5R-6

教師なしメール分類手法を用いたメールボックス自動生成システムの提案

平岡 佑介 † 大囿 忠親 † 伊藤 孝行 † 新谷 虎松 †
名古屋工業大学 大学院工学研究科情報工学専攻 †
e-mail: {hiraoka, ozono, itota, tora}@ics.nitech.ac.jp

1 はじめに

本稿では，メールボックス自動生成システムの提案および本システムを実現するためのメール分類および分類ルール生成手法を示す．メールが一般的に利用されるようになり，連絡手段だけでなく，知識収集にメールが利用されるようになった [1]．その際，ユーザは興味のある多くのメーリングリストに登録し，大量のメールを受信する．現在では，大量のメールを整理するため，ユーザにメールの分類ルールの作成を行わせることで本問題を解決している．本研究では, メール分類ルールの作成における負荷を軽減する目的で，メールボックス自動生成機能を持つメール閲覧支援システム WisdomMail を試作した．本システムは教師なし分類手法である Self Organizing Map[2]（以下 SOM）を用い，メールの内容に基づくメールのメールボックスの生成を行い，メール分類ルールの生成を行なう．

2 メール閲覧支援システム WisdomMail

図 1 にメール閲覧支援システム WisdomMail のスナップショットを示す．本システムは INBOX に含まれる英語メールをメールボックス生成機構に入力し，出力された新規メールボックス及びメール分類ルールをユーザに提示する．メール分類ルールは複数の条件の組み合わせで作成されており，ユーザは必要に応じてメール分類ルールを変更及び管理できる．

3 Self Organizing Map（SOM）

SOM は教師なし競合強化学習及び近傍学習により，ある分布に従う n 次元ベクトルで表現された入力データの特徴を抽出し，その分布を近似した特徴マップを生成する．SOM のネットワークは，入力層と２次元平面マップ上にノードを格子上に配置された出力層の２層からなり，入力されたデータが出力ノードの１つに出力される．各ノードは２次元平面上に表記されるため，類似した特徴を持つデータはマップ上の近い位置に出力される．従って，データの出力された位置を用いることでデータの分類を行うことが可能である．

The Proposal the automatic mailbox creating system by using unsupervised mail clustering
Yusuke HIRAOKA, Tadachika OZONO, Takayuki ITO, and Toramatsu SHINTANI
Graduate School of Engineering, Nagoya Institute of Technology Gokiso, Showa-ku, Nagoya 466-8555 JAPAN

図 1: WisdomMail のスナップショット

表 1: メール分類ルールにおける条件

条件	説明
inSubject(w)	メールのサブジェクトに語 w が含まれる
inFrom(add)	メールの From に宛先 add が含まれる
inTo(add)	メールの To に宛先 add が含まれる
inCc(add)	メールの Cc に宛先 add が含まれる

4 メールボックス生成手法

4.1 メール分類ルールの定義

メール分類ルールとは，メールを指定のメールボックスへ移動するためのルールであり，１つ以上の条件の AND または OR で表現される．表 1 に本システムで利用した条件の一覧を示す．本ルールに従い，人工知能学会のメーリングリスト[1]から Call for Paper に関するメールをメールボックス Mailbox1 へ移動するルールを記述した例を次に示す．

inFrom(”admin@ai-gakkai.or.jp”) ∧
inSubject(”CFP”) → move(”Mailbox1”)

4.2 メールボックス生成手法の流れ

図２にメールボックス自動生成手法の流れを示す．本システムは入力としてメールを受け取り，メールボックス及びメールの分類ルールを出力する．

4.3 メールスレッド分類処理

メールにおいて返信を行う際には，返信対象となるメールを前提とした内容のメールが送られることが多く，誤分類の原因となる．本手法では，返信関係のあるメールをスレッドとしてまとめ，スレッドに対して分類処理を行うことで誤分類を防いだ．現在の実装では，「Re:」を除いたサブジェクトが同一のメールを返信関係を持つメールとして１つのスレッドにまとめた．また，スレッドのサブジェクトとしてスレッドが開始

[1]http://www.ai-gakkai.or.jp/jsai/ml/

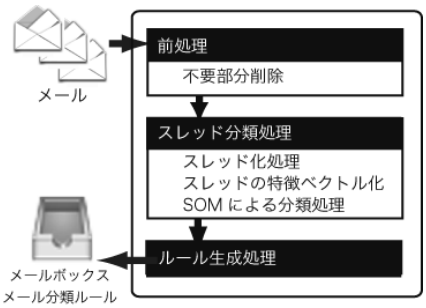

図 2: メールボックス生成手法の流れ

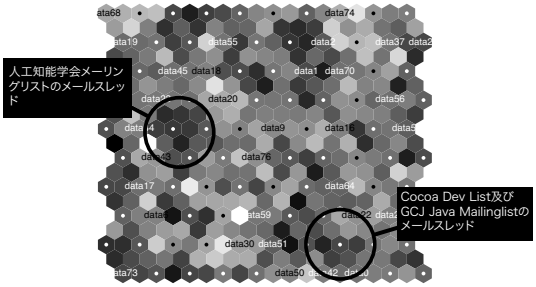

図 3: SOM によるメールスレッド分類処理結果例

されたメールのサブジェクトを用い，スレッドの本文にはスレッドに含まれる全てのメール本文を結合した文字列を用いた．

次に，得られたスレッドの特徴ベクトル表現を取得する．ベクトル表現への変換処理を以下に示す．1．スレッド本文又はサブジェクトに単語 w が出現するスレッドの数を全て調べる．2．単語を出現するスレッド数の多い順に並べ替えて，上位 200 の単語 $(w_1, ..., w_{200})$ を特徴ベクトルの要素とする．3．スレッドの特徴ベクトルの n 番目の要素 f_n を次式で求める．

$$f_n = \begin{cases} 1 & \text{サブジェクトに単語 } w_n \text{ が含まれる} \\ 0.5 & \text{本文にのみ単語 } w_n \text{ が含まれる} \\ 0 & \text{それ以外} \end{cases}$$

得られたベクトル表現を SOM 分類器に入力し，分類処理を行う．本システムでは 10×10 の SOM を用いた．実際に分類処理を行い，2 次元平面に視覚化した結果を図 3 に示す．本例では実験として，人工知能学会のメーリングリスト，Java に関するメーリングリストである Cocoa Dev List[2]，GCJ Java Milinglist[3]を含めた 150 通のメールを対象に分類処理を行った．図 3 に示すマップにおいて丸で囲った部分に含まれるノードにそれぞれ，人工知能学会のメーリングリストのメール及び Java に関するメーリングリストのメールが出力され，本結果から 2 つのトピックに関するメールを分類できたと考えられる．

4.4 ルール生成処理

SOM によって分類された結果を用いてルールの生成を行う．以下に OR 結合したルールの生成処理を示す．

1. マップ中のある (x,y) 座標に分類されたスレッドのリスト T_{xy} を取得する．
2. $t_{xyn}(t_{xyn} \in T_{xy})$ の満たす条件 c を取得する．
3. 全メールスレッド中から c を満たすスレッドの数を N(c) とし，c を満たす T_{xy} 中のスレッドの数を n(c) とする．
4. $n(c)/N(c) > \alpha$ のとき条件 c を OR 条件の一つとして採用する．現在の実装では，$\alpha = 0.4$ を用いた．

本手法を用い，図 3 に示す分類結果から Java に関するメーリングリストからのメールを Mailbox1 に分類する inCc("cocoa-dev@lists.apple.com") ∨ inTo("java@gcc.gnu.org") → move("Mailbox1") のルールが出力された．

5 おわりに

本稿では，ユーザのメール分類ルール作成にかかるコストの軽減を目的として，メールボックス自動生成機能について述べた．既存にも，ベイジアンネットワークを用いたメールの分類システムとして POPFile[4]があるが，ユーザが分類のための教師データを用意する必要があり，ユーザへの負担が必要となる．教師なし分類を用いてメールの分類を行うシステムとして ACEMS[3] があるが，また，検索結果の文書をクラスタリングして提示を行うシステム[5]があるが，分類ルールの生成は考慮していない．本システムはメール分類の結果を用いてメールボックス及び分類ルールを生成し，ユーザに提示することで，ユーザがメール分類を行う際の負荷の軽減を行うことができる．

参考文献

[1] Steve Whittaker, Candace Sidner: "Email overload: exploring personal information management of email", Human factors in conputing systems(1996)

[2] T.Kohonen, 徳高平蔵, 岸田悟, 藤村喜久郎:"自己組織化マップ", シュプリンガー・フェアラーク東京 (1996)

[3] Olle Balter, Candace L. Sidner: "Bifrost Inbox Organizer: Giving users control over the inbox", In Proc of the Second Nordic Conference on Human-Computer Interaction(2002)

[2]http://lists.apple.com/mailman/options/cocoa-dev/
[3]http://gcc.gnu.org/ml/java/

[4]POPFile:http://popfile.sourceforge.net/
[5]Find.com:http://www.find.com/