

異分野共同研究履歴分析の事例

An Analysis on Collaboration Record of Dissimilar Research Fields

田中克明*1 濱崎雅弘*2
Katsuaki Tanaka Masahiro Hamasaki

*1 一橋大学 情報基盤センター

Center for Information and Communication Technology, Hitotsubashi University

*2 産業技術総合研究所

National Institute of Advanced Industrial Science and Technology (AIST)

In this paper, we proposed two systems to analyze archived e-mails. First system visualized the difference between messages sent by a person and messages mentioned him or her along with time passes. Second one extracted series of discussions for an object from messages. We applied them to analyze mail archives of collaboration project of dissimilar research fields.

1. はじめに

分野の異なる個人が協働する際、まずはそれぞれに期待を持つのだが、実際に活動してみるとその期待に応える成果をその分野ですぐには出せないことがわかり、失望感を味わう[吉野 12] ことがしばしばある。

たとえば、情報システムに情報デザイン分野の手を加えるというプロジェクトでは、表層表現の親しみやすさやシステムの利用のしやすさなど、人に触れる部分のデザインを著しく改善できるのではないかと期待が持たれる。しかし、実際には情報デザインの専門家はインタラクションデザインも専門としているわけではなく、結果、情報システムの設計が迷走する、といったことが起こりうる。だが、迷走しつつも、作成した情報システムを利用する実践活動を繰り返すことを通し、相互にどこまで期待することができるかを把握し、新たな活動を続けることができる。

このように、個人が外部に対していただく期待と個人が外部からいだかれる期待が相互に噛み合い、安定すれば、失望感を克服して次の一步を踏み出すことができる。そこで本稿では、期待すること・されることの相違と変動を抽出することを目的として、異分野が共同した活動履歴において、個人がプロジェクトに対して発信している内容、プロジェクト全体から個人が受け取っている内容の方向性の一致具合、および、それらがプロジェクト進行の中で安定した状態にあるかの分析を行うシステムの構築を行った。

2. 分析対象

分析の対象には、発信者と、言及されている者とが明らかであり、プロジェクトに並行して生成される文書が望ましい。メールであれば、発信者を特定することができ、プロジェクト期間中随時作成され、時刻が付与されていることからプロジェクトの進展にそって流れを追いやすい。

本稿では、科学技術振興機構 CREST の一環として 2006 年 10 月から 2012 年 3 月まで進められた「情報デザインによる市民芸術創出プラットフォームの構築」の後期に行われた、「か

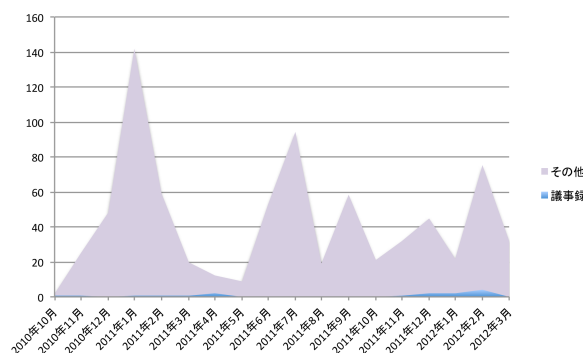


図 1: メール数の推移

えりみちの「アートスペース」プロジェクト [小早川 12]*1において扱われたメール、2010 年 10 月 8 日から 2012 年 3 月 7 日の 768 通を対象とした。768 通には、打ち合わせ議事録が 16 通含まれる。月ごとのメール数の推移を図 1 に示す。

以後の処理にあたり、各メールからは添付ファイルを取り除き、シグネチャ、引用部分などもできる限り取り除いた。

3. 分析手法

3.1 発信・言及内容による分析

今回取り上げたメールは、そこに参加している個人が発信している。ある個人が発信したメールはその個人から周囲に対する言及である。一方、ある個人の名前が書かれているメールはその個人に対する言及である。前者は個人から周囲への期待を、後者は周囲から個人への期待を含んだものであると考えられる。

そこで、個人が発信したメール（発信メール）と個人が言及されたメール（言及メール）の傾向の変化を、一定時間間隔ごとに表示することにより、個人が持つ期待と、個人が持たれている期待の変化を提示することができないかと考え、傾向と変化の提示を行うシステムを構築した。システムの画面例を図 2 に示す。

連絡先: 田中克明, 一橋大学情報基盤センター, 〒 186-8601, (042)580-8440, jsai2013@katsuaki-tanaka.net

*1 <http://kaerimichi.sunagaland.com/>



図 2: 発信・言及分析システム画面例

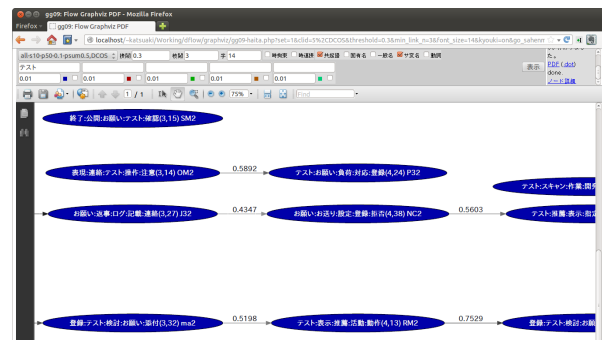


図 3: 議論対象言及分析システム画面例

なお、個人からの発信とその個人への言及ではなく、個人と相手となる個人の間におけるメール内容を分析し、関係の変化を追跡する方法も考えられる。しかし、分析対象となる要素数が参加するメンバー数の組み合わせに対応して増えるため、個人の発信内容・個人への言及内容を分析対象とした。

本システムでは、ある個人に着目し、期間ごとの発信メールと言及メールを集約、それぞれを長さがあるベクトルとして線を描画する。描画にあたっては、発信メールと言及メールの内容が近ければ、2つの線の向きが近く、発信メールと受信メールの表記内容が異なれば、2つの線の向きが遠ざかるように表示されることを狙った。

まず、すべてのメールを pLSI [Hofmann 99] を用いて 3 トピックに縮約した。次に、メールがやり取りされた期間をメールの送信時刻により $N(N \geq 1)$ 分割し、期間 n に属するある個人が発信したメールを d_{fn} 、言及があるメールを d_{mn} とする。この上で、pLSI にて求めた各トピックを $z_k (k = 1, 2, 3)$ とし、期間 n の発信メールを示すベクトル V_{fn} 、言及メールを示すベクトル V_{mn} を

$$V_{fn} = \sum_{k, d_{fn}} p(d_{fn}|z_k), \quad V_{mn} = \sum_{k, d_{mn}} p(d_{mn}|z_k)$$

として求めた。発信メールは From ヘッダのメールアドレスにより選択し、言及メールは名前など個人を示す単語が出現するメールを該当するものとして選択した。

最後に、それぞれのベクトルを方向、すなわち角度に変換する。変換には、RGB 色空間で表された色を HSV 色空間へ変換する際の手法を用いた。色は、RGB 色空間では一定の幅を取る 3 つの値の組として表現され、HSV 色空間では、1 つの角度と 2 つの一定の幅を取る数値として表現される。3次元のベクトルである V_{fn} 、 V_{mn} にこの変換を当てはめ、角度として得られた値を発信メール、言及メールを示す方向とした。なお、HSV 色空間では、角度となる H が色相環上の位置(何色か)を表し、S が彩度、V が明度を表す。

3.2 議論対象に関する言及の分析

様々の分野の人間が議論を行うと、同じ言葉がもつ位置づけが異なっていることがある。すなわち、メーリングリストなどである一定の期間にわたって議論をこなう場合、議論の対象となる言葉の扱われ方を抽出することで、分野を超えた議論がどのように進展しているかを追うことができると考えられる。

そこで、メーリングリストで議論されているトピックの変化を時系列に沿って抽出し、ある議論対象に対する言及の変化を抽出するシステムを構築した。画面例を図 3 に示す。

本システムでは、メール本文に着目し、メールが交わされた期間を時刻により N に分割、各期間までに交わされたメールを pLSI を用いて K トピックに分類する。この際、新たなメールを含まなかったトピックは古いトピックとして扱い重みを減らし、古い内容が徐々に忘れ去られるようにした。次に、着目する言葉 w に対し、 $p(w|z)$ がしきい値以上であるトピックを選択、トピック中で w と共起する語をラベルとしてトピックの遷移を表示したものが図 3 である。

4. 考察

4.1 発信・言及内容による分析

図 2 に示したように、本システムでは、メーリングリストの一定期間ごとのメールについて、発信メールと言及メールの方向を表示する。まず、個別の個人・期間において、発信メールと言及メールについて本システム表示上の方向が近い場合、メール本文を確認すると、内容の近いメールが発信メールと言及メール双方に多く含まれていることを確認できた。たとえば、着目した個人がある要請を送りそれに対して他のメンバーが回答を寄せている場合、着目個人とその他のメンバーとの間で議論が交わされている場合などであり、相互の立場が噛み合っている。ただし、期間を分割したために、前の期間への返答が次の期間で行われていると、発信と言及の向きがかみ合わないように表示されてしまう場合もあった。

一方、発信メールと言及メールの表示が遠い場合、その期間の当該のメールから、両者の内容にあまり相関がないことを確認できた。例えば、ある個人からの、ワークショップにて用いる計算機システム実装の細かい部分の問いかけに対し回答が得られておらず、当該個人を含む宛先に対して全体のスケジュール調整についての呼びかけのみがなされている場合などがあった。なお、両方の線が 180 度あけて反対側にあっても、発言内容と言及内容が「逆」とは限らず、両者の内容が異なっていることを示しているだけである。すなわち、発信者としての立場と言及される立場が噛み合っていないことを表しているとは限らない点に、注意が必要である。

続いて、複数の期間に渡った表示について確認を行ったところ、ほぼ同じような内容のメールを発信している個人は、発信に関する線が同じ方向を維持し続けていることを確認できた。例えば、書記を担当し主として会議の議事録を発信しているメンバーの場合、発信メールを示す方向がほぼ同一の方向を維持し続けることが確認できた。逆に、発信メールを示す方向が変化するときには、内容自体が変化していた。計算機システムの実装に関する議論が落ち着き、打ち合わせの会場調整を行って

いる場合などである。

以上より、本システムを用いて、個人がメーリングリスト上で発信している内容と、その個人が言及されている内容にどの程度の差があるかの確認と、この差がプロジェクトの進行にともなってどのように推移したかの確認を行うことが可能である。

4.2 議論対象に関する言及の分析

本分析システムでは、蓄積されたメールをあらかじめ複数のトピックの遷移として構造化した後、本システムを利用して分析を行う者が選択した「単語」に関連するトピックの表示を行う。例えば、単語として「テスト」を選択すると、テスト用のデータ登録に関する議論、人を集めてのプロジェクトテストなどワークショップ実施のテストに関する議論と進んでいることを確認できた。また、pLSIを用いてトピック抽出を行ったことにより、バックエンドのシステムに関する議論、フロントエンドのシステムのテストに関する議論が別々に抽出できることや、テスト環境における負荷に関する議論、テストデータに関する議論、テスト表示に関する議論が並行して行われている様子も確認できた(図3)。

このように、本システムによる表示を用いて、ある対象について、別々の視点から行われている議論の推移を時間経過に沿って確認を行うことが可能である。

4.3 手法を組み合わせでの分析

続いて、発信・言及内容による分析システムにおいてメールの発信・言及内容に相違がある部分から相違内容の根拠となる単語を見つけ出し、議論対象に関する言及の分析システムを用いて議論の変化を追うことで、プロジェクトの進捗に伴う相互理解状況の遷移を抽出できないか試みた。

そのために、発信・言及内容による分析システムにおいて、個人各期間で発信したメールあるいは言及したメールをひとまとめの「文章」とみなし、その「文章」に出現する単語をTFIDFにより重み付けし、相違の元となっている単語の探索を試みた。しかしながら、発信・言及に大きな差がある部分では、重複する単語が少なく、重複している部分があったとしてもその単語に関する立場の相違が議論として継続する部分を見つけることが困難であった。

これには、議事録など全体を包括する内容で個人が言及されている場合、個人の発信内容と差異があってもプロジェクト上の「相違」はないこと、発信・言及内容による分析ではpLSIにより次元縮約し単語の確率分布として方向性を得ており、単語レベルでの相違を見つけることが難しいことなどが、理由として考えられる。

5. おわりに

本稿では、メールのアーカイブを分析するために、個人が発信したメールと個人へ言及したメールの方向性を時間経過にそって提示するシステム、および指定された単語に対する言及内容を時間経過にそって提示するシステムを提案した。分析はまだ完了していないが、構築したシステムにより、発信と言及のおよその重なりと相違や、指定した語に関する議論内容を時間経過にそって把握できることを確認した。

本稿では単純に発信、言及内容について分析を扱ったが、はじめに述べた、期待すること・されることの相違と変動を抽出するためには、議事録や日程調整など作業のコーディネーションに関するやりとり[中小路 09]をアーカイブから除外し、問題解決に関するメールを選んで処理する必要があるだろう。

また、現在は個人のみに着目して発信と言及を扱っているが、プロジェクトには同じ分野に属する複数の個人が参加していることから、単一の個人のみではなく同分野の人物をまとめるなど、粒度を可変とした分析が必要であろう。

また、メーリングリストの分析に関連し、話題の自動抽出[斉藤 98]、議論構造の抽出[中山 08]などの研究がなされている。本稿同様に投稿者を中心とした分析では、オープンソースソフトウェア開発のメーリングリスト中に現れる単語をもとに投稿者の心理的な状態を分析する研究[Rigby 07]がある。また、本稿と同じプロジェクトについて発表論文の共著関係の推移を中心に分析した研究を先行して行っている[濱崎 12]。本稿では、メール全文をpLSIにより分類した結果を利用し、個人に関連するメールの傾向を表示すること、ある対象についての言及内容の推移を表示することにとどまったが、メール中に現れる表現と期待状態や内容と登場する個人・グループとの関係の対応付けを同時に行えるようにし、実時間を手がかりにこれらを同期させることで、メンバー相互の期待の一致とずれ、およびその変動をより細かく抽出する手法が考えられるだろう。

分析を通し、異なる文化との共同において、現在の失望は次への期待を生み出すことを明らかにできれば幸いである。

参考文献

- [Hofmann 99] Hofmann, T.: Probabilistic Latent Semantic Indexing, in *Proc. of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 50–57 (1999)
- [Rigby 07] Rigby, P. C. and Hassan, A. E.: What can OSS mailing lists tell us? A preliminary psychometric text analysis of the Apache developer mailing list, in *Proc. of the 4th International Workshop on Mining Software Repositories*, p. 23 (2007)
- [吉野 12] 吉野 孝: 学と民の協働によるシステム開発, 情報処理学会誌, Vol. 53, No. 4, pp. 365–368 (2012)
- [小早川 12] 小早川 真衣子, 高見 知里, 濱崎 雅弘, 古堅 真彦, 家近 詠子, 新野 佑樹, 須永 剛司: 表現ネビュラ: ミュージアム来場者のための表現の場, デザイン学研究作品集, Vol. 17, No. 17, pp. 14–19 (2012)
- [斉藤 98] 斉藤 典明, 水澤 純一, 山本 平一, 山口 英: 話題の自動抽出による電子メールの情報組織化手法, 情報処理学会論文誌, Vol. 39, No. 10, pp. 2907–2913 (1998)
- [中山 08] 中山 祐貴, 宮寺 庸造, 横山 節雄, 中村 勝一: 電子メール中の議論過程抽出手法の提案, 電子情報通信学会技術研究報告. ET, 教育工学, Vol. 108, No. 354, pp. 35–40 (2008)
- [中小路 09] 中小路 久美代, イェ コンウエン, 山本 恭裕: ソフトウェア開発における知識コミュニケーションのためのインタラクションデザイン, 第 23 回人工知能学会全国大会予稿集 2I3-1 (2009)
- [濱崎 12] 濱崎 雅弘, 沼 晃介, 田中 克明: 異分野越境型プロジェクトにおけるコミュニケーションとコラボレーションに関する一考察, 第 26 回人工知能学会全国大会予稿集 1P1-OS-14-3 (2012)