

テキストマイニングと強化学習を用いた 電子メール自動分配

上田芳弘* 成田仁志** 加藤直孝* 林克明* 南保英孝** 木村春彦**

研究の背景

顧客からの問い合わせに対応する企業のコールセンター等では、電話や FAX に換えてホームページで問い合わせを受け、メールで回答するシステムの構築が望まれている。そこで、問い合わせメールを適切な担当者に自動分配するシステムを開発し、顧客の問い合わせの入力から担当者の回答、並びにこれらのデータ管理までを効率的に行うことができるシステムを構築する。

研究内容

提案した手法は、まず回答者となる各担当者が作成した文書ファイルを収集して、この中の出現単語の重要度として $tf \cdot idf$ 値と $idf/conf$ 値を算出し、2種類の辞書を担当者ごとに作成する。さらに、従来の学習理論に代えて Profit Sharing を応用し、これらの重要度を時間経過とともに学習することが特徴である。システムは、問い合わせメールと2種類の辞書を照合して、単語の重要度と文字の一致率から担当者ごとにスコアを算出し、このスコアが高い担当者を回答者として推定する。

従来の学習理論では、辞書作成時の文書ファイルや分配先の担当者の変更・更新されると、一般に初期状態から再学習しなければならず、システムの再構成に時間や工数がかかる。よって、問い合わせサービスを一時停止するなどの対応が必要となる。そこで、システムを運用しながら逐次、時系列での変化を学習できる方法が求められ、本研究では強化学習の Profit Sharing を適用した。強化学習は、時系列に得られる報酬を手がかりとして、できるだけ多くの報酬を早く獲得することを目的とした環境適応のための機械学習理論である。

(注) $tf \cdot idf$ 値 : ある人が他の単語より多く使用し、かつ他人が使用しない単語に大きい値が与えられる。

$idf/conf$ 値 : ある人が1文中に同時に使用する2単語で、かつ他人が使用しない単語に大きい値が与えられる。

研究成果

工業試験場に寄せられた実際の問い合わせメールを用いて評価実験を行い、以下の結果を得た。

- (1) 問い合わせメールを分配している専属担当者の分配精度を調査して実用上必要な目標精度を求めた。
- (2) 提案手法により(1)の目標精度と同等以上、かつ従来理論以上の分配精度を得た。
- (3) (1)の目標精度を得るために辞書構築時に必要となる文書ファイル数を求め、このファイル中のノイズが精度に与える影響を評価した。

論文投稿

電子情報通信学会論文誌 Vol.J87-D- I No.10, 2004. p.887-898.

NO.	メールアドレス	候補者名	ヒット上位単語
1	hhhh@irii.go.jp	平野(半導体)	メッキ 分析 鉛 イオン 部品
2	dddd@irii.go.jp	堂元(表面処理)	電極 部品 鉛 分析 ある
その他の担当者(42名)			
3	iii@irii.go.jp	伊藤(触媒)	メッキ イオン 酸化 鉛

図 回答者候補の表示

* 電子情報部 ** 金沢大学工学部