メール到着時間の規則性推定

畠田　幹太　　三浦　孝夫
法政大学理工学部創生科学科

1. 前書き

インターネットを介した連絡や情報入手手段として多様な手法のひとつに電子メールがある。本研究では、電子メールデータを分析し、到着時間の規則性を検討する。特徴を確率的にモデル化し、モデルと送信者（グループ）との対比を試みる。

2. 電子メール

各電子メールの属性には、受信日時・時間、曜日、時間帯、受信間隔、送信者種別などがあり、個々のメールに付随するものや、外部から付与するものまでを検討する必要がある。確率論的な観点からは、到着時間や到着間隔のモデル化のために、ポアソン分布や指数分布を使うことが多い[1]。

3. 電子メールの分析

ポアソン分布について検討する。収集した各メールデータの時間、メール数から λを推定する。

パラメータ λの最尤推定 $\lambda = \sum_{i=1}^{k} x_i / n$

求めた λを用いてポアソン確率 $P(x) = \frac{e^{-\lambda}}{k!}\lambda^k$ を得る。

指数分布についても検討する。λを推定する。

パラメータ λ の最尤推定 $\lambda = \frac{n}{x_1 + \cdots + x_n}$

求めた λを用いて指数確率 $P_\lambda(x) = \lambda e^{-\lambda x}$ を得る。

4. 実験

4.1 準備

本研究でデータとして用いるのは、法政大学Ｇメール、私用 Yahoo!メール、約 2 年 9 か月間分、19658　通である。

4.2 実験手順

各メールに対し、送信者、グループを抽出し、メール到着時間の分布がポアソン分布に従うかを検討する。実測値と推定値を、検定を用いて比較する。

同様に、指数分布についても検討する。実測値と推定値を、検定を用いて比較する。

4.3 結果

表 1　実験結果

・ポアソン分布に従う
t-news、GAP、amazon、楽天市場、マイナビ
・指数分布に従う
[法政大学工体連サッカー部OB]
・どちらにも従わない
全メール、大学管理者、三浦先生、Yahoo!ショッピング
リクナビ、[ショッピングサイト]、[就活サイト]

4.4 考察

ポアソン分布に従うメール分布の特色は、t-news、GAP からのメールのように、1 日 1 件以下で特定の時間にメールを受信する。amazon、楽天市場からのメールのように、連続性、ばらつきはあるものの、やはりどちらも 24 時間周期性が強い。次のメールが来るまでの時間が 19～24 時間の確率が高い。

表 2　t-news、GAP、amazon、楽天市場からのメール受信間隔

メール受信間隔	t-news	割合	GAP	割合	amazon	割合	楽天市場	割合
0～6	0	0.0%	0	0.0%	103	18.8%	33	9.9%
7～12	0	0.0%	0	0.0%	15	2.7%	3	0.9%
13～18	4	0.5%	2	0.6%	41	7.5%	14	4.2%
19～24	673	86.5%	174	54.7%	206	37.6%	86	25.8%
25～30	88	11.3%	26	8.2%	101	18.4%	28	8.4%
31～36	0	0.0%	0	0.0%	7	1.3%	4	1.2%
37～42	0	0.0%	0	0.0%	7	1.3%	7	2.1%
43～48	8	1.0%	91	28.6%	15	2.7%	65	19.5%
49～54	4	0.5%	11	3.5%	13	2.4%	21	6.3%
55～60	0	0.0%	0	0.0%	0	0.0%	6	1.8%
61～66	0	0.0%	0	0.0%	4	0.7%	9	2.7%
67～72	0	0.0%	10	3.1%	7	1.3%	18	5.4%
総メール	778		318		548		333	

指数分布に従うメール分布の特色は、法政大学工体連サッカー部 OB のように、24 時間周期性よりも連続性が強い。CM メールとは違い、個人とのやりとりでは、次のメールが来るまでの時間が 0～6 時間の確率が高い。

表 3　[法政大学工体連サッカー部 OB]からのメール受信間隔

メール受信間隔	OB	割合
0～6	49	30.2%
7～12	12	7.4%
13～18	13	8.0%
19～24	19	11.7%
25～30	10	6.2%
31～36	4	2.5%
37～42	6	3.7%
43～48	6	3.7%
49～54	4	2.5%
55～60	4	2.5%
61～66	2	1.2%
67～72	3	1.9%
総メール	162	

5. 結論

本稿では、メール到着時間の規則性を検討した。楽天市場と amazon は類似した傾向が見られた。t-news、GAP のように 1 日 1 通以下、決まった時間にメールを送信するサイトは、ポアソン分布に従い、1 日のメール受信回数、次のメール到着時間を推測することが可能である。工体連サッカー部 OB のような個人的やり取りの場合、時間帯にばらつきがあり、次のメール到着時間の推測は難しい。

参考文献

[1] 杉山将(2016) 『機械学習のための確率と統計』(機械学習プロフェッショナルシリーズ) 講談社

Copyright © 2018 IEICE