

ユーザ操作に基づく電子メール分類方式

6Bb-7

田中みどり 工藤正人 古関義幸

日本電気（株） C&C研究所

1 はじめに

近年、業務・教育・プライベート等様々な目的で電子メールが利用されるようになった。それに伴い個人の扱うメール数が増え整理や検索に費やされるコストが増大し、電子メールの本来の利用目的である業務やコミュニケーションに支障をきたすようになってきている。このような大量のメールの整理のコストを軽減するためには、メールをある基準にしたがって自動的に分類することが有効である。

電子メールの自動分類では、「正確性」「ユーザ適応性」が要求される。なぜならば、上述したように電子メールは日常業務のためのツールとして利用されており、正確に分類することができなければ業務遂行の弊害となってしまふからである。また、電子メールでやりとりされる話題や業務内容は時間を追って変化しておりそれに伴ってユーザの分類基準も変化していく。このような変化にすばやく適応できなければユーザの満足する分類はおこなうことができない。

従来のテキストデータの自動分類の研究には、ユーザが定義した分類ルールを利用するもの [1] と、分類ルールを自動的に学習するものがある [2]。ユーザ定義型の分類は正確性は優れているがユーザの手間が大きい。一方、学習型の分類では、正確性も低く、ユーザ適応のための学習時間が長くかかってしまい、電子メールのようにすばやく適応性を要求される応用分野には不適合である。したがって、電子メールを対象とした自動分類にはユーザ定義型分類が適していると考えられるが、「いかに簡単に分類ルールを入力保守できるか」という問題が重要になる。特に、プログラミングやルール記述に不慣れな非技術者による利用が増えると、この問題の解決が不可欠である。

本稿では、この問題を解決するために『ユーザの意

図する分類がどのようなものなのか、その一例をユーザ自身に入力させ、そこから分類カテゴリごとに共通するメールパターンを作成する』という手法を提案する。本手法によれば、ユーザは特別なプログラミング言語やルール記述形式を覚える必要はなく、分類例を示すだけで自分の意図する分類ルールを作成することができる。また、分類基準を変更しなければならない場合にも一回事例を示すだけですぐに分類ルールを修正することができる。

2 分類パターン表現

各分類カテゴリには、そのカテゴリに分類されるべきメールの特徴を表すメールパターンの集合が分類パターンとして定義される。分類パターンは、肯定メールパターン集合と否定メールパターン集合からなり、あるメールが、肯定メールパターン集合の中のいずれかに一致し、かつ、否定メールパターン集合のいずれにも一致しない場合にそのカテゴリに分類されることを意味する。図1は分類パターンの例を示し、『差出人が midori か yukari であるメールで、サブジェクトが prj1 ではないメールはカテゴリ prj-member に分類する』ことを表す。

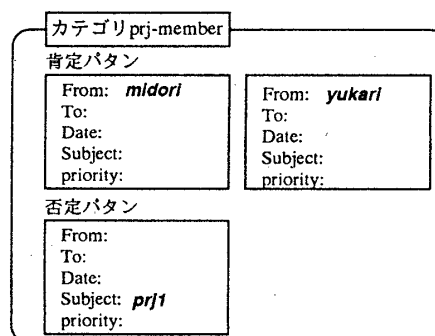


図1 分類パターン例

新たなメールが受信されると、このメールに対してすべての分類パターンを照合し、分類カテゴリを決定する。複数のパターンとマッチした場合には、そのメールは複数のカテゴリに分類される。逆に、マッチするパ

タンがひとつもなかった場合、そのメールは「未分類」とされる。

3 ユーザ操作に基づく分類パターン作成

図2に分類パターン作成のプロセスを示す。

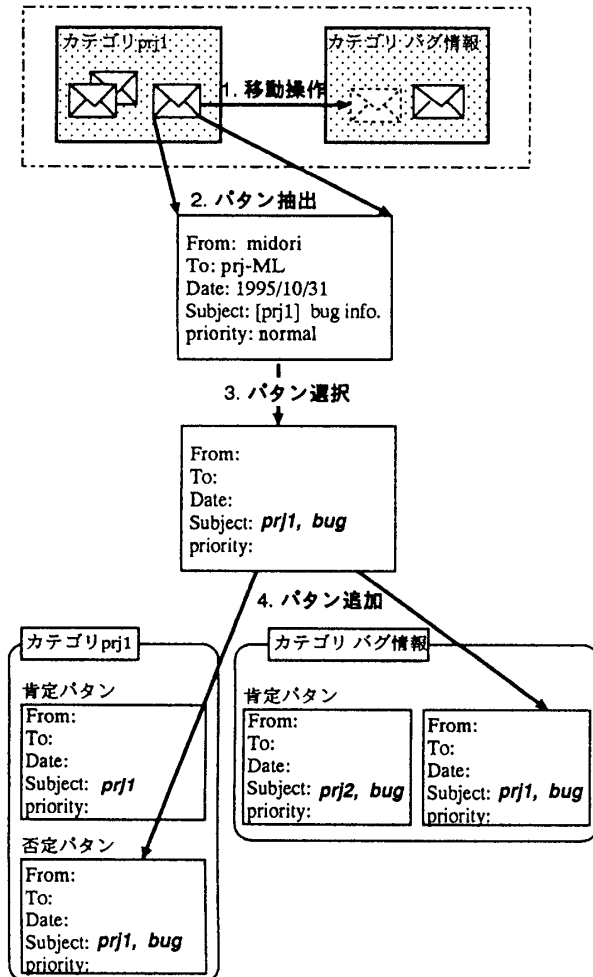


図2 分類パターン作成フロー

(1) ユーザによる分類操作

画面上には、全メールの分類されている様子が表示されている。ユーザは、現在の分類基準を変えたいと考えた場合に、画面上のメールを直接操作して自分の意図する分類を例示する。たとえば、「これまで prj1 に関するメールはひとつにまとめていたが、その中でバグ関連のものは、バグ情報を集めたカテゴリに分類するように変更したい」場合には、カテゴリ"prj1"の中から該当するメールを選びカテゴリ"バグ情報"に移動する。

操作するメールは複数でもよいが、移動先は同一カテ

ゴリとする。

(2) パタン抽出

ユーザの操作したメールからメールパターンを作成する。操作されたメールがひとつの場合には、ヘッダ情報をすべて抽出する。操作されたメールが複数であった場合には、それらに共通する項目があるかどうかを調べ、共通項目があるならばそれをパターンとする。共通項目がないならば、操作された各メールからひとつずつパターンを作成する。

(3) パタン選択

抽出したパターンをユーザに提示し、パターンの中で必要な情報だけを選択してもらう。

(4) パタン追加

ユーザの選択したパターンを、移動元カテゴリの否定メールパターン集合と、移動先カテゴリの肯定メールパターン集合に追加する。

4 試作システム

Windows 上の電子メールクライアント試作システムに本分類機能を実装した。分類例を示すだけで必要な情報を抽出してユーザに提示するので面倒なタイプ入力をおこなう必要がなく、また、否定メールパタンの利用により複雑な分類基準も簡単にパターン化でき、非技術者だけでなくコンピュータやプログラミングに慣れた技術者にとっても有効であることを確認できた。

5 終わりに

本稿では、ユーザの分類操作例に基づいた電子メールの自動分類手法について報告した。今後は、さまざまな状況・目的に合った情報整理を実現するために、自動分類だけでなくソート機能、検索機能、自動クラスタリング機能等についても検討を進める予定である。

参考文献

- [1] E. Lutz, H. v. Kleist-Retzow and K. Hoernig, "MAFIA - An Active Mail-Filter-Agent for an Intelligent Document Processing Support," *SIGOIS Bulletin*, 11:4, 1990, pp. 16-32.
- [2] 溝口文雄, 柳田正博, 大和田勇人, "帰納学習による分類機能をもった電子メールクライアントの開発," 情報処理学会第51年全国大会, 1995, pp. 3-213-214.