

# 電子メールにおける帰納的学習を用いた 返信文自動生成手法の性能評価

6P-8

植松 智 荒木 健治 柄内 香次  
北海道大学大学院 工学研究科

## 1 はじめに

電子メールによる情報のやり取りが頻繁に行われている現在、その返信文作成に多大な時間・労力が費やされることとなった。そこで、このような問題を解決する方法として、受信メールに対する返信文の下書きを自動生成する手法が求められている。

ここで、受信文に対する返信文の生成を行うシステムは、対話システム的应用として考えることができる。ところで、これまで開発されてきた対話システムの多くは、応答文を生成する規則が予め組み込まれている[1][2]。そのため、用意されたパターン以外の状況や、ユーザ及び内容に適応した応答の生成が困難になるという問題点がある。

そこで、本稿では実例からそこに内在する生成ルールを動的に獲得することができる、帰納的学習[3]を用いた電子メール返信文自動生成手法を提案する。これは、形態素解析により得られた形態素の字面情報とその形態素の品詞情報から生成ルールを獲得し、これを用いて返信文を自動生成する。本稿では、実験を通じて本手法の性能評価を行い、実験結果から考察をする。

## 2 処理過程

まず、受信メールを形態素解析ツール JUMAN[4]を用いて形態素解析をおこない、獲得された生成ルールとの比較を行う。このとき、受信メールに対して適用できる生成ルールが複数あった場合、尤度に基づき使用する生成ルールを決定し返信文生成を行う。さらに生成ルールが適用された部分を受信文から除外し、残りの文字列に適用できる生成ルールがあるかを繰り返し比較していき生成ルールの複数適用を行う。この生成ルール適用の例を表 1 に示す。なおこの際、誤った語順の出力を防ぐために返信文から得た品詞列辞書を参照する。出力された返信文に対して、ユーザが校正処理を行い最終的な返信文を完成させる。この時削除・修正の必要があった生成ルールは尤度を下げ、そのまま用いた生成ルールは尤度を上げるというフィードバック処理を行う。なお、完成された返信文も JUMAN で形態素解析を行い、受信メール・返信メールの組から生成ルール獲得処理を行う。また、この

組をメール対データに保存する。ここでメール対データとは、入力された受信メールと最終的に作られた返信文の組を保存しておくテキストデータのことである。各処理の詳細を次章以降に述べる。

表 1 ルール適用の例

|  |
|--|
| (受信) (生成ルール一回目の適用)<br>/まつ:名詞/さん:接尾辞/は:助詞/どんな:指示詞/映画:名詞/が:助詞/好き:動詞/? :記号/         |
| (生成ルール) (尤度最大とする)<br>(受信) /まつ:名詞/さん:接尾辞/は:助詞/<br>(返信) /僕:名詞/は:助詞/_<br>↓ (繰り返し適用) |
| (受信) (生成ルール二回目の適用)<br>/どんな:指示詞/映画:名詞/が:助詞/好き:動詞/? :記号/<br>(上記の文字列に対して適用できるかを調べる) |
| (返信) (一回目の適用による出力)<br>僕は<br>↓<br>僕は (二回目以降の生成ルール適用)                              |

### 2. 1 生成ルール獲得処理

入力された受信メール・返信メールの組と、メール対データに保存されている受信メール・返信メールの組を、受信側同士・返信側同士形態素単位での比較を行う。ここで、それぞれに一致する自立語を含んだ部分文字列が存在した場合、これを生成ルールとして獲得する。(表 2 参照)

### 2. 2 品詞列辞書

返信メールにおける返信文の形態素解析結果から、その品詞列部分を品詞列辞書として保存する。複数の生成ルールを一つの文に組み合わせ適用する際、過去ユーザがどのような品詞列の並びを持った返信文を作成したかを参照することで、組み合わせの際の基準に用いることとする。

### 2. 3 生成ルールの優先度

複数の生成ルールが適用可能であった場合の優先度を以下に示す。

まず、各生成ルールは正適用回数・誤適用回数という値を保持していて、この値から正適用回数－誤適用回数を計算しこれを尤度とする。この尤度が一番大きい値のルールを選ぶ。なお、尤度が同じ生成ルールがあった場合は、生成ルールの登録が早いものを優先することとする。

\*Evaluation of Automatic Generation of a Reply  
Using Inductive Learning on Electronic Mail

\*Satoshi Uematsu, Kenji Araki, Koji Tochinai

\*Graduate School of Engineering, Hokkaido University

表2 生成ルール獲得の例

|   |
|---|
| (受信 現在)<br>/まつ：名詞/さん：接尾辞/は：助詞/どんな：指示詞/映画：名詞/が：助詞/好き：動詞/?：記号/                                      |
| (受信 過去)<br>/まつ：名詞/さん：接尾辞/は：助詞/もともと：副詞/札幌：名詞/の：助詞/人：名詞/?：記号/                                       |
| (受信 一致部分)<br>/まつ：名詞/さん：接尾辞/は：助詞/_   |
| (返信 現在)<br>/僕：名詞/は：助詞/サスペンス：名詞/系：名詞/が：助詞/結構：副詞/好き：動詞/。：句点/  |
| (返信 過去)<br>/僕：名詞/は：助詞/本州：名詞/の：助詞/人間：名詞/で：助詞/、：読点/大学：名詞/入学：名詞/から：助詞/札幌：名詞/に：助詞/住んで：動詞/まず：接尾辞/。：句点/ |
| (返信 一致部分)<br>/僕：名詞/は：助詞/_   |
| (生成ルール)<br>(受信) /まつ：名詞/さん：接尾辞/は：助詞/_<br>(返信) /僕：名詞/は：助詞/_<br>正適用回数__回； 誤適用回数__回；                  |

### 3 実験

実験には、著者の使っているメーラーに蓄えられたものから28組を任意に選んだ。再現率・適合率は以下の計算で求めることとする。入力に伴う獲得ルール数の推移を図1に、生成ルールの再現率・適合率を図2に示す。

$$\text{再現率} = \frac{\text{生成ルールが適用されたメールの個数}}{\text{入力されたメールの個数}}$$

$$\text{適合率} = \frac{\text{生成ルールが正適用された回数}}{\text{生成ルールが適用された回数}}$$

#### 3.2 評価方法

再現率の計算はメール単位で行う。入力された受信メール中の文に対して、一つでも生成ルールの適用があった場合は、生成ルールが適用されたメールの個数を加算する。

また適合率は、入力された受信メール中の文に対して生成ルールが適用された際、その生成ルールが出力した文をそのまま残して返信文として使えるかを判断し、そのまま使える場合は生成ルールが正適用されたとみなし正適用回数を増やす。

### 4 考察

実験の最終的な結果、再現率は46%、適合率は50%となった。再現率は生成ルール数の増大に伴い向上して行った。しかし、文と文の対応関係を網羅的に行った為に誤った生成ルールを獲得してし

まい、適用された生成ルールの内50%が誤適用されてしまった。

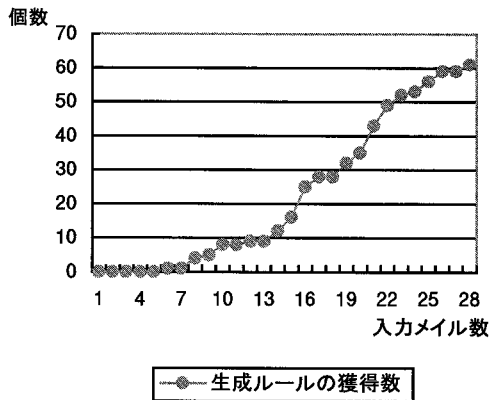


図1 入力に伴う生成ルール獲得数の推移

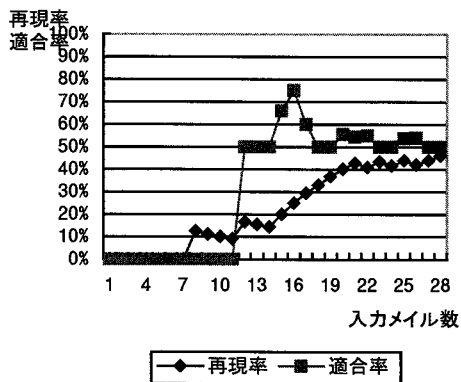


図2 生成ルールの再現率・適合率

### 5 おわりに

実験の結果、再現率46%、適合率50%となり、電子メールにおける返信文自動生成に、本手法が有効である可能性が示せた。しかし、誤りルールの獲得や、ルールの誤適用がおこってしまった。

今後の課題として、誤りルール獲得の防止及び適用の仕方について検討をしていく。

#### 参考文献

- [1]人工知能学会(編):人工知能ハンドブック,オーム社(1990).
- [2]Weizenbaum, J.: ELIZA - A Computer Program for the Study of Natural Language Communication Between Man and Machine, Communications of the Association for Computing Machinery vol.9, No.1, pp.36-45(1966).
- [3]荒木健治: 帰納的学習を用いた自然言語処理の有効性について, 信学技報, TL99-41, pp.33-40(2000).
- [4]京大工学部 長尾研究室 日本語形態素解析システム JUMAN. <http://www.nagao.kuee.kyoto-u.ac.jp/>