

## 対話の自動要約における認識誤りの影響

福西 克文<sup>\*1</sup>、亀山 恵<sup>\*2</sup>、有馬 純<sup>\*1</sup>

<sup>\*1</sup>NTTデータ通信(株) 情報科学研究所

<sup>\*2</sup>SRI International Artificial Intelligence Center

会議室予約をターゲットとした対話自動要約システム(MIMI)において、システムの入力となる音声認識の性能によって要約システムの性能がどのように変化するのかを、入力テキストに対して擬似的な認識誤りを含ませる、エラーシミュレーションを行って、対話自動要約システムの、音声認識の誤認識に対するロバスト性を検証した。

## Influence of speech recognition errors in automatic dialogue summarization

Katsufumi FUKUNISHI<sup>\*1</sup>, Megumi KAMEYAMA<sup>\*2</sup>, Isao ARIMA<sup>\*1</sup>

<sup>\*1</sup>NTT Data Comm. Systems Corp.      Laboratory for Information Technology  
<sup>\*2</sup>SRI International      Artificial Intelligence Center

We investigated the extent to which speech recognition errors affect the performance of an automatic dialogue summarization system (MIMI), but simulating word recognition errors in input dialogues. The result showed MIMI's robustness against recognition errors.

### 1. はじめに

これまでに会議室を予約するときの受付係(Clerk)と依頼者(Client)との対話の自動要約というタスクを対象として、要約精度の向上に取り組んできた。これまでの取り組みにおいては、あらかじめ収録した受付係と依頼者の模擬対話を人手によりテキストに書き起こし、そのテキストをシステム入力として実験を行ってきた<sup>[1][2]</sup>。これらは、音声認識を統合した対話自動要約システム構築の前段階とし

ての取り組みである。

音声認識の認識率が100%であれば、書き起こしテキストと全く同じ入力が得られるが、実際には認識率が100%となることは考えられず、自動要約システムの入力には常に誤りを含んでいることを考慮しなければならない。

そこで、音声認識の認識誤りに対する要約システムのロバスト性を検証するため、書き起こしテキストに対してシミュレーションによる変更を加えることにより、誤認識を含む

テキストをシステム入力とした場合の実験を行った。

## 2. 対話データ

### 2. 1. 収録方法

本実験で使用した対話データの収録は、まず最初にClientとなる被験者に対し、表 1に示すような予約をしたい日時・会議室の希望や、予約したい会議室が既に予約されていた場合の対処などの、依頼時に必要となる情報を書いた依頼内容のメモをあらかじめ伝えておく。次にその情報に基づいて、受付係と対面で会話をしてもらい、それを収録したものである。表 1のシナリオ例に基づいて行った対話の例を付録に示す。

表 1. シナリオ例

処理内容	予約
名前	安部
担当名	開発企画
日付	来週のいずれかの日
時間帯	13:00～17:00
会議室	プレゼンルームまたは 第5・6会議室
OHP	使う
備考	なし

### 2. 2. 要約項目

対話の要約項目としては、本実験では「会議室の予約に関する対話」をターゲットドメインとしていることを前提にして表 2に示す9項目を設定した。これらの項目は、予約を行うときに必ず必要な情報のための6項目（項目1～6）と、会議室を選択する際の目安になる付帯情報の2項目（項目7、8）、さらには、Clientの依頼した要求がシステムによって実現されたかどうかという項目から成る。

表 2. 要約項目一覧

1	処理内容(予約かキャンセルか)
2	予約者氏名
3	予約者担当名
4	会議室
5	日付
6	時間帯
7	会議人数
8	OHP使用の有無
9	処理要求が受理されたかどうか

### 2. 3. 処理要求による分類

対話内容は、1つの対話中においてClientが要求した処理数によって、

(1) 単一要求 (Single)

(2) 複数要求 (Multi)

という2種類に分類した。

つまり、(1)はClientが1対話中において要求した予約・キャンセル処理が1つのみの場合である。また(2)は「毎週水曜日の予約をしたい」といった複数の処理要求があった場合である。

従って、複数要求の対話の場合には、それぞれの処理要求について表 1に示すような要約項目が存在し、正しく要約が行われた場合には、各要求処理ごとに要約項目が出力される。

### 2. 4. 収録データ

模擬対話は50人の被験者に対し、各3対話の計150対話を収録した。150対話の内訳としては、

(1) 単一要求 …… 96対話

(2) 複数要求 …… 54対話

であった。また、1つの対話に含まれる単語数に注目したときの解析結果を表 3に示す。

以上のような150対話を75対話ずつ2つのデータセットに分割し、一方を学習データセットとして用いることで対話自動要約シ

表 3. 対話に含まれる単語数

処理	対話数	平均	最大	最小	分散
单一	96	212	475	70	80.4
2	15	302	498	197	84.6
3	5	212	312	130	74.8
複数	4	33	401	684	172
	5	1	393	393	0
全体	150	264	684	70	264

システム(MIMI)の構築を行い、もう一方を評価データセットとして実験を行った。

### 3. 対話自動要約システム(MIMI)

本実験で使用した自動要約システム(MIMI)は、テキストからの情報抽出技術であるFAST US<sup>[3]</sup>をベースとしたCascaded Finite Transducerで、単語・句・複合句・ドメインパターンを認識する4つのフェーズからなり、「会議室の予約・キャンセルに関する受付係と依頼者の対話の要約」をターゲットドメインとして構築されている。

MIMIはドメインに関係のある情報しか認識せず、未知の単語は無視する。また、MIMIでは対話の順序に従って、1文ごとに処理を行い、処理が進むごとに要約結果が随時更新されるようになっている。

しかし自然対話を対象とするため、文法により規定できる「文」が入力として与えられるとは限らない。実際の発声内容を見てみると、対話のような自由発話においては、

- ・「文」そのもの、あるいは「文」の境界が明確ではない。
  - ・相手の発話により発話が中断されてしまうことがある。
  - ・発話の中斷は特定の位置で起こるものではなく、任意の位置で発生する。
  - ・発声間違いによる言い直しがある。
- などの現象が起こっていることがわかる。

また複数要求の対話においては、1対話の中でいくつの予約・取消が出ているか、現在、どの予約・取消について話しているか、をMI

MIは認識しなければならず、単一要求の処理に比べて非常に複雑な処理を要求される。

そこでこれらの問題点に対処するために、MIMIでは次の3点の拡張が行われている。

#### (1)柔軟な処理単位

文、あるいは発話の単位はシステムの想定する文／発話単位とはかぎらないため、入力がシステムにとって複数文だったり、1文に満たなくても処理ができるようにした。

#### (2)1文先読み

さらに、1文に満たない入力に対してうまく処理させるため、MIMIは1文先の文を処理したあとに要約を更新するようにした。

#### (3)要約の上書き

対話の進行に伴う状況の変化に応じて、Clientの要求が変更されることがある。また、言い直しも起こる。これらの場合、対話中のより後方の情報が正しい結果となる。このような現象にもとづき、要約の更新は上書きを基本とした。

## 4. 認識誤りのシミュレーション

### 4. 1. 音声認識における誤認識

連続音声認識などの文章の認識では、有限状態オートマトンなどによる文法制約を用いて、認識時の探索空間や認識対象語彙の削減などを行っている。このような場合には、ある単語の誤認識結果として得られる語彙を、文法に基づいて限定することが可能である。

しかし、本実験で使用したような対話音声は文法的制約を受けない自由発声であるため、発声された文章が必ずしも文法的に「正しい文」であるとは限らない。このことは、音声認識において文法制約を適用することが困難であることを表している。つまり、対話音声

の認識を考えた場合、文法に基づいた形での、誤認識対象となる語彙の限定が困難であることを表している。

そこで、本実験ではすべての語彙が、ある語彙に対して、誤認識結果となる可能性があると考えて実験を行った。

音声認識における誤認識を大きく分けると、(1)置換誤り、(2)削除誤り、(3)挿入誤りの3種類に分けることが可能である。

今回の実験では、これらの誤りの中でも実際に取り組みやすく、かつ、最も発生率が高いと考えられる(1)の置換誤りを対象として実験を行った。

#### 4. 2. エラーシミュレーション

エラーシミュレーションを行うにあたり、認識手法としては、我々がこれまでに行ってきました、会議室予約を対象ドメインとした連続音声認識の手法（音韻モデルに基づいて作成された Hidden Markov Model による認識手法）を想定した。

この認識手法では、単語モデルは対応する音韻モデルを連結して作成される。また、各音韻モデルは自己ループを持つ2状態あるいは3状態で構成される。

エラーシュミレーションは、単語を構成す

るモデルの状態数の総計に注目し、この総計をもとに誤認識時の出力を作成することとした。

そこで、まず最初にすべての認識語彙について、各単語を構成するのに必要な状態数を求め、次に、与えられた誤認識率になるように、乱数により誤認識を発生させる単語を選択した。

選択された単語の状態数をもとに、誤認識結果となる単語を選択した。

音声認識における認識率は、大量の評価データに対して認識を行った結果であることを考慮して、今回の実験では150対話全体での誤認識率が与えられた誤認識率（想定誤認識率）となるようにした。そのため、各対話ごとで見た誤認識率は、想定誤認識率とは異なっている。

#### 5. 実験結果および考察

以上のことにより、評価データセット75対話を対象に行った実験結果を、図1、図2に示す。

図1は単一要求、複数要求のそれについて、シミュレートした誤認識率と再現率の関係を示している。再現率とは、システムが

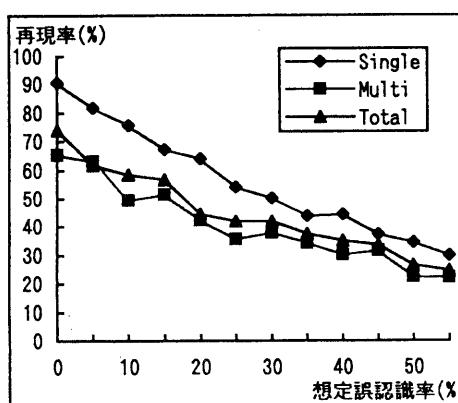


図1. 誤認識率と再現率の関係

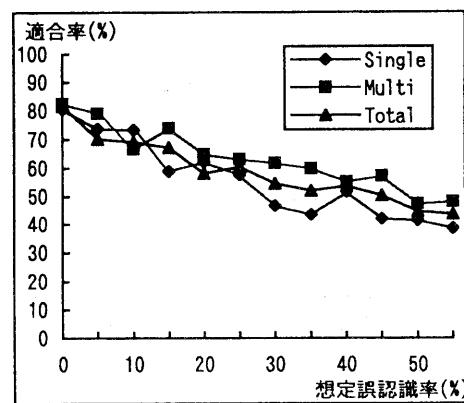


図2. 誤認識率と適合率の関係

要約項目として、正しく抽出できた項目の割合を表しており、システム理解度の指標となる。

図2は各要求に対する適合率を表したものである。適合率とは、抽出できた項目のうち、要約内容が正しいものの割合を表しており、システム正確度の指標となる。

これらの再現率、適合率のグラフについて、線形回帰分析を行い、その傾きを求める結果となつた。

また、9つの要約項目ごとの再現率を図3に、線形回帰分析によって求めた傾きを表5に示す。

実験結果を全体的に見ると、0~55%まで想定誤認識率を線形に変化させていった場合に、MIMIの性能（再現率・適合率）も、同じように線形に低下しており、性能が極端に悪くなるような境界点は存在しない。さらに、その傾きに注目した場合、急峻な傾きを示しているものがない。

これらことは、音声認識の性能の劣化が、システムの性能劣化に直接的な影響を及ぼしているが、対話要約というMIMIの基本性能は維持されていることを表している。

適合率に注目すると、誤認識率の増加に比較して適合率の低下が少なく、誤認識の影響

表4. 線形回帰分析（傾き）

	再現率	適合率
単一要求	-1.08	-0.75
複数要求	-0.77	-0.59
全体	-0.81	-0.61

を受けづらさを示している。これは、特に複数要求の対話において顕著に表れている。

劣化性能を項目別に見た場合では、処理内容・OHP使用の有無・処理の受理などの性能が良く、日付・時間帯などの性能が悪くなっている。これは、ClientとClerkのネゴシエーションによって、要求内容に変更が起こりうるような項目では、そうでない項目に比べて劣化が大きくなることを表していると考えられる。

## 6. まとめ

本報告では、従来より我々が構築してきた対話要約システムMIMIにおいて、入力となる音声認識結果の誤認識がシステムに与える影響を、認識誤りをシミュレーションすることにより実験を行い、MIMIの認識誤りに対する robust性を検証した。

今後は、実際の音声認識結果を使用した実験による検証を行う必要がある。

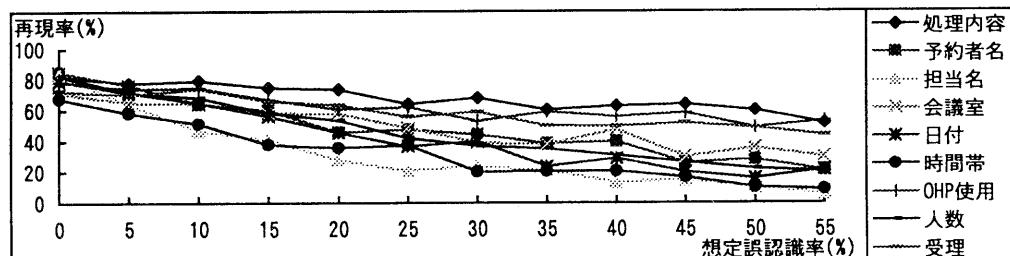


図3. 要約項目ごとの再現率

表5. 項目別の傾き

	処理内容	予約者名	担当名	会議室	日付	時間帯	OHP使用	人数	受理
傾き	-0.49	-1.08	-1.20	-0.76	-1.16	-1.04	-0.48	-1.14	-0.55

## 7. 参考文献.

- [1] M.Kameyama and I.Arima "Coping with aboutness complexity in information extraction from spoken dialogues." In the Proceedings of the ICSLP-1994
- [2] M.Kameyama and I.Arima "A minimalist approach to information extraction from spoken dialogues." In the Proceedings of the ISSD-1993
- [3] D.Appelt, J.Hobbs, J.Bear, D.Israel and M.Tyson, "FASTUS: A Finite state Processor for Information Extraction from Real World Text" In the Proceedings of the IJCAI-1993

付録 1. 対話例

A:Client, B:Clerk	
B:はい、こちら会議室予約係です。	A:で、時間はですね。
A:開発企画の安部です。	B:はい。
B:安部さんですか。	A:1時から5時、13時から17時がいいんですけれど。
A:えーと、来週、会議室を予約したいんですが。	B:1時から5時。
B:はい。えーとどんな会議室がよろしいんでしょうか。	A:はい。
A:えーとですね、プレゼンテーションルームか。	B:わかりました少々お待ち下さい。
B:はい。	A:はい。
A:第5会議室か第6会議室のどちらを使いたいです。	B:そうですね、そうしましたら、第5会議室は月曜か、金曜があいてますね。
B:第5か第6。	A:はい。
A:はい。	B:第6会議室でしたら、月曜と火曜が空いてます。
B:ですね。えーと人数はどのくらいですか。	A:はい。
A:人数は25人です。	B:あと、プレゼンテーションルームでしたら木曜と金曜が空いてますね。
B:25人。	A:あ、じゃあですね、月曜日、第5会議室でお願いします。
A:で、ひにちなんですが、いつでもけっこうです。	B:あ、わかりました。月曜第5会議室を1時から5時という事で。
B:あ、来週ならいつでも。	A:はい。そうです。
A:はい。	B:わかりました。それでは予約します。
B:いいんですね。	

付録 2. 要約例

処理内容	予約
予約者氏名	安部
予約者担当名	開発企画
会議室	第5会議室
日付	8月31日
時間帯	13:00~15:00
会議人数	25人
OHP使用	不明
処理要求の受理	受理