

改行・空行挿入によるメールテキスト整形のための特徴分析

村田 匡輝*, 大野 誠寛, 松原 茂樹(名古屋大学)

Characteristic Analysis for Formatting Mail Texts by Linefeed and Blank Line Insertion

Masaki Murata, Tomohiro Ohno, Shigeki Matsubara (Nagoya University)

1. はじめに

電子メールは日常的なコミュニケーションツールであり、送信者は、受信者に配慮して読みやすいメールを書くことが望まれている。改行・空行を適切に挿入することはそのためのテクニックの一つであるが[1,2]、このようなノウハウをすべての利用者が習得しているわけではなく、図1のような読みにくいメールを受け取ることも多い。受信メールに対して、改行や空行を適切な位置に自動的に挿入することができれば、読みやすさが増し、効率的にメールを閲覧できるようになる。

本稿では、メールテキストに対して適切な位置に改行や空行を挿入するシステムの実現を目指し、送信者により意識的に改行・空行が挿入されているメールテキストを対象に、その挿入傾向を分析する。

2. 分析データ

本研究では、意識的に改行が挿入されているメールでは、改行や空行は送信者自身が読みやすいと思う位置に挿入されていると想定する。ただし、人による傾向の違いを考慮するため、不特定多数の送信者により書かれた大量のメールを分析に用いることにより、多くの人に受け入れられるような改行や空行の挿入傾向を捉えることを試みる。

分析には、多人数からメールが送られている大学教員1名の受信メールのうち、意識的な改行が行われていると思われるメール48,357通を用いた。なお、意識的な改行の有無は、行長が同一の行が同一段落内に連続して出現するかといった簡単なルールを作成し判定した。また、挨拶や署名、引用部分、転送部分を簡単なスクリプトにより除去し、送信者によって書かれたメールの内容部分のみを分析の対象とした。

3. メールテキストにおける改行・空行挿入の傾向

<3・1>行長の分布 行長が何文字程度で改行が挿入される傾向にあるかを明らかにするため、行長の分布を調査した。行長を横軸に、その頻度を縦軸にとったグラフを図2に示す。行長は、全角文字を1文字、半角文字を0.5文字として算出し、グラフ上では小数点以下切り捨てで表示している。グラフ上の曲線は多項式近似曲線を示す。4~36文字で書かれた行の出現頻度が全体の90.3%を占め、特に11~21文字を頂点として山なりに分布していることが分かる。

改行の自動挿入戦略として、行長が11~21文字に収まるように改行することが望ましいが、状況によっては36文字程度になることも許すことが考えられる。

<3・2>行長のバランス 改行の挿入位置は他の行の行長に影響を受け、行長のバランスが考慮されるかを明らかにするため、前行の行長との差を調査した。差の絶対値を横軸に、その頻度を縦軸にとったグラフを図3に示す。差が大きくなるにつれてその出現頻度が少なくなっている。これは、一つ前の行とおおよそ同じ長さになるように改行が挿入されていることを示している。

改行挿入戦略として、行長のバランスを考慮し、行長のばらつきが大きくなるように改行することが考えられる。

<3・3>段落の行数 空行はどの程度の頻度で挿入される傾向にあるのかを明らかにするため、段落を構成する行数とその頻度の分布を調査した。その結果、段落内の行数が多くなるにつれ、その出現頻度は低下していた。3行以下の段落

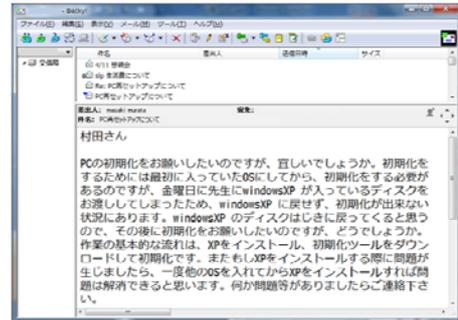


Fig 1. Sample of unreadable mail text

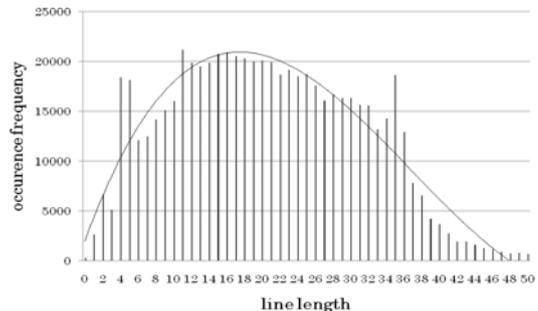


Fig 2. Distribution of line length

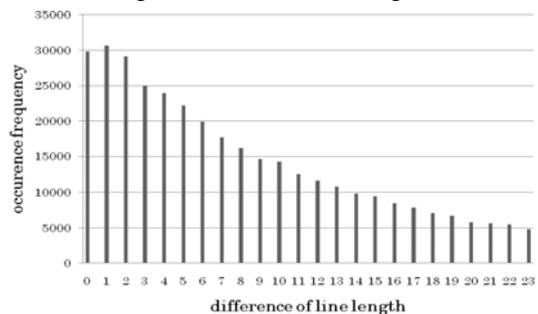


Fig 3. Differences of line length

の出現頻度は全体の81.4%を占めた。

空行挿入戦略として、段落の行数を3行程度以下にすることが考えられる。ただし、行数が4行以上の段落が出現することも許容し、行数が多くなるほど空行を挿入しやすくすることが考えられる。

4. おわりに

本稿では、48,357通のメールデータを用いて適切な改行・空行挿入についてその傾向を分析し、メールテキスト整形のための改行・空行挿入戦略について検討した。今後は、本研究で得られた成果をもとに、メールテキスト整形のための統計的な改行・空行挿入手法を開発する予定である。

文 献

- (1) 藤田, メール文章力の基本, 日本実業出版社, 2010
- (2) 村田 他, 信学論, Vol. J92-D, No. 9, pp.1621-1631, 2009