プライバシ保護のための音声中の人名除去手法の検討

Investivation of Elimination Method of Person Names from Speech to Ensure Privacy

土屋 雅稔 *, 遠藤 翔子 †, 藤井 康寿 †, 中川 聖一 †
豊橋技術科学大学
* 情報メディア基盤センター / † 情報工学系

1. はじめに

事故防止や犯罪防止を目的として，道路，駅構内，建物や街中など，人間が生活している実世界に多数のカメラやマイクなどのセンサを設置し，これらをネットワークで結合したユビキタスセンサネットワークの構築が進んでいる．目的指向で設計されたユビキタスセンサネットワークを設置し，そこから得られた情報のみを個々のユビキタスセンサネットワーク毎に利用するという閉じた運用は，適用分野や有用性が明確であるが，センサの設置者以外は，これらのセンサによってどのような情報が収集されているかを伺い知ることができない．このような性質は，一般市民のプライバシに対する脅威となる可能性がある．

美濃は，このようなユビキタスセンサネットワークの問題点を解決するため，センシングウェブという概念を提案している[1]．センシングウェブとは，**プライバシを考慮したセンサ**をネットワーク化し，センサの設置者だけでなく，一般市民の誰もがセンサを利用できるようにすることによって，ユビキタスセンサネットワークの問題点を解決しようとする提案である．例えば，画像センサ (カメラ) によって撮影された画像について，画像中に含まれる人物のみを対象として黒塗りを施すことにより，撮影地点に「誰が居るか」というプライバシ情報は隠蔽した上で,「何人の人間が居るか」「撮影地点の天候」などの情報のみが得られるプライバシを考慮した画像センサを実現できると考えれられる[2]．また，音センサ (マイクロホン) によって収録された音情報から，言語的，パラ言語的または非言語的に個人を特定できるような情報を取り除くことができれば，プライバシを考慮した音センサを実現できると考えられる[3]．

本稿では，プライバシを考慮した音センサを作成するため，音センサによって収録された音情報から，言語的に個人を特定できる情報を削除するための方法について検討する．後述する通り，言語的に個人を特定できる情報のみを直接取り扱うことは困難であるため，特に重要と考えられる人名の除去方法について検討する．

2. 音情報・言語情報に含まれるプライバシ情報

マイクロホンで収録される音は，プライバシ情報を扱う上では，背景音と音声の 2 つに大別できる．背景音には，どういう場所で話しているかというプライバシ情報は含まれるが，話者そのもののプライバシ情報は含まれないので，一般にプライバシ処理を行わずにそのまま用いることができる．音声も，複数人の音声が重なり合うと，雑音のような性質を持ち，誰が何を喋っているか個々に識別できなくなる[4]．プライバシ情報を扱う上では，このような音声についても，背景雑音と同等に扱うことができる．そのため，本稿では**音声**を，**誰が何を喋っているか識別可能な人の声**とする．

音声により表現され，伝達される情報は，

(1)　言語的情報 (linguistic information)
(2)　パラ言語的情報 (para-linguistic information)
(3)　非言語的情報 (non-linguistic information)

の 3 種類に大別できる[5]．パラ言語的情報は、音声に込められた意図，発話者の態度等，言語的情報以外で発声者が意識的に制御できる情報（主に韻律的特徴）を指す．非言語的情報は声質，性別などの発話者の個性や感情などの話者が意識的に制御しない情報である．すなわち，非言語的情報はその全てがプライバシ情報となり得る．

言語情報に含まれるプライバシ情報は，発話内容に含まれるプライバシ情報と，発話スタイルに含まれるプライバシ情報の 2 種類に大別することができる．発話内容には，発話者本人のプライバシ情報と会話中で言及されている人についてのプライバシ情報が含まれる．そのプライバシ情報は，以下のような情報全てである．

- 個人情報 (氏名，住民基本台帳番号など)
- 身体特徴 (年齢，身長，性別など)
- 居住地 (住所，電話番号など)
- 社会的地位 (職業，勤務地，会社名，学校名など)
- その他 (成績，収入など)

なお，上記のような情報であっても，広く一般に知られている情報 (例えば，有名野球選手の身長) はプライバシ情報とは言わない．しかし，その情報が知られている範囲を推測することは難しいため，安全のためには，上記のような情報を全てプライバシ情報として扱うのが望ましい．これらの多くは人名や地名などの固有表現であるが，通常の固有表現には含まれない表現も多い．これらを取り除くことが，プライバシ情報の保護のためには必要である．

ここで，音声認識処理とプライバシ情報の抽出が独立であると仮定すると，プライバシ情報の抽出は，音声認識結果として得られたテキストに対するチャンキングとして定式化できる．固有表現を対象とするチャンキングには，統計的機械学習に基づく手法が有効である．しかし，プライバシ情報の抽出を統計的機械学習に基づくチャンキングとして定式化する場合，新規プライバシ情報に対して十分な量のラベルありデータを用意することが事実上不可能であるという問題がある．この問題に対応するには，大量のラベルなしデータと少量のラベルありデータを併用する半教師あり学習が有効である．Miller らは，ラベルなしデータからクラス言語モデルを作成し，ラベルありデータにクラスタリング結果を素性として追加して学習を行う方法を提案している[6]．鈴木らは，Conditional Random Field のポテンシャル関数と学習方法を拡張し，ラベルありデータとラベルなしデータを直接に併用する手法を提案している[7]．筆者らは，ラベルなしデータから求めた類似語を素性として追加して学習を行う方法を提案している[8]．この方法については，次節で詳しく述べる．

発話スタイルに含まれるプライバシ情報とは，類義語の選択，機能語の選択，フィラーの挿入頻度などにおける地域的・個人的な癖である．典型的な例としては，ある地方の方言に特有の文末表現などが使用されていると，発話者の出身地を推測することが可能である．このようなプライバシ情報を保護するには，類義語や機能語の言い換えを行い，発話スタイルの個人性を取り除くことが必要である．鍛治らは，待遇表現に着目して話し言葉コーパスおよび書き言葉コーパスをウェブから自動収集し，コーパスに基づいて，書き言葉特有の用言を話し言葉でも用いられる用言に言い換える手法を提案している[9]．下岡らや秋田らは，統計的機械翻訳の枠組みに基づいて，発話スタイルを変換する手法を提案している[10),11]．この手法では，同一の内容に対する発話であるが，発話スタイルが異なるテキストからなるパラレルコーパスから変換モデルを学習し，その変換モデルを用いて話し言葉テキストを書き言葉テキストに整形したり，その逆の処理を行う．これらの方法により，話し言葉テキストを書き言葉テキストに変換すると，発話スタイルの個人性はかなり取り除かれる．

3. 未知語に対して頑健な固有表現抽出

増加し続けている固有表現を網羅したラベル付きコーパスを用意することは，非現実的である．筆者らは，固有表現ラベルは付与されていないものの，大量に利用可能なラベルなしコーパス (例えば，新聞記事データ) を併用して，ラベル付きコーパスに頻出しない (または，出現しない) 語を含む固有表現を頑健に抽出できる固有表現抽出法を提案している[8]．提案手法は 2 段階からなる．最初に，ラベル付きコーパスに頻出しない語に対して，ラベルなしコーパスから求めた周辺ベクトルに基づいて，固有表現ラベル付きコーパスに頻出し，かつ，その前後の文脈の出現が良く類似している語を対応付ける．次に，元々の語と，新たに対応付けた類似語の両方を素性として，従来からの機械学習手法を適用する．例えば，**図 1** の左端列のような文があり，この文に含まれる「石狩」という形態素が，ラベル付きコーパスには頻出しないとする．この形態素「石狩」に対して，良く類似していると同時に，ラベル付きコーパスに頻出する形態素「関東」を対応付ける．そして，元々の形態素「石狩」と，対応付けられた形態素「関東」の双方を素性として用いて機械学習を行う．以下，提案手法について説明する．

3.1 類似形態素の対応付け

ある形態素に対して，その前後の文脈の出現が最も類似している形態素を求める方法を，以下に述べる．

ある形態素 m の周辺ベクトル V_m は，あらゆる可能な unigram, bigram を次元とし，その unigram, bigram が形態素 m の直前直後に出現した頻度を各次元の値とするベクトルである．形式的には，次式によって定義される．

$$V_m = \begin{pmatrix} f(m,m_0), & \cdots & f(m,m_N), \\ f(m,m_0,m_0), & \cdots & f(m,m_N,m_N), \\ f(m_0,m), & \cdots & f(m_N,m), \\ f(m_0,m_0,m), & \cdots & f(m_N,m_N,m) \end{pmatrix},$$

ここで，$M \equiv \{m_0, m_1, \ldots, m_N\}$ は，ラベルなしコーパスに出現する全ての形態素からなる集合である．また，$f(m_i, m_j)$ は，形態素 m_i と形態素 m_j がラベルなしコーパスに連続して出現した頻度であり，$f(m_i, m_j, m_k)$ は，形態素 m_i, m_j, m_k がラベルなしコーパスに連続して出現した頻度である．

ラベル付きコーパスに頻出する形態素からなる集合を M_F とする．この時，ある非頻出形態素 $m_u \in M \cap \overline{M_F}$ に対して，周辺ベクトルの観点から最も類似した頻出形態素 $\hat{m_u}$ は，以下の式を解くことによって得られる．

$$\hat{m_u} = \underset{m \in M_F}{\operatorname{argmax}}\, sim(V_{m_u}, V_m), \qquad (1)$$

ベクトル間の類似度を求める関数 sim としては，様々なものが利用可能であるが，本稿では cosine 類似度を用いる．

3.2 素　　性

本稿では，i 番目の形態素 m_i に対する素性 F_i を，形態素素性 $MF(m_i)$，類似形態素素性 $SF(m_i)$，文字種素性 $CF(m_i)$ の 3 つ組として定義する．

$$F_i = \langle\, MF(m_i),\ SF(m_i),\ CF(m_i)\, \rangle$$

形態素素性 $MF(m_i)$ とは，その形態素 m_i の表層形

形態素素性 MF		類似形態素素性 SF		文字種素性	チャンク
表層形	品詞	表層形	品詞	CF	ラベル
今日	名詞–副詞可能	今日	名詞–副詞可能	$\langle 1,0,0,0,0,0\rangle$	O
の	助詞–連体化	の	助詞–連体化	$\langle 0,1,0,0,0,0\rangle$	O
石狩	名詞–固有名詞	**関東**	名詞–固有名詞	$\langle 1,0,0,0,0,0\rangle$	B-LOCATION
平野	名詞–一般	平野	名詞–一般	$\langle 1,0,0,0,0,0\rangle$	I-LOCATION
は	助詞–係助詞	は	助詞–係助詞	$\langle 0,1,0,0,0,0\rangle$	O
晴れ	名詞–一般	晴れ	名詞–一般	$\langle 1,1,0,0,0,0\rangle$	O

図 1 学習データの例

表 1 NHK コーパスに含まれる固有表現

種類	頻度 (%)
DATE	755 (19%)
LOCATION	1465 (36%)
MONEY	124 (3%)
ORGANIZATION	1056 (26%)
PERCENT	55 (1%)
PERSON	516 (13%)
TIME	101 (2%)
計	4072

と品詞の組である．類似形態素素性 $SF(m_i)$ は，形態素 m_i に対して最も類似した頻出形態素の形態素素性であり，次式のように定義される．

$$SF(m_i) = \begin{cases} MF(\hat{m}_i) & \text{if } m_i \in M \cap \overline{M_F} \\ MF(m_i) & \text{otherwise} \end{cases}, \quad (2)$$

$\hat{m}_i$ は，形態素 m_i に対して周辺ベクトルの観点で比較して最も良く似ていると同時に頻出する形態素であり，式 (1) によって求められる．文字種素性 $CF(m_i)$ は，6 個の 2 値のフラグからなる．フラグはそれぞれ，形態素 m_i の表層形が，漢字・平仮名・片仮名・アルファベット・数字・その他の文字を含むか否かを表す．

3.3 訓練条件

本稿における提案手法の訓練条件は，以下の通りである．固有表現ラベル付きコーパスとして，IREX ワークショップ実行委員会によって公開されているコーパス (以後，このコーパスを **IREX コーパス**と呼ぶ) を用いた．IREX コーパスは，1995 年 1 月 1 日から 1 月 10 日までの間に発行された 1,174 件の毎日新聞記事からなり，その記事中の 18,677 個所の固有表現がタグ付けされている．周辺ベクトルを求めるラベルなしコーパスとしては，毎日新聞データ (1993 年 ∼1995 年) を用いた．分量は，3.5M 文 ·140M 形態素である．また，頻出語集合 M_F を，IREX コーパスに 5 回以上出現した全ての形態素の集合と定義する．統計的固有表現抽出器の機械学習には，CRF++[⋆1] を用いた．

4. 実験および考察

音声情報としては，日本放送協会 (NHK) によって 1996 年 6 月 1 日から 12 日にかけて放映された「おはよう日本」などのニュース番組を収録したコーパス (以後，このコーパスを **NHK コーパス**と呼ぶ)[⋆2] を用いた．NHK コーパスの発話内容を人手で書き起こしたテキストに対して，ARTIFACT 以外の種類の固有表現について，大学院生 1 名が IREX コーパスと同一の基準で固有表現ラベル付けを行った結果を**表 1** に示す．

次に，SPOJUS(1 パストライグラム) を音声認識用デコーダとして用いて，NHK コーパスの音声情報をテキストに変換した．音響モデルとして，CSJ から学習したコンテキスト依存音節モデル (928 音節) を，言語モデルとして，毎日新聞 (75ヶ月分) を用いた．語彙サイズは 20K 語とした．言語モデルの作成に用いた毎日新聞の発行時期は，1991 年 1 月から 1994 年 9 月まで，および 1995 年 1 月から 1997 年 6 月までである．よって，NHK コーパスのニュース放映時期と，言語モデルの新聞発行時期は重複しており，未知語が比較的少なく，かつ，ドメインが類似していることが期待される．ただし，発話スタイルは大きく異なっていると考えられる．書き起こし (正解テキスト) と比較したところ，単語正解率は 0.764，単語正解精度は 0.736 だった．

音声認識結果に対する人名抽出精度を求めるには，音声認識結果に対して固有表現ラベルを付与する必要がある．本稿では，書き起こしと認識結果を，文字を単位とする編集距離を最小とするように照合し，得られた文字対応関係に基づいて，音声認識結果に固有表現ラベルを付与した．例を**図 2** に示す．また，固有表現の種類別の文字正解率と文字正解精度を**表 2** に示す．表 2 より，人名は，固有表現の中でも音声認識が困難なカテゴリに属することが分かる．

このようにして作成した固有表現ラベルつき書き起こしと固有表現ラベルつき認識結果を対象として，3.3 節で述べた固有表現抽出器を用いて人名を抽出した結果を**表 3** に示す．なお，ベースラインは，類似形態素素性 SF を用いない手法である．表 3 より，提案手法を書き起こしに対して適用した場合には，ベースラインより性能が改善されているにも関わらず，提案手法を認識結果に対して適用した場合には，殆んど性能の改善が見られないことが分かる．

次に，音声認識用辞書と人名抽出結果の関係を**表 4**

⋆1 http://crfpp.sourceforge.net/

⋆2 NHK 技研によって作成されたコーパスであり，一般には公開されていない．

固有表現ラベルつき書き起こし	··· 支持を得ていた<PERSON>チャムロン</PERSON>元知事 ···
音声認識結果	··· 支持を得ていた 八分ごろ 元知事 ···
固有表現ラベルつき音声認識結果	··· 支持を得ていた<PERSON> 八分ごろ </PERSON>元知事 ···

図 2 音声認識結果に対する固有表現ラベルの付与 (例)

表 3 人名抽出結果

	ベースライン			提案手法		
	Rec.	Prec.	$F_{\beta=1}$	Rec.	Prec.	$F_{\beta=1}$
書き起こし	0.804	0.806	0.769	0.855	0.761	0.805
音声認識結果	0.571	0.537	0.554	0.562	0.547	0.555

表 2 固有表現種類別の文字正解率と文字正解精度

種類	Cor.	Acc.
DATE	0.754	0.714
LOCATION	0.734	0.711
MONEY	0.919	0.909
ORGANIZATION	0.755	0.738
PERCENT	0.535	0.518
PERSON	0.506	0.477
TIME	0.470	0.453
固有表現以外	0.783	0.772
全体	0.775	0.762

表 4 認識辞書と人名抽出性能 ($F_{\beta=1}$)

	ベースライン	提案手法
認識辞書に登録されている人名	0.759	0.779
認識辞書に登録されていない人名	0.493	0.477

に示す．認識辞書に 1 単語として登録されていない人名が，認識辞書に登録されている人名の組み合わせとして認識されていた場合は，以下の 2 例のみだった．

- 町長の<PERSON>木村光雄</PERSON>容疑者
- 人道問題担当の<PERSON>明石康</PERSON>事務次長

この 2 例以外は，認識辞書に登録されていない人名は全て，認識辞書に登録されている適当な語 (人名以外の語を含む) の組み合わせとして認識されていた．さらに，音声認識結果と人名抽出結果の関係を**表 5** に示す．表 5 より，認識誤りを含まない場合には，人名以外の語として認識されている場合を含めて，提案手法はベースラインよりも良い性能を示している．

以上より，筆者らの固有表現抽出手法は，固有表現ラベル付きコーパスにとっての未知語には対応できるが，音声認識誤りには脆弱であることが分かる．よって，適切なプライバシ保護を実現するためには，固有表現ラベル付きコーパスの不足に対応するだけでなく，音声認識誤りに対して頑健な手法が必要である．

表 5 認識誤りと人名抽出性能 ($F_{\beta=1}$)

	ベースライン	提案手法
認識誤りを含まない人名	0.815	0.856
認識誤りを含む人名	0.448	0.420

参 考 文 献

1) 美濃導彦：センシングウェブ : 概念と課題，人工知能学会誌， Vol.24, No.2, pp.179–184 (2009).
2) 角所　考，満上育久，美濃導彦：カメラ映像におけるプライバシ対処のためのアプローチ，人工知能学会誌， Vol.24, No.2, pp.196–201 (2009).
3) 中川聖一，山本一公，土屋雅稔：音声に含まれるプライバシ情報の保護，人工知能学会誌， Vol.24, No.2, pp.190–195 (2009).
4) 小林大祐，梶田将司，武田一哉，板倉文忠：ヒューマンスピーチライク雑音における音声的特徴の分析，電子情報通信学会技術報告，No.SP95-105, pp.85–92 (1995).
5) 藤崎博也：音声の韻律的特徴における言語的・パラ言語的・非言語的情報の表出，電子情報通信学会技術報告，No.HC94-37, pp.1–8 (1994).
6) Miller, S., Guinness, J. and Zamanian, A.: Name Tagging with Word Clusters and Discriminative Training, *Proc. of HLT-NAACL 2004*, pp.337–342 (2004).
7) Suzuki, J. and Isozaki, H.: Semi-Supervised Sequential Labeling and Segmentation Using Giga-Word Scale Unlabeled Data, *Proc. of ACL'08-HLT*, pp.665–673 (2008).
8) Tsuchiya, M., Hida, S. and Nakagawa, S.: Robust Extraction of Named Entity Including Unfamiliar Word, *Proceedings of ACL-08: HLT, Short Papers*, Columbus, Ohio, Association for Computational Linguistics, pp.125–128 (2008).
9) 鍜治伸裕，岡本雅史，黒橋禎夫：WWW を用いた書き言葉特有語彙から話し言葉語彙への用言の言い換え，自然言語処理， Vol.11, No.9, pp. 19–37 (2004).
10) 下岡和也，南條浩輝，河原達也：講演の書き起こしに対する統計的手法を用いた文体の整形，自然言語処理， Vol.11, No.2, pp.67–83 (2004).
11) 秋田祐哉，河原達也：統計的機械翻訳の枠組みに基づく言語モデルの話し言葉スタイルへの変換，No.2005–SLP–127, 社団法人情報処理学会，pp. 109–114 (2005).