情報が提供される。提示されるのは「開示する相手名」「送受信日時」「件名」「本文」「添

<sup>1</sup><http://lucene.apache.org/solr/>

表 2: 被験者によるアンケート結果

(a) 拒絶理由				(b) 有用性と関連性			
分類	開示可否	説明	件数 (%)	分類	有用性	関連性	件数 (%)
開示可	可	提供できる	125 (77.2)	○	役に立つ	関連あり	25 (22.5)
機密 不正確 理由なし	不可	機密を含む 誤りを含む 理由なし	2 ( 1.2) 9 ( 5.6) 1 ( 0.6)	△-関係あり	どちらともいえない	関連あり	27 (24.3)
回答なし	不明	未承諾	25 (15.4)	×-既知	既知で役に立たない	関連あり	21 (18.9)
				×-期待はずれ	役に立たない	関連あり	4 ( 3.6)
				×-関連なし	役に立たない	関連なし	24 (21.6)
				トリガーが不適切	推薦要否の判断が適切でない		10 ( 9.0)

付ファイル名」である。メールアドレスの所有者はこれらの情報を参考に開示の可否を判断する。

## 4 実証検証

### 4.1 参加者と検証期間

開発したメール推薦システムを研究開発部門に展開して実証検証を行った。検証に参加した被験者は11名である。2017年3月から2017年9月までの約7ヶ月間の検証を行った。推薦対象となるメールアドレスは108535件であった。運用期間で推薦されたメールアドレスは162件であった。

### 4.2 被験者によるアンケート

開発したシステムを評価するために、被験者に対して以下の2種類のアンケートを実施した。

- メールデータを開示しなかった拒絶理由
- 推薦結果の有用性および関連性

前者のアンケートで機密度の高いメールアドレスを共有することに対する障壁を調査する。そして、後者のアンケートで構築したシステムの有効性を検証する。被験者は回答候補から選択する形でアンケートに答える。回答候補に該当しない場合や補足情報を入力するために自由記述欄を設けた。自由記述欄の内容を加味してアンケート結果を分類したものを表2に示す。次節以降では被験者から集めたアンケート結果を分析する。

### 4.3 拒絶理由

表2aから開示可が125件(77.2%)であり、大部分のメールアドレスは他者に公開可能であることが明らかになった。加えて開示不可の中で機密情報を含むものは2件(1.2%)しか存在しないことが確認できた。また、情報が古い、情報が不足しているなど不正確さを理由に挙げたものが9件(5.6%)であった。ただし、検証実験の被験者の所属部門は同じである。面識のない利用者間における共有については再考を要する。

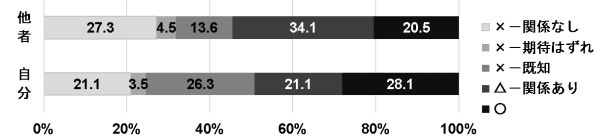


図 2: 所有者による層別結果

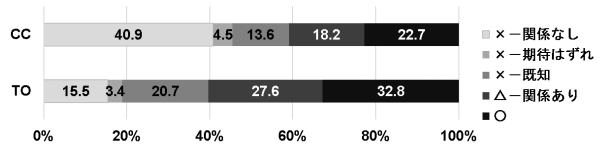


図 3: 受け手の立場（宛先）による層別結果

### 4.4 有用性および関連性

有用性は推薦結果が業務活動に役立つか否かを、関連性はトリガメールと推薦メール間の関連の有無を表している。開示可となった125件のうちで、未評価を除いた111件に対する有用性と関連性を評価した結果を表2bに示す。表2bから全推薦のうち22.5%が「○」であり、業務の役に立っていることがわかった。関連があるものは「○」「△-関係あり」「×-既知」「×-期待外れ」を合計した77件あり、全体のうち69.4%で関連したメールアドレスが推薦されているという結果が得られた。これらの結果は関連するメールアドレスが必ずしも有益でないことを示している。

### 4.5 考察

被験者が有益と判断したメールアドレスの特徴を考察するためにアンケート結果を層別し、以下の調査を実施した。

- 推薦メールの所有者による有用性を調べる
- トリガメールの受け手の立場による有用性を調べる

推薦メールの所有者は自分と他者に、トリガメールの受け手の立場はTOとCCに層別した。TOは自分宛てに送られたものを、CCはカーボンコピーとして受信したものを指す。TOとCCの層別ではメーリングリスト宛てに送信されているケースを目視で除去し

た。表 2b からトリガーが不適切となったものを除去して層別した結果を図 2 および図 3 に示す。

図 2 から、自分の送受信ボックスから推薦された場合において「○」の割合が 28.1% と高いことがわかった。自分の送受信ボックスからの推薦は一度は知った情報の再提示であるが、必ずしも被験者が過去の情報を記憶しているわけではなく、有益となったのではないかと考えられる。この結果は自分の送受信ボックスを検索対象とすることの妥当性を示している。その一方で「×-既知」に 26.3% の評価結果が集中していることから、推薦が余計であった事例を多く含んでいると考えられる。また、他者から有益な情報が提供されるのは 20.5% であった。自分が所有しているメールアドレスには及ばないものの、他者とのメールアドレスの共有が有効であることが確認できた。

続いて TO と CC で比較する。図 3 から TO では役に立つものが 32.8% であり、CC に比べて高くなっている。自分向けに送られたメールは業務の当事者として直接的な対応が求められ、推薦に対する有用性が高まったと考えられる。言い換えれば、対応を行うために有益な参考情報を求めていると考えられる。

#### 4.6 事例

推薦が有益な事例では、過去の類似事例が推薦されていた。事例を図 4 に示す。事例 (a) は海外出張の申請が必要となった状況で、他者が作成した申請書が提供された。事例 (b) では詳細な実施計画を立てる段階で、当初の依頼内容の全容が再確認できる文書が推薦された。

### 5 おわりに

本稿では、事務手続き業務を支援するシステムの事例として、メールアドレスを活用した社内情報の推薦システムを開発した。そして被験者に対してアンケートを実施することによりシステムの有効性を検証した。その結果、多くのメールアドレスが利用者間で共有できることがわかった。また、自分および他者のメールアドレスの推薦は業務に有益となることを示した。さらに被験者が有益と感じたメールアドレスの特徴について考察を行った。

構築したシステムは Apache Solr による検索を用いて、トリガーマールに類似するメールアドレスを抽出している。そのため、トリガーマールと類似しているだけで、解決策が記載されないメールアドレスが推薦される事例がみられた。実証検証で利用したアンケート機能を活用すれば、評価結果のフィードバックデータを利用者から収集することが可能である。今後はこれら

事例(a)	
トリガーマール	
件名: 海外出張申請依頼	添付ファイル: なし
本文: お疲れ様です。論文のAcceptおめでとうございます。早速、海外出張申請書を提出してください。	
推薦メール	
件名: 【承認済】海外出張申請書	添付ファイル: 海外出張申請書.xxx
本文: お預かりしていました海外出張申請につきまして、承認をいただきましたので申請書(一式)を添付いたします。.....	

事例(b)	
トリガーマール	
件名: 件名: 研究依頼の件	添付ファイル: なし
本文: さて研究依頼の件ですが、いつごろ始めるのがよろしいでしょうか?ご都合など教えて...	
推薦メール	
件名: 研究依頼ドラフト	添付ファイル: 研究依頼書ドラフト.xxx
本文: モジュールの改善を依頼内容として、研究依頼書のドラフトを作成いたしました。ご確認いただき、希望納期が妥当かどうか...	

図 4: 有益であった事例

のフィードバック結果を活用して、システムの改善に取り組みたい。

## 6 商標について

Apache Solr は Apache Software Foundation の登録商標または商標です。

### 参考文献

- [1] M. Bagdouri and Douglas D. W. Oard. Building Bridges across Social Platforms: Answering Twitter Questions with Yahoo! Answers. In *Proc. of SIGIR'17*, pp. 1181–1184, 2017.
- [2] P. Nakov, D. Hoogeveen, L. Márquez, A. Moschitti, H. Mubarak, T. Baldwin, and K. Verspoor. SemEval-2017 Task 3: Community Question Answering. In *Proc. of SemEval'17*, pp. 27–48, 2017.
- [3] D. Graus, D. van Dijk, M. Tsagkia, W. Weerkamp, and M. de Rijke. Recipient Recommendation in Enterprises using Communication Graphs and Email Content. In *Proc. of SIGIR'14*, pp. 1079–1082, 2014.
- [4] C. van Gysel, B. Mitra, M. Venanzi, R. Rosemarin, G. Kukla, P. Grudzien, and N. Cancedda. Reply With: Proactive Recommendation of Email Attachments. In *Proc. of CIKM'17*, pp. 327–336, 2017.
- [5] A. Kannan, K. Kurach, S. Ravi, T. Kaufmann, A. Tomkins, B. Miklos, G. Corrado, L. Lukacs, M. Ganea, P. Young, and V. Ramavajjala. Smart Reply: Automated Response Suggestion for Email. In *Proc. of SIGKDD'16*, pp. 955–964, 2016.