

メーリングリストに投稿された回答メールに対する 返信メールを利用した回答の内容確認

西村 俊介 西村 涼 渡辺 靖彦 岡田 至弘

龍谷大学 理工学部 情報メディア学科

{s_nishimura,r_nishimura}@afc.ryukoku.ac.jp, {watanabe,okada}@rins.ryukoku.ac.jp

1 はじめに

how 型の質問に答えるために、メーリングリストに投稿されたメールを知識源とした質問応答システムの研究を進めている [1]。メーリングリストには、質問や回答、その回答に対する報告や補足などのメールが頻繁に投稿されている。われわれはこれまでに、Vine Users ML *1 に投稿されているメールから質問や回答の中心となる文 (重要文) を抽出することで、how 型の質問に答えるための知識を獲得できること、そしてその知識の内容の正しさを回答メールに質問者が返信したメール (質問者返信メール) を利用して確認できることを示した [2] [3]。

本研究では、抽出した知識の内容の正しさを回答メールに質問者以外が返信したメール (非質問者返信メール) を利用して確認する方法について検討する。非質問者返信メールでは、回答メールの内容の確認や補足、質問メールへの回答など、さまざまな情報が表現されている。そこで、メーリングリストで行われる問題解決のためのやり取りを調査し、回答メールの内容を確認することができる情報を非質問者返信メールから取り出す方法について検討する。

2 メーリングリストで行われる問題解決のためのやり取り

われわれは、メール間の参照関係と投稿者のメールアドレスを利用してメーリングリストに投稿されたメールを以下の 4 種類に分類した [1]。

質問メール ある問題について、最初に投稿された質問のメール

直接回答メール 質問メールに直接回答するメール

質問者返信メール 直接回答メールに質問メールの投稿者が直接返信するメール

その他 上記以外のメール

その他に分類されたメールのうち直接回答メールへの返信メールを、**非質問者返信メール**と呼ぶことにする。質問メール、直接回答メール、質問者返信メール、および非質問者返信メールの参照関係を図 1 に示す。本研究では、この非質問者返信メールを利用して、直接回答メールの内

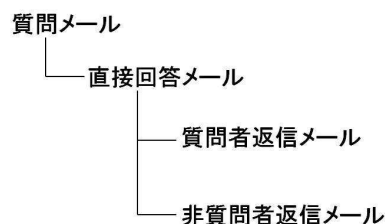


図 1 質問メール、直接回答メール、質問者返信メール、および非質問者返信メールの参照関係

容を確認する方法について検討する。以下では、直接回答メールに返信する質問者返信メールと非質問者返信メールについて述べる。

2.1 質問者返信メールの特徴

質問者返信メールでは、質問者にとって直接回答メールの内容が正しいか、誤っているかの情報が表現されていることが多い。質問者返信メールの文章には以下の特徴がある。

特徴 1 直接回答メールの一部を引用して、返事を書くことが多い。

特徴 2 回答の内容を肯定あるいは否定する典型的な表現がある。

特徴 3 1 通の質問者返信メールで複数の内容を取り扱うことがある。

これらの特徴の例を図 2 に示す。図 2 の質問者返信メールでは、最初に質問メールを投稿した質問者が直接回答メールの内容について 2 箇所引用して返答している (特徴 1、特徴 3)。そして、その返答には「できました」や「存在しないようです」など、典型的な表現が使われている (特徴 2)。図 3 に直接回答メールの内容を肯定あるいは否定するのに用いられる典型的な表現を示す。

2.2 非質問者返信メールの特徴

非質問者返信メールは、以下の 4 種類に分類できる。

1. 内容確認メール

直接回答メールの内容について肯定あるいは否定

*1 <http://vinelinux.org/ml.html> (Vine linux に関心のある人たちが情報を交換しているメーリングリスト)

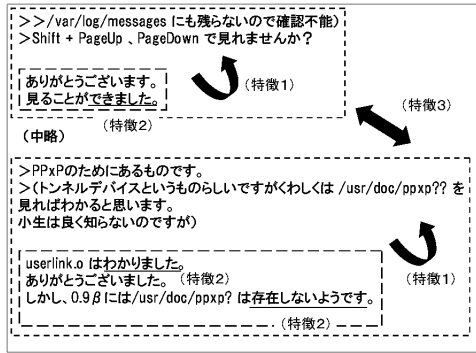


図2 質問者返信メールの特徴

- できました
- (うまく) 이었습니다
- なるほど
- わかりました
- 確かに～ました
- 安心～ました
- とおり (副詞的名詞、「以下のとおり」は除く)
- だめ
- ませんでした。
- しました。(「書きました。」「言いました。」は除く)
- ません。
- ありがとうございました

図3 直接回答メールの内容を肯定あるいは否定するのに用いられる典型的な表現

の返答が行われているメール(ただし、後述する内容補足メールをのぞく)。2.1節で述べた質問者返信メールと同じ特徴をもち、図3の表現を用いて直接回答メールの内容を肯定あるいは否定することが多い。

2. 内容補足メール

直接回答メールの内容を補足しているメール。内容補足メールは、直接回答メールの内容(例:「version 3ではインストールできた」)を確認するのではなく、関連する情報(例:「version 2ではインストールできない」)を追加するものが多い。ただし、非質問者返信メールの投稿者が直接回答メールの投稿者(回答者)でもある場合、その非質問者返信メールは直接回答メールの内容を否定する内容補足メールであることが多い。これは、回答者が直接回答メールで述べていた内容を非質問者返信メールで自ら補足して修正することが多いからである。例えば、図4(a)の直接回答メールの投稿者(回答者)は、その内容の誤りに気づいて図4(b)の非質問者返信(内容補足)メールを投稿して直接内容メールの内容を否定して修正している。

3. 質問回答メール

質問メールの内容について回答しているメール。直接回答メールと同じ特徴をもつ。以下の非質

FSSTND的にどこが好ましいかは分からないのですが、Vine(RedHat)を使う上で、
make install するなら /usr/local/ 下に置くのが管理するのに楽ですね？
なので、cmail でも自分で rpm 化してインストールするなら/usr/ 直下に入れば良いのでは。
私も cmail 使っておりますが rpm 化してないので /usr/local/mule/site-lisp/ に入れちゃってます。

(a) 内容に誤りのある直接回答メールの例

>FSSTND的にどこが好ましいかは分からないのですが、Vine(RedHat)を使う上で、
>make install するなら /usr/local/ 下に置くのが管理するのに楽ですね？
>なので、cmail でも自分で rpm 化してインストールするなら/usr/ 直下に入れば良いのでは。
>私も cmail 使っておりますが rpm 化してないので /usr/local/mule/site-lisp/ に入れちゃってます。

そういう問題じゃ無いですね。
ディストリビューション固有のパッケージで無いものは/usr/local/ 以下が好ましい というのが
正解かもしれません。

(b) (a) の直接回答メールの内容を否定する非質問者返信(内容補足)メールの例

図4 直接回答メールの内容を否定する非質問者返信(内容補足)メールの例

問者返信メールは、質問回答メールとなることが多い。

- 直接回答メールの投稿者が質問メールの投稿者(質問者)である場合
この場合にあってはまるのは、最初の質問メールで質問を十分に説明できなくて、質問者が追加の説明を行った直接回答メールに対して、非質問者返信メールで回答が行われている場合である。
- 非質問者返信メールへ質問メールの投稿者(質問者)が返信している場合
この場合にあってはまるのは、非質問者返信メールで行われた回答について、質問者が内容の確認やお礼を述べている場合である。
- 非質問者返信メールへの返信のメールに以下に示すような表現が含まれている
 - ー できました/できませんでした
 - ー うまくいきました/うまくいきませんでした
 - ー しました
 この場合にあってはまるのは、非質問者返信メールで行われた回答について、その返信メールで内容の肯定あるいは否定が行われている場合である。

4. その他

上記以外のメール

このように、内容確認メールと内容補足メールは直接回答メールの内容確認に利用できる可能性がある。一方、質問回答メールは直接回答メールの内容確認には利用できない。質問回答メールと内容確認に利用できる内容補足メールは、メールの投稿者を確認することで取り出すことができる可能性がある。

優先度	手がかり表現
高	できました (p) うまくいきました (p) なるほど (p) わかりました (p) 確かに～ました (p) 安心～ました (p) とおり (副詞的名詞) (p) (「以下のとおり」は除く) だめ (n) ませんでした。 (n) ました。 (p) (「書きました。」「言いました。」は除く) ません。 (n) ました (p)
低	ありがとうございました (o)

図 5 内容確認メールを抽出するのに用いる手がかり表現

3 非質問者返信メールを用いた直接回答メールの内容確認

非質問者返信メールを用いた直接回答メールの内容確認の方法を以下に示す。

step 1 [直接回答メールの内容を否定する内容補足メールの抽出]

直接回答メールの投稿者(回答者)が投稿した非質問者返信メールを、直接回答メールの内容を否定する内容補足メールとして取り出す。取り出した非質問者返信メールが返信している直接回答メールに、否定(negative)の内容確認ラベルを与える。

step 2 [質問回答メールの削除]

以下のいずれかの条件をみたす非質問者返信メールを質問回答メールとして取り除く。

- 直接回答メールの投稿者が質問メールの投稿者(質問者)である
- 非質問者返信メールへ質問メールの投稿者(質問者)が返信している
- 非質問者返信メールへの返信のメールに以下の表現が含まれている
 - － できました/できませんでした
 - － うまくいきました/うまくいきませんでした
 - － しました

step 3 [形態素解析]

非質問者返信メールの本文を形態素解析する。形態素解析にはJUMAN[4]を用いた。

step 4 [手がかり表現を用いた内容確認メールの抽出と内容確認文の抽出]

図 5 に示す手がかり表現を含む非質問者返信メールを内容確認メールとして取り出す。取り出した内容確認メールに含まれる図 5 の手がかり表現のうち、最も優先度の高いものを含む文を内容確認文として取り出す。

表 1 抽出した内容補足メールと削除した質問回答メールの内訳

	内容確認	内容補足	質問回答	その他	合計
内容補足	0	11	0	0	11
質問回答	1	4	10	11	26

さらに、図 5 の各手がかり表現に与えられている記号 p (positive)、n(negative)、o(other) にしたがって、非質問者返信メールが返信している直接返信メールに肯定(positive)、否定(negative)、その他(other)の内容確認ラベルを与える。

4 実験結果と評価

1999 年 3 月から 1999 年 4 月までに Vine Users ML に投稿されたメールから、非質問者返信メールを無作為に 100 通取り出した。この 100 通のメールから、3 章で述べた step 1～4 の処理によって直接回答メールの内容を確認することができるかどうか実験を行った。

step 1 で抽出した内容補足メールと step 2 で削除した質問回答メールの内訳を表 1 に示す。step 1 では内容補足メールとして 11 通を取り出したが、それらはすべて正しく取り出していた。一方、step 2 では質問回答メールとして 26 通を取り出したが、そのうち正しく抽出できたのは 10 通であった。誤って取り出したものには 4 通の内容補足メールがあったが、それらは直接回答メールの内容を確認するものではなかった。step 4 では内容確認メールとして 33 通取り出したが、そのうち正しく取り出したのは 8 通であった。さらに、内容確認文を正しく取り出したのは 6 通であった。したがって、直接回答メールの内容確認に成功したのは、

- 内容確認メールから取り出した 6 通
- 内容補足メールから取り出した 11 通

の合計 17 通であった。図 6 に内容確認メールの抽出には成功したが、内容確認文の抽出に失敗した例を示す。この例では「音が出るようになりました。」という文が内容確認文であるが、誤って「私もタワー PC で音が出ずに困っていました。」という文を抽出した。なお、質問回答メールと判定して step 2 で削除した 26 通のメールには、手がかり表現が含まれているメールが 11 通あった。そのうち 2 通は質問回答メールであったが、残りの 9 通は質問回答メールではなかった。しかし、その 9 通のうち、内容確認メールであったのは 1 通だけであった。

5 まとめ

4 章の結果より、非質問者返信メールから回答メールの内容を確認することができる内容確認メールを精度よく取り出すにはさらなる検討が必要である。一方、非質問者

私もタワーPCで音が出ずに困っていました。
サウンドカードはSB16(PnP)です。
吉岡さんの以下の情報を参考にして、`/etc/conf.modules`を作成してみたところ、
音が出るようになりました。
ありがとうございました。

図6 内容確認メールの例

返信メールと直接回答メールの投稿者が同じである内容補足メールは精度よく取り出すことができ、内容確認の結果もよかった。また、非質問者返信メールが質問メールの内容に回答している質問回答メールも多く含まれていて、それらへの返信メールによる内容確認についても今後の課題である。

参考文献

- [1] 渡辺靖彦, 横溝一哉, 西村涼, 岡田至弘: “メーリングリストを利用した質問応答システムのための知識獲得.”, 自然言語処理, vol.12, No.6, pp.25-44, (2005).
- [2] 西村 涼, 渡辺 靖彦, 岡田 至弘, “メーリングリストに投稿されたメールを利用してあいまいな質問に問い返す応答システムの作成”, 言語処理学会第13回年次大会, E5-2, (2007).
- [3] 渡辺靖彦, 西村涼, 岡田至弘: “メーリングリストを利用した質問応答システムのための知識の内容確認.”, 電子情報通信学会技術研究報告, NLC2006-10, (2006).
- [4] 黒橋 禎夫, 河原 大輔: “日本語形態素解析システム JUMAN version 5.1 使用説明書.”, 京都大学, (2005).