マルチメディアビッグデータ解析のための音声中のプライバシー情報の秘匿化

Privacy Anonymization in speech data for multimedia big data analysis

中村　哲（ Satoshi NAKAMURA,　Dr. Eng. ）

奈良先端科学技術大学院大学　先端科学技術研究科

**　情報科学領域　　教授**

(Professor, Information Science Division, Graduate School of Science and Technology,

Nara Institute of Science and Technology、)

IEEE、 ISCA、 電子情報通信学会、情報処理学会、日本音響学会　他

受賞：文部科学大臣表彰 (2010 年) 、総務大臣表彰(2011 年)、情報処理学会山下記念研究賞(2007 年)、喜安記念業績賞(2008 年)、電子情報通信学会論文賞(2015)、日本音響学会栗屋学術奨励賞(1992 年)、日本音響学会技術開発賞(2008 年)、Antonio Zampolli 賞 (2012 年) 、ドコモモバイルサイエンス賞(2007 年)、IEEE Fellow(2016 年)、情報処理学会フェロー(2018 年)他

著書：音響サイエンスシリーズ 18　音声言語の自動翻訳　コンピュータによる自動翻訳を目指して、 コロナ社出版 (2018 年) 、Spoken Dialogue Systems Technology and Design、 Springer 社出版 (2011 年) 、 多言語音声翻訳技術の最前線　朝日出版 (2016 年)、音声言語処理 -コーパスに基づく音声言語処理-、 森北出版、 1997 、音声・音情報のディジタル信号処理、昭晃堂、1998

研究専門分野：情報学、音声言語処理学

あらまし

インターネットの普及によるソーシャルデータ、マルチメディアデータ、センサーの普及による地震、気象などのセンシングデータや、DNA 等の科学データの増大により、ゼタバイトを上回るスケールの超大規模データが出現している。最近は、IoT データが注目され始めているが、これまで、主に扱われてきたデータは、購買履歴、インターネットのソーシャルデータなども含めてオープンなものが殆どであった。一方で、個人性を含むテキストデータの解析や、コールセンタの対話音声ログデータの解析のニーズが高まっている。テキストだけでなく、音声データが利用できれば、音声認識で誤った部分の書き起し、感情の分析、サービスの質の分析など、より高次の分析が可能となる。本研究は、オープンでないマルチメディアビッグデータの解析を進めるため、テキスト、音声情報の中のプライバシー部分について一貫性を維持しながら、秘匿化を行う技術の確立を目指す。

１．研究の目的

本研究では、オープンでないマルチメディアビッグデータの解析を進めるため、非構造化テキストと音声チャネルの中のプライバシー部分について一貫性を維持しながら、同一話者の異なる音声に置換することで秘匿化を行うための要素技術の確立を目指した。具体的には、非構造化データとしての自然言語文の匿名化、および、ニューラルネットワークを用いた文字レベルの音声認識による置換すべき固有名詞認識の容易化について検討を行った。

２．研究の背景

インターネットの普及によるソーシャルデータ、マルチメディアデータ、センサーの普及による地震、気象などのセンシングデータや、DNA 等の科学データの増大により、ゼタバイトを上回るスケールの超大規模データが出現している。このビッグデータは、3V と言われる Volume、Velocity、Variety の点において従来技術で扱えない規模に達し、その成長速度でみれば、昭和 35 年から 1 兆倍に成長し、加速度的に大きくなっている。この急速な変化は解析のための技術や計算システムの立ち後れを引き起こしており、これらの技術開発は喫緊の課題である。今後、データを新しい重要な資源と捉え、それら資源の確保とその加工技術、すなわちデータの高度な分析・解析技術を世界に先駆けて確立することが極めて重要となる。ビッグデータの解析としては、大量の顧客データに基づき、顧客分析、トラブル分析、販売予測、あるいは、E-commerce サイトでの自動推薦などに利用されており、その重要性が認識されつつある。

一方で、コールセンタの対話音声ログデータなどの非構造テキストデータ解析、ムービーファイルの音声の解析などのニーズが高まっている。このようなデータが利用できれば、データの構造化の手間が省け、感情の分析、サービスの質の分析など、より高次の分析が可能となる。また、これまで、主に扱われてきたビ

マルチメディアビッグデータ解析のための音声中のプライバシー情報の秘匿化

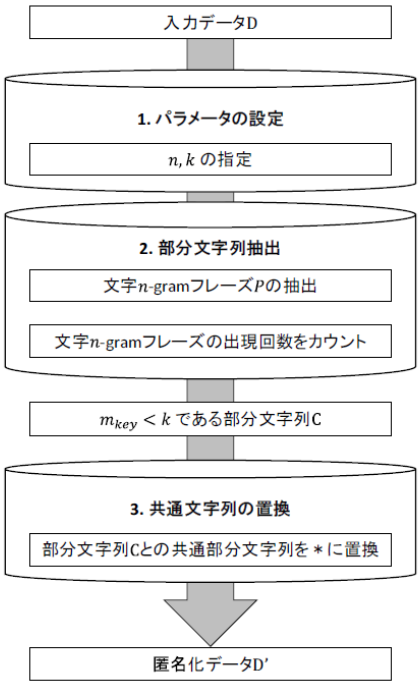
Privacy Anonymization in speech data for multimedia big data analysis

ッグデータは、インターネットのソーシャルデータなどオープンなものが多かったが、今後は個人情報を含むデータの取り扱いも必要になる。

３. 研究の方法

講演・会話音声ログデータ、ムービーファイルの音声データには、人の名前、住所、利用者のＩＤ、カード番号、電話番号などのプライバシーや権利に関わる情報が含まれている。これまで、当該発話部分を雑音で置き換える処理も試みられてきたが、たとえば名前の一貫性を保ったりすることはできなかった。本来なら、秘匿すべき性質の固有名の識別、あるいは、複数の人が登場して、それらの名前を順番にＡ氏、Ｂ氏と置換し、複数回出現しても一貫性を保って、Ａ氏、Ｂ氏と置き換えること等が必要である。そこで、本システムを構成するために必要な音声認識（KALDI）、形態素解析（Mecab）、秘匿情報の抽出、声質変換、音声合成（Openjtalk）の要素技術の整備、構築を行った。

まず、秘匿情報の抽出と、N-gram に基づく部分文字列単位の秘匿化技術の検討を進めた。



予備実験として、福岡市の事業所データを用い、提案法である N-gram に基づく k-匿名化法について、匿名化率、文字匿名化率を用いて人手評価を行った。匿名化アルゴリズムを図１に示す。

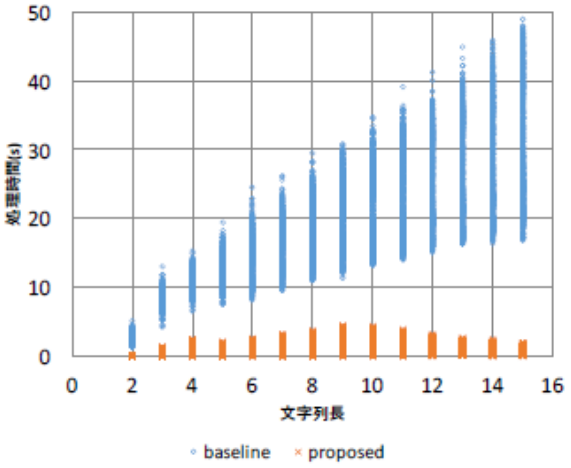


図２　処理時間の散布図

表１　処理時間の比較

	Baseline	Proposed	差	高速化率
平均時間	18.1063	0.9968	17.1095	0.0541
最大時間	49.0515 (15 文字)	4.5296 (9 文字)	15.8004 (15 文字)	0.3173 (9 文字)
最小時間	1.3129 (2 文字)	0.0007 (2 文字)	5.1386 (2 文字)	0.0003 (2 文字)
STD	3.9819	0.5649	0.1429	0.0333

この結果、全体の結果として、特定率は平均 25%、事業内容の理解ができているのは 62%であった。このことから、全体の情報の理解は保持しつつ、匿名化を実現できていることが判明した。

さらに、匿名化する文字列（黒塗りパターン）が最小になるように匿名化箇所を選定した。しかし、実際には、k 匿名化算出には時間がかかるという問題がある。wi の文字列長が l のとき、匿名化パターンは 2l-1 であり、2l-1 回 k 匿名化を計算する必要がある。そこで、本手法では転置索引を用いて k 匿名化を算出するための探索範囲を限定した。さらに、集合論も用いて探索範囲を限定し計算量を削減した。本手法の利点

は、参照リストを用いた従来の全文字匿名化と比較して残存文字数の減少を抑制できる点である。本手法は部分文字列の黒塗りパターンそれぞれについて同一データの出現回数 n を算出する必要がある。最適な黒塗りパターンを選ぶためには、各黒塗りパターンの数 n を記録しておく必要がある。そのため、本手法では黒塗りパターンと n を記録する匿名化辞書を構築する。ただし、n の算出には時間がかかるため、匿名化辞書構築には時間がかかる。そのため、本手法では転置索引と集合論の考えを用いて n の算出を高速化する。

図 2 は文字列長ごとの処理時間の散布図である。すべての項目について、提案手法は、高速化を行わない匿名化法である従来法（baseline）より処理時間が速かった。表 1 に、従来法と提案手法の平均処理時間、従来法と提案手法の処理時間の差、高速化率（提案手法の処理時間と従来法の処理時間）を示している。処理時間について従来法の処理時間に対して平均的に 17.1 秒速く処理ができており、平均 5.4%に削減することができた[1、2、3]。実際にテキストの匿名化を行う時間については、1000 文字のテキストを一つのパラメータを設定して匿名化を行ったとき、0.5 秒以下で処理を行うことが可能となった。

次に、音声認識について、文字単位の認識を行うための、ニューラルネットワークを用いた音声認識の高度化を進めた。これまでは、言語モデルを含んだニューラルネットを用いて計算される音素の事後確率を基に隠れマルコフモデルと単語列の連鎖確率（N-gram）により単語列を計算する音声認識法が主流であった。しかし、秘匿すべき文字列は固有名詞である可能性が高く、従来の辞書を事前に準備した単語単位の音声認識では秘匿すべき文字列を正しく抽出できないため、文字を出力するニューラルネットワークの手法について研究を進めた。さらに、波形からスペクトログラムを抽出していたが、この部分もニューラルネットワークに置き換え転移学習を行うことにより波形からの特徴抽出を可能とし、全体として、注意機構付きの双方向性 LSTM(Long Short Term Memory)再帰型ニューラルネットワークを用いて音声認識を構成することで、波形から直接音素列を出力する音声認識システムを構成することに成功した。

英語の Wall Street Journal の音声に対して行った標準評価においても表 2 に示すように、性能的にこれまでの時系列型音声認識（CTC: Connectionist Temporal Classification）より高い性能であることが明らかとなった。

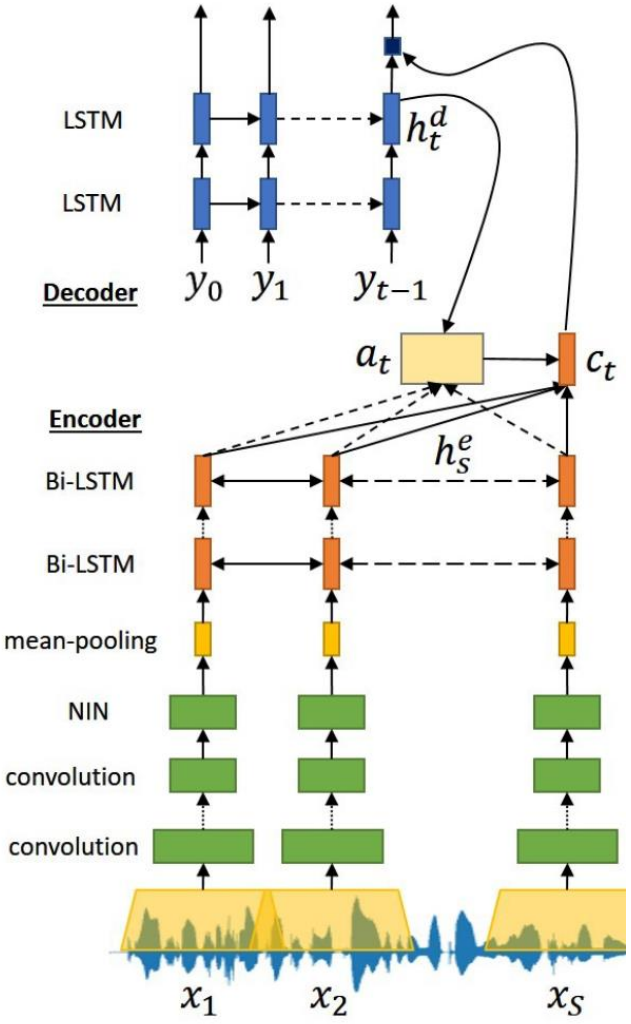


表 2 波形入力文字出力音声認識性能

方法	特徴量	文字誤り率
従来法（CTC）	スペクトル特徴量	8。97%
提案法	波形	6。54%

おわりに

非構造化テキスト、音声情報の中のプライバシー部分について一貫性を維持しながら、秘匿化を行う技術の要素技術の研究開発を進めた。文字単位の匿名化を実現する手法の研究、文字単位で音声認識を行う研究について新たな方法の開発と実験による有効性の確認を行った。今後は、自然言語文における固有表現抽出と匿名化を統合し、実際の非構造化テキストの匿名化および、音声認識結果の個人性への抽出、匿名化の研究開発を進めていきたい。

参考文献

[1] 前田若菜、鈴木優、吉野幸一郎、Graham Neubig、and 中村哲。ソーシャルメディアにおける非構造化テキストデータの k-匿名化によるプライバシー保護、第 162 回情報処理学会データベースシステム研究会、 2015 年 11 月　東京。

[2] 前田若菜、鈴木優、吉野幸一郎、中村哲。文脈からの推定を考慮した非構造化テキストデータの匿名化手法、第 15 回情報科学技術フォーラム、2016 年 9 月　富山。

[3] Wakana Maeda, Yu Suzuki, and Satoshi Nakamura. “Fast Text Anonymization using k-anonymity”, The 18th International Conference on Information Integration and Web-based Applications & Services, Singapore, Nov. 2016.

[4] Andros Tjandra, Sakriani Sakti, Satoshi Nakamura, “Attention-based wave2text with feature transfer learning”, IEEE ASRU 2017

この研究は、平成２６年度ＳＣＡＴ研究助成の対象として採用され、平成２７～２９年度に実施されたものです。