

行動履歴の Topic モデルによるモデル化とメール配信への適用

木虎 直樹 久保 征人

シナジーマーケティング株式会社 研究開発グループ

{kitora.naoki, kubo.masato}@synergy101.jp

あらまし

マーケティング活動の一環で多くの企業がメール配信を行っている。しかし、開封率やクリック率に課題を抱えている企業は少なくない。本稿では、Topic モデルを用い、行動履歴でメール受信者をクラスタリングし、クラスタごとにメールの件名や本文を最適化することにより、開封率やクリック率の向上を試みた事例を紹介する。

1 はじめに

マーケティング活動の一環で多くの企業がメール配信を行っている。今回検証した企業においても月に4回程度のメール配信を実施している。配信の際は経験則をもとに属性でセグメント分けして配信しているが、どういった属性をもとにセグメント分けするのが良いのかという明確な答えを持っていない。受信者が不要なメールを受信すると、メールがジャンクな情報として捉えられ、有益な情報があっても埋もれてしまい、結果として機会損失につながり、メールの開封率やメール内のリンクのクリック率が低いのではないかという問題意識があった。このように開封率やクリック率に課題を抱えている企業は少なくないのではないだろうか。

本稿では、Topic モデルを用い、行動履歴でメール受信者をクラスタリングし、クラスタごとにメールの件名や本文を最適化することにより、開封率やクリック率の向上を試みた事例を紹介する。

2 前提と基本的なアイデア

利用できるデータとしては、メール受信者のデモグラフィック属性とメール本文、メールクリック履歴の3つであり、メール本文の構成は図1に示すとおりである。

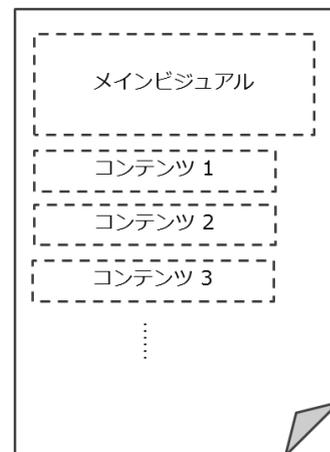


図1 メール本文の構成

コンテンツは25文字前後のタイトルと60文字前後の説明文からなる。送信するメール本文内のテキスト内容およびその量を受信者ごとに変更することはできず、メールコンテンツの並べ替えのみ可能であるという前提があった。

そのためクリック率の向上についてはメールコンテンツの並び順最適化以外に選択

肢がなかった。開封率については件名を最適化することで向上を図ることとした。開封前に受信者が目にするのは差出人と件名くらいだからである。

最初に、メール受信者には興味のあるトピックがあり、メールコンテンツは何らかのトピックに属するという仮説をたてた。このようにすることで、メール受信者一人一人とメールコンテンツ一つ一つはトピック分布で表され、メール受信者とメールコンテンツをトピックを軸とする同一の座標空間内で表現することが可能になる。

メール受信者は距離の近いメールコンテンツに興味を持つとし、メールコンテンツを距離の近い順で並べることとした。しかし、受信者ごとに並び順を最適化して送りわけるとすると、配信パターンが多くなり現実的ではないため、受信者をクラスタリングしクラスタごとにメールを最適化することとした。

3 具体的手法

クラスタリング

クラスタ作成の流れを図2に示す。

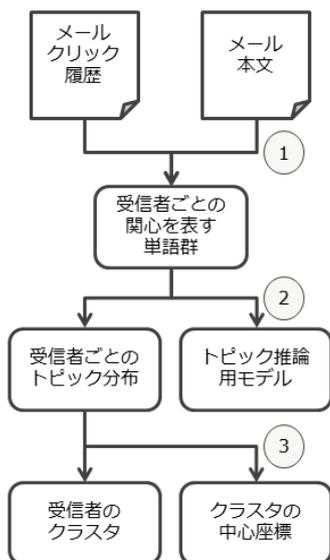


図2 クラスタ作成の流れ

- ① メール本文中のクリックしたリンク付近のテキストから単語を抽出し、メールクリック履歴をもとにメール受信者ごとの関心を表す単語群を生成する。
- ② メール受信者ごとの関心の高い単語群をそれぞれ文書とみなして、トピックモデルを適用し「メール受信者ごとのトピック分布」と「トピック推論用モデル」を得る。トピックモデルには LDA [1]を使用した。
- ③ トピック分布をベクトルとみなし K-means によりクラスタリングし、受信者のクラスタとクラスタの中心座標を得る。

メールコンテンツ並び替え

メールコンテンツ並び替えの流れを図3に示す。

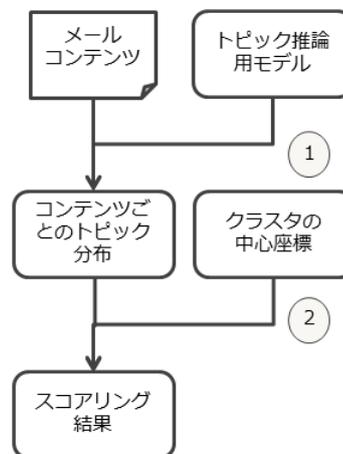


図3 メールコンテンツ並び替えの流れ

- ① クラスタリング時に作成したトピック推論用モデルを利用し、メールコンテンツのトピック分布を求める
- ② トピック分布をベクトルとみなし、クラスタリング時に求めたクラスタの中心座標との距離を計算し、近いもの順に並び替える

4 実験

内容

最初に2011年6月から2013年9月までの2年間強のメールマガジン約100通の内容と配信結果をもとにメール受信者のクラスタリングを実施した。クラスタ数は人数バランスとクラスタの特徴を考慮し4とした。

効果検証は2013年の10月と11月に配信したメールマガジンのうち2通にて実施した。10月配信時に行動履歴が存在しクラスタリング可能であったメール受信者は24,334人であり、11月配信時は24,689人であった。これを4つのクラスタに分割し、更に各クラスタをランダムに4分割することで、図4のように配信群を作成し、件名最適化とコンテンツ並び順最適化の有無によりA/Bテストを実施した。

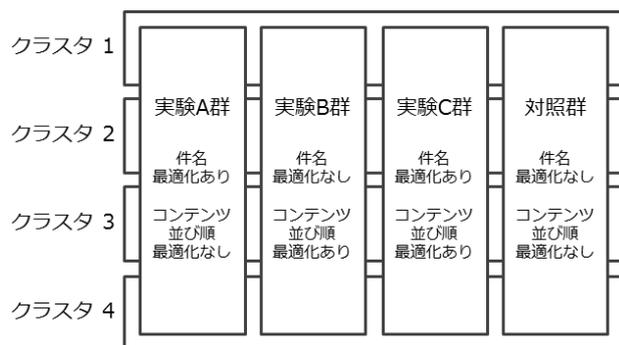


図 4 配信群

結果

10月の配信結果を表1に、11月の配信結果を表2に示す。

表 1 2013年10月の配信結果

	実験A群	実験B群	実験C群	対照群
開封率	22.3%	21.1%	20.9%	20.5%
開封後のクリック率	7.9%	7.6%	9.1%	7.0%
クリック率	1.8%	1.6%	1.9%	1.4%

表 2 2013年11月の配信結果

	実験A群	実験B群	実験C群	対照群
開封率	21.1%	21.1%	22.7%	20.1%
開封後のクリック率	5.7%	6.1%	5.9%	5.0%
クリック率	1.2%	1.3%	1.3%	1.0%

ここで、開封率、開封後のクリック率、クリック率とは以下の数式で定義される。

$$\text{開封率} = \frac{\text{開封数}}{\text{配信数}}$$

$$\text{開封後のクリック率} = \frac{\text{クリック数}}{\text{開封数}}$$

$$\text{クリック率} = \frac{\text{クリック数}}{\text{配信数}}$$

表中で背景色が付いているセルは、対照群との比較において有意水準10%で統計的な有意差が見られる箇所である。群間比較の検定はカイ二乗検定にて行った。

クリック率は1~2%であるということもあり、すべての組み合わせで想定した結果が得られたわけではないが、おおよそ以下の傾向が見られた。

- ・ 件名の最適化は開封率を向上させる
- ・ コンテンツ並び順の最適化はクリック率を向上させる
- ・ 件名とコンテンツ並び順を共に最適化することで相乗効果が期待できる

5 おわりに

今回の検証で、メール受信者の行動履歴に基づき、クラスタごとにメールの件名や内容を最適化することでメールの開封率やクリック率が向上するという傾向を確認することができた。今後はメールクリック履歴だけではなく、Webアクセス履歴のデータを加えることで、より良いクラスタの作

成とコンテンツの最適化を目指したい.

参考文献

- [1] Blei et al, “Latent Dirichlet Allocation”,
Journal of Machine Learning Research,
Vol. 3, pp. 993-1022, 2003.