

表現の特徴を用いた議事録自動生成システムの性能評価

阿部里美

Satomi Abe

sato@media.eng.hokudai.ac.jp

荒木健治

Kenji Araki

araki@media.eng.hokudai.ac.jp

北海道大学大学院 情報科学研究科

Graduate School of Information Science and Technology Hokkaido University

1. はじめに

議事録とは、会議での決定事項、話し合った内容、出された意見等をまとめたものであり、会議の記録を残すためにはその作成が必須である。しかし全ての発言の中から重要な部分を見極め、それをまとめあげる作業には多大な労力と時間が要される。このため、人間の手間をかけずに議事録を自動的に生成することが望まれている。そこで、本稿では議事録を自動的に作成するシステムを提案する。

議事録は、会議の会話履歴という文書から重要な箇所を見つけ出し、それをまとめて提示したものであると言う事ができるが、これは、ある文書から重要な部分を抜き出してまとめる自動要約の一種とみなすことができる。近年、自動要約の分野では、自然な要約や複数テキストを対象とした要約に注目が集まっているが、多くの研究の基礎にあるのが重要文抽出である[5]。重要文抽出とは、文書中から重要と考えられる文を抜き出す技術であり、重要な部分を絞り込むための処理としても利用される。そこで本研究においてもこのような自動要約の手法にならない、まず重要文の抽出を行い、その後抽出した文を適切な形に変換することで、議事録を生成するというアプローチをとった。

自動要約における重要文抽出は、文の長さや文章中での出現位置といった素性を基に、文の重要度を決定するのが一般的だが[4]、本研究ではそれに加えて、話者が誰かといった会話特有の情報や、重要文によく現れる表現パターンを重要文抽出ルールとして自動的に獲得し、それらを用いて重要文の抽出を行う。

議事録自動生成の他研究には、会議コーパス[1]の会議会話履歴から議事録を生成する研究[2]や、チャット会議ログから議事録を生成する研究[6][7]等があるが、その多くは一般的な議事録のようにまとめた形ではなく、ほぼ抜き出した文そのままに近い形で重要文を提示している。我々がこれまで開発してきたシステムにおいても、重要文をそのままの形で提示してきたが[3]、本稿で提案するシステムは、抽出した重要文の文体を変換し、文の表す内容によってまとめ、より議事録の形に近づけた状態で出力を行う。

以下では、議事録自動生成システムの概要について述べ、実際の会議の書き起こしを用いてシステムの性能評価実験を行った結果について報告する。

2. システムの概要

本システムは、まず会議の会話履歴から重要文を抽出し、次に抽出した文の文体を適切な形に変換することで議事録を生成する。本章では、重要文抽出処理及び文体変換処理の概要について述べる。

2.1. 重要文抽出

重要文抽出の大まかな流れとしては、システムはまず会議の会話を書き起こしたものを入力として受け取る。受け取ったデータに対して、茶釜[9]を用いて形態素解析を行い、品詞情報を付加する。次に、重要文抽出ルールを用いて各文の重要度を計算し、重要度の大きな文を重要文として抽出する(詳細は 2.1.3 で述べる)。システムが抽出した文が重要文として正しいか否か、抽出すべきであるのに抽出できなかった文はないかを学習用の正解データ(正解重要文)と比較することで評価を行う。続いて、正解重要文と既に獲得済みのルールを基に、新たなルールを生成する(詳細は 2.1.2 で述べる)。最後に、正しい重要文を抽出するのに用いたルールの尤度を増大させ、誤った文を抽出するのに用いたルールの尤度を減少させることにより、ルールの尤度を更新する。

以上のように、データの入力、抽出、学習、フィードバック、といった処理を繰り返すことでシステムは学習を進め、性能が向上していくことになる。また、本システムにおいて、重要文は1文だけでは情報が不足することが多いため、隣接する2文を単位として扱う。

以下ではシステムが使用する重要文抽出ルール、及び各処理の詳細について述べる。より細かな内容については、我々のこれまでの研究である文献[3]に譲る。

2.1.1. 重要文抽出ルール

重要文抽出ルールには、次の2種類のルールを用いる。

素性ルール

以下に示す5つの素性をそれぞれ0, 1, 2, ...といった値で数値化する。それら5つの数字を並べたものを文の状態パラメータとし、これを素性ルールとして用いる。

・ 全文中での出現位置

会話履歴全体の中での文の出現位置
0:序盤 1:中盤 2:終盤 (全体を3等分する)

・ 一つの発言中での出現位置

話し始めてから別の人が発言するまでの発言を1回の発言とすると、1回の発言中での文の出現位置
0:1文目 1:序盤 2:中盤 3:終盤 (全体を3等分)

・ tf*idf

次式で求められる各単語のtf*idf値の和

$$tf * idf(t, d) = \frac{\text{文書dに単語tが出現する回数}}{\text{文書中の全ての単語の出現回数}}$$

$$* \left\{ \log \left(\frac{\text{全文書数}}{\text{単語tが出現する文書数}} \right) + 1 \right\}$$

0:0.2以下 1:0.4以下 2:0.6以下 3:0.8以下
4:1.0以下 5:1.2以下 6:1.4以下 7:1.6以下
8:1.8以下 9:それ以上

- ・ 文の長さ
含まれる単語の数
0:0~40 語 1:41~80 語 2:81 語以上
- ・ 話者情報
その発言の話者が誰か
0:その他参加者 1:議長

パターンルール

「@ で/助詞 よろしい/形容詞 です/助動詞 か/助詞 . /記号」といった表現のパターンからなる (@は変数)。

2.1.2. 学習処理部

学習処理部では、正解重要文から 2.1.1 で述べた 2 種類のルールを重要文抽出ルールとして獲得する。

素性ルールの獲得

重要文に対して、2.1.1 で示す文の状態パラメータを求め、それを新たな素性ルールとして獲得する。

パターンルールの獲得

正解重要文そのもの、及び既に獲得したルールと正解重要文との組から、字面・品詞共に共通する部分を残り、異なっている部分を変数に置き換えたものを、新たなルールとして獲得する。例えば、@を変数、アルファベットを品詞情報付きの単語としたとき、正解の重要文「a b c d e f g h」と既に獲得したルール「@ b k e f g h」の組の場合、「@ b @ e f g h」というパターンが生成される。生成したパターン、及び正解重要文そのものをルールとして獲得する。

ここで、各正解重要文には、表 1 に示す 8 種類のカテゴリのうちのいずれに属する文か、といった情報を付加している。ルールの生成基である正解重要文のカテゴリが、生成したルールの属するカテゴリとなり、ルールはカテゴリ毎に獲得される。カテゴリに分類することで、より適切なルールの生成、重要文の抽出が行われる[3]。

2.1.3. 抽出処理部

重要文の抽出は文の重要度を基に行う。重要度の計算は、ルールを適用した際のスコアを基に計算する。スコア計算式は文献[3]で示したものと同等のものを用いるため、本稿では詳細を省略し、概要のみを述べる。

素性ルールのスコア

その素性ルールの尤度はどのくらいか、素性ルールと対象文の各素性がどのくらい一致しているか、そして各素性が他の素性に比べてどの程度重要か、を基にスコアを計算する。ルールの尤度とは、ルールが適用されたとき及びルールが獲得されたときにカウントされる値（出現回数）、そのルールを適用して正しい重要文が抽出されたときにカウントされる値（正適用回数）、そのルールを適用して誤った文を抽出したときにカウントされる値（誤適用回数）から求められる。

パターンルールのスコア

対象文にマッチしたルールについて、素性ルールの尤度と同様にして尤度を求め、それをパターンルールのスコアとする。

以上のようにして全てのルールのスコアを求める。そして閾値を越えたルールについて、カテゴリ毎に、パターンルールのスコアの和と素性ルールのスコアの和の重み付き和を求め、最も大きな値となったカテゴリのルールを適用し、そのスコアを文の重要度とする。

2.2. 文体変換処理

2.1 の重要文抽出によって抽出された文を、議事録として適切な文体に変換する。システムが抽出した文は、その文の抽出に用いたルールと同じカテゴリに属するものとみなすことができる。そこで、そのカテゴリを利用して、表 1 に示すテンプレートの対応表に従い、文体を変換する。表 1 の"~"に入れるべき部分を抽出した文の中から特定し、その部分を抜き出してテンプレートに当てはめることで文体変換を行う。

処理の流れは、まず抽出箇所特定ルールを用いて、文中のどの部分を抜き出すべきかを特定する。そして、抜き出した部分をテンプレートに当てはめることで文体変換を行う。全ての文の文体変換が終了したら、それらを表 1 のカテゴリの順に並べ替え、議事録として提示する。抽出が終わると、抽出箇所特定ルールの学習に入り、正解重要文と正解変換文の組から新たなルールを獲得し、さらに、新しく獲得したルールと既に獲得済みのルールとの組からより一般化されたルールを生成する。その後、2.1 の重要文抽出と同様、フィードバック処理により正しい箇所を抜き出すのに用いたルールの尤度を上げ、誤った箇所を抜き出すのに用いたルールの尤度を下げることによってルールの尤度を更新する。

以下ではルール、及び各処理の概要について述べる。

2.2.1. 抽出箇所特定ルール

抽出箇所特定ルールは、抽出すべき箇所の前後 3 単語の情報からなる。例えば「<文頭> それでは/接、/記号 ~ と/助 いう/動 こと/名」というルールがあった場合、これは「<文頭> それでは/接、/記号」と「と/助 いう/動 こと/名」に挟まれる部分があれば、それが抽出箇所である、ということを意味している。

2.2.3. 学習処理部

本システムは学習によりルールを獲得するため、正解データが必要となる。そこで予め、抽出すべき重要文である正解重要文と、その文を文体変換したものである正解変換文の対を用意し、それらを用いてルールの獲得を行う。手順は次のとおりである。

表 1 カテゴリとテンプレートの対応

カテゴリ	テンプレート
前回までの話題	(そのまま表示)
報告	~について~から報告があった。
説明	~について~から説明があった。
審議内容	~について審議した。
確認内容	~について確認した。
決定事項	~こととなった。
意見	(そのまま表示)
今後の予定	~こととなった。

1. 正解変換文と正解重要文の組に共通して現れる単語を基に、重要文のどの部分を抜き出しているかを判断する。ただし、助詞や助動詞等の付属語は頻繁に出現するため、それらを共通部分とみなすと抽出範囲の特定を大きく誤ってしまう可能性がある。そこで、自立語のみを共通部分の判断に用いる。
2. 抜き出された箇所前後 3 単語を重要文から抽出し、カテゴリを付加した上でルールとして獲得する。
3. 1～2の処理によって獲得したルールと、既に獲得された同カテゴリのルールとの組から、共通部分、差異部分を基に、より一般化されたルールを獲得する。ただし、共通部分、差異部分の判断は字面、品詞に分けて行う。

例) 獲得したルール：

ませ/助動 ん/助 . /記 ～ と/助 言う/動 こと/名

既に獲得済みのルール：

, /記 それでは/接 , /記 ～ と/助 言う/動 こと/名

生成されるルール：“@”は変数を表す)

@1 @2 @3/記 ～ と/助 言う/動 こと/名

2.2.2. 文体変換処理部

システムは対象の文に対して、全ての抽出箇所特定ルールとのマッチングを行い、マッチしたルールを用いて抽出すべき箇所の特定を行う。マッチしたルールが複数あった場合は、以下の優先順位に基づいて最も優先度の高いルールを適用し、抽出する。

優先順位

1. 字面・品詞共に固定の単語(“と/助”等)の数が多
2. 品詞のみが固定の単語(“@/助”等)の数が多
3. 抽出される部分の tf*idf が大きい
4. 正適用回数が多い
5. 誤適用回数が小さい
6. 出現回数が多い

また、文の簡潔化や、テンプレートへの不適切な接続を防ぐため、抽出した箇所に対し次の処理を加える。

- ・ 文を句読点で分割し、tf*idf の平均値が閾値(ここでは 0.01 とする)よりも小さい部分を削除する
 - ・ テンプレートが「～について」の場合、抽出箇所の最後尾に名詞が出現するまで抽出範囲を広げる
 - ・ テンプレートが「～こととなった」の場合、抽出箇所の最後尾に動詞・または名詞が出現するまで抽出範囲を広げ、活用形を原形に変換する
- 最後尾が名詞の場合は「とする」を付加する

2.2.4. フィードバック部

重要文抽出ルールと同様に、抽出箇所特定ルールについてもそれぞれ 2.1.3 で述べたような出現回数、正適用回数、誤適用回数を保持している。これらの値を更新することで、2.2.2 で述べたルールの優先順位を更新する。

3. 実験

システムの性能を評価するため、18 会議分の会話書き起こしデータ[8]を用いて実験を行った。実験に用いたデータは 1 会議平均約 381 文であり、全データの総文字数は 376,686 文字(1 会議平均 20,927 文字)であった。

実験は、素性ルール、パターンルールいずれも空の状態から開始した。学習に使用する正解重要文は、文献[8]において会話データと共に公開されている議事録を基に、対応する文を正解重要文として第一著者が定めた。ただし、広い範囲の文章をまとめあげたものや、会話中に直接述べられていない内容を含んだものは重要文を定義することができないため、今回は扱っていない。正解重要文は 1 会議平均 4.6*2 文(2 文単位で扱うため)である。また正解変換文についても、会話データと共に公開されている議事録をもとに、第一著者が定めた。

重要文抽出で用いる閾値等は、様々な値で実験を行ったところ、最も良好な結果を示した値を用いた(重要度閾値=1.9, 素性ルール閾値=0.6, パターンルール閾値=0.5, 素性ルールに対するパターンルールの重み=0.4)。

4. 結果と考察

重要文抽出の評価には再現率と適合率の両方を考慮した値である F 値を用いた。3 会議毎の F 値平均の推移の様子を図 1 に示す。横軸はシステムに入力した会議のデータの数である。実験が進むにつれ、F 値が上昇傾向にあることがわかる。最終 3 データにおける平均 F 値は 0.36 であった。チャット会議ログからの議事録生成の試作検討を行っている小林の研究[6]の F 値は 0.84 を得ており、本システムはそれに比べ、かなり低い値である。しかし、小林[6]の要約率が 31%であるのに対し、本システムは最終 3 データの平均要約率が 1.3%とかなり小さく、また小林[6]が少数のデータ(7 会議文)から人手によりルールを作成し、クローズドデータで実験を行っているのに対し、本システムはルールを自動的に学習し、オープンデータで実験を行っている。これらの点を考慮すると、比較的良好的な結果であると考えられる。

システムが生成した議事録、生成目標とする議事録の例を図 2、図 3 に示す。下線の文は重要文として正しく抽出できた文である。各文とも意味が大きく変わってしまう等の問題は見られず、重要文そのままに比べ、短くまとめることができている。しかし、いずれの文もやや冗長になってしまっており、議事録として自然な形であるとは言い難い。実下線の文については「せっかく集計表をつくっていただいていますので、これを見ながら」をより短くし、「集計表を見ながら」と短縮した方がより簡潔である。本システムは現在、文内短縮を行って

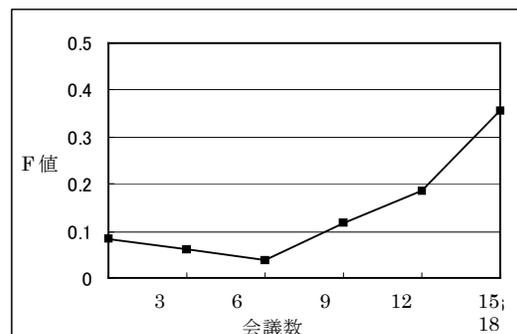


図 1 3 会議毎の F 値平均の推移

<p>■ 決定事項</p> <ul style="list-style-type: none"> ・(前回宿題でお出して、○×をつけていただいて、集計したA3の資料が届いていると思います。前回の議事録もあわせて届けられましたので、皆さん、前回の議論もトレースしていただけたかと思う)こととなった。 ・(せっかく集計表をつくっていただいていますので、これを見ながら、どうやって指標を選定していくかという話を先にさせる)こととなった。 ・(これも特になくていいのではないですか。それから、調査研究の推進ということで、117も残しておいてよろしいのではないかと思う)こととなった。 ・(一通り来ましたが、次は総会になってしまいますので、事務局の方で整理して、総会以後の話を要約して、報告書の文案をつくらせていただいて、事前に配付する)こととなった。

図2 システムが生成した議事録

<p>■ 前回までの内容</p> <p>これまでの部会で挙げた123の評価項目に対し、評価項目として採用すべきか、各委員が項目ごとに○×を付加した。 <u>評価項目について前回集計した表を見ながら、どのように指標を選定していくかについて審議した。</u></p> <p>■ 決定事項</p> <p><u>抽出された指標項目について、事務局で整理し、報告書の文案をつくり、各委員に総会前に一度配布することとなった。</u></p> <p>■ 今後の予定</p> <p>3月24日の総会において報告書案の了承を求めることとなった。</p>
--

図3 生成目標とする議事録

ないため、このような冗長性に対処することができない。今後、文内短縮の方法についても検討する必要がある。

また、図2の1文目と3文目のテンプレートへの抽出箇所は2文にまたがってしまっており、文として不適切である。重要文は1文だけでは情報が不足することがあるため、2文を単位として抽出を行っているが、1文だけで十分な情報を持つ場合も存在する。このような場合、予め不要な文の削除を行うことで、不適切な箇所の抽出を防ぐことができる可能性がある。tf*idfによって文が重要な情報を含むか否かを判定するという方法や、2文を繋ぎ1文化する等の処理の追加を検討している。

この他、重要文抽出時点での問題として、図1に示すように学習が進むにつれて精度が上がってきてはいるものの、正解重要文の数自体が平均4.6文と少ないこともあり、正しく抽出できた重要文の絶対数がかなり小さくなっている。文体変換部のフィードバック処理では、正しく抽出できた重要文に対してのみ、使用したルールの正否を判断するため、正しく抽出できた重要文の数が少ないことが原因で、正適用回数や誤適用回数の更新によるフィードバックがうまく機能しないといった問題がある。十分な量の学習データを確保して実験を行うことにより、ルールの優先順位が適切になることを確認する必要がある。また、抽出した文のカテゴリが誤っていたために、図2の2文目のように審議内容を表す文が、決定事項として認識されてしまった場合、文の意味としてはそれほど不自然ではないが、議事録としてはやや不自然な形になってしまう。文体を適切に変換するためにも、重要文抽出やカテゴリの精度の向上が必須である。

5. まとめ

本稿では、会話履歴から議事録を自動的に生成するシステムを提案し、評価実験を行った。本システムは実例から学習を行うため、議事録作成者の習性に沿った学習が行われるという特徴がある。生成された議事録の精度

にはまだ課題もあるが、今後は十分な量の実験データを確保し、実験を行う予定である。また、会話は音声で交わされるものであるため、音声認識を行った結果からの議事録生成についても今後検討する予定である。

参考文献

- [1] A. Janin, D. Baron, J. Edwards, D. Ellis, D. Gelbart, N. Morgan, B. Peskin, T. Pfau, E. Shriberg, A. Stolcke, and C. Wooters. The ICSI meeting corpus. In Proc. IEEE ICASSP, pp. 364-367, 2003.
- [2] G. Murray, S. Renals, J. Carletta, and J. Moore. Evaluating automatic summaries of meeting recordings. In ACL 2005 Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization, pp. 33-40, 2005.
- [3] 阿部 里美. 汎用的要約システムの議事録に対する適応性能の評価. 平成17年度電気・情報関係学会北海道支部連合大会講演論文集, 179, 2005.
- [4] 奥村 学, 難波 英嗣. テキスト自動要約に関する研究動向(巻頭言に代えて). 自然言語処理, Vol. 6, No. 6, pp. 1-26, 1999.
- [5] 奥村 学, 難波 英嗣. テキスト自動要約に関する最近の話題. 自然言語処理, Vol. 9, No. 4, pp. 97-116, 2002.
- [6] 小林 竜己. チャット会議ログからの議事録自動生成—領域分析と試作検討—. 人工知能学会研究会資料, SIG-SLUD-A202-02, pp. 7-12, 2002.
- [7] 小林 竜己. 談話の局所・中位構造を利用したチャット会議ログからの議事録自動生成. 人工知能学会研究会資料, SIG-SLUD-A203-05, pp. 29-34, 2003.
- [8] 札幌市男女共同参画施策ホームページ. <http://www.city.sapporo.jp/shimin/danjo/>
- [9] 茶筌. <http://chasen.aist-nara.ac.jp/hiki/ChaSen/>