

目的に類する送信履歴を用いたメール作成支援システム

大谷 紀子 研究室

1672034 グエンフーズン

1. 背景と目的

現代社会では、メールがさまざまな場面で使われており、時候の挨拶、顧客とのやりとり、打ち合わせの報告、就職活動の応募など、似たようなメールを作成する場面が多くある。人間は一般的に同じ作業を繰り返すことに負担を感じるので、似たような文面のメールを何度も書く作業にはまったく新しい文面のメールを書くよりも多くストレスを感じる傾向がある。対策手段として、前に書いたメールを検索し、コピーペーストして内容を書き換える方法がある。しかし、メールを検索する際に、既存メールに書かれた適切な単語を考える必要がある。検索単語を含むメールが大量にある場合、同じ内容のメールが検索結果として多く提示されるので、コピー元とするメールの選択に手間がかかる。また、メール本文のコピーペーストによりメール作成効率が低下する。既存のメールシステムでは、定型メール作成機能が提供されている。しかし、本機能ではユーザがあらかじめテンプレートを設定する必要があるので、定型メールの種類が多い場合には活用が困難である。

本研究ではメール作成の効率を向上させることを目的とする。既存の送信メールに基づいて定型メールを自動決定し、メール作成作業の負担を軽減するシステムを提案する。

2. システム概要

本システムは、メール作成に先立ち、既存のメールを内容に基づいてクラスタリングし、各クラスタを代表するメールを定型メールとして設定する。メール作成の際には、入力されたキーワード

から目的に合う定型メールを検索し、件名、および内容を表す単語とともに提示する。ユーザは、提示された定型メールから1通を選び、内容を修正して新しいメールを完成させる。

2.1 定型メールの自動決定

定型メールの自動決定には、 $tf \cdot idf$ 法と k -means法を適用する。まず、メールの内容を特徴づける重要な単語を $tf \cdot idf$ 法により抽出し、各メールをベクトルで表現する。 $tf \cdot idf$ 法とは、単語の出現頻度 tf と出現文書の偏在度 idf によって単語の重要度を数値化する手法である[1][3]。 $tf \cdot idf$ の値が大きい単語を重要単語とし、各メール中の重要単語の $tf \cdot idf$ を成分とするベクトルを生成する。

生成されたベクトルを k -means法によりクラスタリングする。 k -means法では、データの平均を表す重心により、データを与えられた k 個のクラスタに分ける[2]。まず、 k 個のベクトルをランダムに選び、クラスタの重心にする。すべてのベクトルが、 k 個の重心のうち最も近い重心のクラスタに属するよう、クラスタを生成し、クラスタ内ベクトルの平均ベクトルを新たな重心とする。以上の処理を重心が変化しなくなるまで繰り返す。各重心に最も近いベクトルのメールを当該クラスタの定型メールとする。

2.2 定型メールを活用したメールの作成

メール作成の際、ユーザが入力したキーワードと、キーワードの類義語を検索クエリとして定型メールのデータベースを検索する。検索結果として、定型メールの件名と重要単語を表示する。ユーザは、表示情報からメールの内容を判断し、作

成元とするメールを選ぶ。選んだメールを表示する画面で直接文章を修正して送信する。

3. 評価実験と考察

工作上頻繁にメールを書く社会人5名を被験者として、評価実験を実施した。被験者1~5の既存メールと本システムによる定型メールの数を表1に示す。まず、被験者に、似たようなメールを作成する場面を想定するよう指示し、いつも通りの方法でメールを作成させた。次に、本システムを使用して同じメールを作成させた。メール作成後、アンケートに回答させた。また、アンケートでの回答理由、システム使用上の問題、改善希望についてのヒヤリングを実施した。

本システムによるメール作成作業に関しては、4名が「負担の軽減を感じない」と回答した。理由として、「表示ウィンドウが多くて見づらい」、「機能は便利だがシステムの使い方に慣れない」という意見が多く得られた。以上より、UIやUXを改善する必要があるといえる。

本システムの処理時間に関しては、3名が「全体的に時間がかからなかった」と回答した。今回のように、既存メールが数十通程度の場合には、定型メールの自動決定にかかる時間は問題にならないといえる。

本システムの機能については、4名がメールを表示する画面のままで修正して送信できる点を高く評価した。ヒヤリングでは、「選んだ文章をメール作成画面へコピーペーストすることは不便」という意見が多く得られた。メールの選択から作成への作業を改良するとメール作成の効率が向上すると考えられる。

定型メールの検出の正確度については、被験者全員が「検索したいメールが出た」、または「似たようなメールが出た」と評価した。本研究のクラスタリング手法によってメールが適切なクラスタに属するといえる。

本システムの既存メールシステムへの導入につ

表1：既存メールと定型メールの数

被験者	既存メール	定型メール
1	62	25
2	20	9
3	14	7
4	31	9
5	86	50

いては、4名が「なくても大丈夫だけど、あったら使いたい」と回答した。以上より、現状のシステムの必要性が低いといえる。

本実験の結果より、定型メールの自動決定では類似メールがグループ化されたといえる。また、本システムによるメール作成負荷の軽減に関して、メールの選択から作成への作業方法は最も高く評価された。しかし、全体的にシステムの利便性が低いといえる。利便性を向上させるため、優先的にUIとUXを改善するべきである。「表示ウィンドウが多くて見づらい」という意見が得られたことから、改善案として同じ画面で複数の作業ができるようUIを設計することが挙げられる。また、既存システムの提供者が提唱するデザインのルールに沿って設計すると、ユーザが慣れたインターフェースを実現できる。UXに関しては、ユーザが慣れた手順で使用できるように、処理の流れを既存システムと同様にすることを検討する。

参考文献

- [1]Luhn, H. P., "A Statistical Approach to Mechanize Encoding and Searching of Literary Information", IBM Journal of Research and Development, Vol.1, No.3, pp.309-317, 1957.
- [2]MacQueen, J. B., "Some Methods for Classification and Analysis of Multivariate Observations", Proc. of 5th Berkeley Symposium on Mathematical Statistics and Probability, Vol.1, pp.281-297, 1967.
- [3]Jones, K. S., "A Statistical Interpretation of Term Specificity and Its Application in Retrieval", Journal of Documentation, Vol.28, pp.11-21, 1972.