音声翻訳システム実利用データを用いた統計的機械翻訳のモデル適応

安田圭志　大熊英男　内山将夫　隅田英一郎　磯谷亮輔　河井恒　中村哲

情報通信研究機構　言語翻訳グループ

〒619-0289　京都府「けいはんな学研都市」光台3-5

E-mail: {keiji.yasuda, hideo.okuma,mutiyama,eiichiro.sumita,ryosuke.isotani,
hisashi.kawai,satoshi.nakamura}@nict.go.jp

1. はじめに

近年における音声言語処理技術の発展により，音声翻訳システムの実用化が進んでいる[1, 2]. 今後の更なるシステム性能の改善には，従来から取り組まれている要素技術の改善に加え，音声翻訳システムの実利用データを有効利用したシステム改善への取り組みが重要となってきている.

音声翻訳システムは主に，音声認識，機械翻訳，音声合成の3つの要素技術から構成される．音声認識の従来研究[3]では, 音声認識システムの実利用データを用いたシステム性能の改善方法が提案されており，実利用データの利用が，システム性能の改善に有効であることが示されている.

機械翻訳について注目すると，音声入力を前提とした機械翻訳では，機械翻訳への入力に音声認識誤りが含まれる可能性があり，この点においてテキストを対象とした機械翻訳と異なっている．従来の機械翻訳の研究では，テキスト入力を対象とした実利用データ利用の研究がなされ，その有効性が示されているものの[4], 音声入力の実利用データを用いた研究[1]は少なく, 限られた条件での有効性しか示されていない.

ここで, 実利用データの利用方法について述べる. 最もシンプルな利用方法は, 得られた実利用データの音声の書き起しと, その対訳作成を人手により行ない, 音声認識や機械翻訳システムのモデル学習に加えるという方法(教師あり学習)である．この方法はシンプルかつ効果が大きい反面, 書き起しや対訳作成がボトルネックとなり, 多くのデータを処理することが出来ないという問題がある．これらの問題を解決するため, 音声認識や機械翻訳の結果を, 信頼尺度等でフィルタリングし[4, 5], 正しいと自動判定されたデータのみを利用する方法（教師無し学習）も提案されている. このような方法では, 新出語などの追加は行えないものの, 既存のモデルを実利用データに適応化させる効果があると考えられる.

本研では，音声入力を対象とした機械翻訳システムの教師無し学習に取り組み，2009年度に実施された音声翻訳実証実験のデータを用いた実験結果を示す.

2では音声翻訳実証実験について, 3では提案手法について述べる．4 では機械翻訳システムのモデル適応実験について述べ，最後に5で本論文を結ぶ.

2. 音声翻訳実証実験

本論文では，2009年度に実施された音声翻訳実証実験[2]により収集されたデータセットを用いる．ここでは音声翻訳実証実験について説明する.

2.1 概要

本実証実験は，自動音声翻訳技術の翻訳精度の飛躍的向上及び訪日観光分野における同技術を活用したサービスの早期実用化を図ることを目的しており, 総務省が「地域の観光振興に貢献する自動音声翻訳技術の実証実験」(総事業予算額9.85億円)を民間法人等に委託して実施した.

実証実験は, 日英中韓の4ヶ国語を対象とし, Fig.1に示す通り，全国5地方の観光施設等約370箇所に約1700台の端末を設置して行われた。実験期間中には，約20万件のアクセスが記録された．このように大規模で，実利用に近い条件下での実証実験は，世界的にも類を見ない.

NICT は，実証実験を受託したすべての事業者に対して音声翻訳技術を提供するとともに，実験システム構築，運用，データ分析等の面で全面的にサポートした.

2.2 システム構成

各地方プロジェクトが構築した実証実験システムの簡略化された構成図を Fig.2 に示す。音声翻訳端末は，スマートフォン，ノート PC などからなり，台数は300～500である。端末で入力された音声は，16kHz サンプリングの ADPCM 形式で音声翻訳サーバーに送られる。音声翻訳サーバーは，実際には言語ごとに用意された音声認識，機械翻訳，音声合成用のサーバー群から構成される．翻訳結果は，テキストおよび合成音声の形で端末に送信される. また，入力音声，音声認識結果，翻訳結果は，日時，端末ID，言語指定等の情報とともに利用ログとしてシステム内に蓄積される.

Fig.3 は，音声認識，機械翻訳，音声合成からなる音声翻訳システムの内，機械翻訳部の処理の詳細で

Copyright(C) 2011 The Association for Natural Language Processing.
All Rights Reserved.

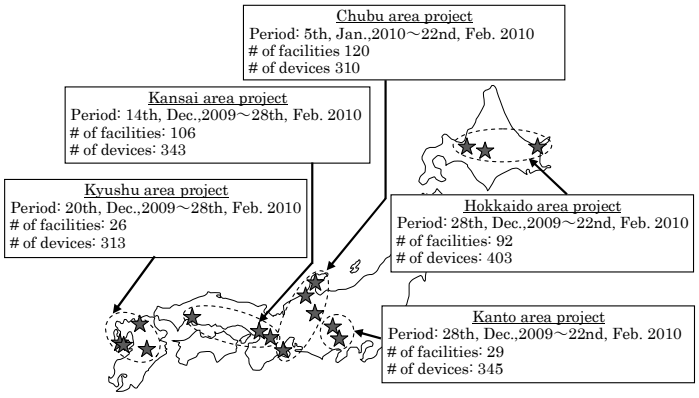

Fig. 1 Overview of the five local projects.

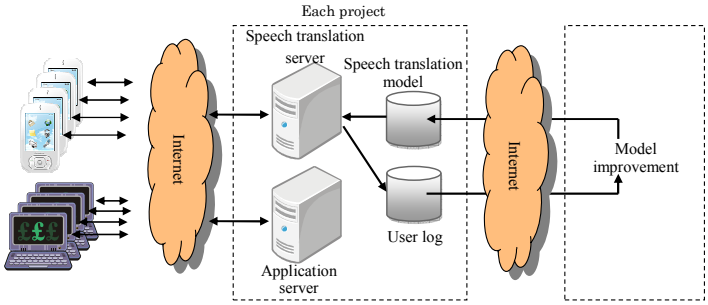

Fig.2 A schematic diagram of system configuration for the speech-to-speech translation experiment.

ある．機械翻訳部は，主に統計的機械翻訳と 2 つの翻訳メモリから構成されている．

統計翻訳システムは，フレーズベース型統計翻訳の枠組みを利用した．本手法は，翻訳対象の原言語の単語列(f)に対する目的言語の単語列(e)の確率を次式により求める．

$$p(f \mid e) = \frac{\exp(\sum_{i=1}^{M} \lambda_i \cdot h_i(e,f))}{\sum_{e'} \exp(\sum_{i=1}^{M} \lambda_i \cdot h_i(e',f))} \quad (1)$$

ここで，$h_i(e,f)$は，目的言語から原言語，原言語から目的言語への単語やフレーズ単位の翻訳確率，目的言語の言語モデル確率等からなる素性関数[6]である．統計的機械翻訳では，式(1)を用い翻訳結果$\hat{e}$を次式により求める．

$$\hat{e}(f,\lambda) = \arg\max_e \sum_{i=1}^{M} \lambda_i h_i(e,f)$$

各モデルの学習には，MOSES[6]ツールキットと SRILM ツールキットとを用いて，翻訳モデルと言語モデルの学習を行っている．

実験実施地方毎に，数千文からなる地域固有の表現（固有表現）とその対訳(英中韓)を事前にテキストで収集した．実際の各モデルの学習には，BTEC コーパス[7]に加え，この固有表現を用いている．データの使い方は，まず，前述のツールキットを用いて，データ毎に個別にモデルを学習し，次に，両モデルを次式により線形結合して利用した．

$$h_{baseline}(e,f) = 0.9 \cdot h_{baseline}(e,f) + 0.1 \cdot h_{regional}(e,f)$$

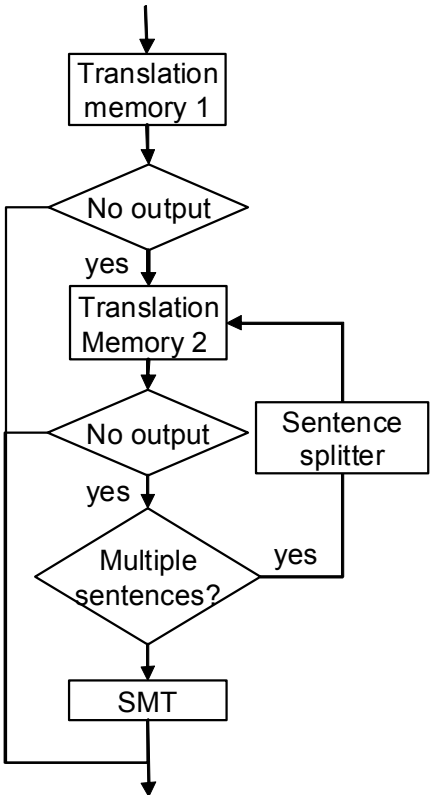

Fig.3 A flow of the Machine Translation subsystem.

ここで，$h_{baseline}(e,f)$は線形結合後のモデル，$h_{btec}(e,f)$は BTEC コーパスを用いて学習したモデル，$h_{regional}(e,f)$は固有表現を用いて学習したモデルをそれぞれ表す．

Fig.3 の翻訳メモリ 1 では，前述の BTEC コーパスを，翻訳メモリ 2 では，各地方ごとに収集した固有表現をそれぞれ用いている．

3. 提案手法

提案手法では，機械翻訳結果を逆翻訳した結果を用い，モデルのアダプテーションに用いるデータの取捨選択を行う．3.1 ではデータの取捨選択方法について述べ，3.2 では取捨選択されたデータをモデルアダプテーションに用いる方法について説明する．

3.1 アダプテーションデータ選択手法

提案手法では，まず，順方向の機械翻訳結果を，再度原言語に機械翻訳する．次に，順方向の機械翻訳への入力である音声認識結果を参照訳とみなし，逆翻訳の結果の PER(Position independent word Error Rate)を計算し，この値が閾値以下の場合は，アダプテーション用データとして利用する．

3.2 アダプテーション手法

選択された実利用データは， 2 で述べた固有表現データとともに，次に述べる方法で用いる．

Step1: 得られた実利用データと 2 で述べた固有表現とを結合し，アダプテーション用コーパスとする．

Copyright(C) 2011 The Association for Natural Language Processing.
All Rights Reserved.

Table1 Evaluation results of supervised and unsupervised adaptation (without data filtering)

Project Area	System type	Additional field data			Ratio (%)			
		Transcription	Translation	Size (# of sentences)	S	S, A	S, A, B	S, A, B, C
Hokkaido	Baseline	N/A	N/A	0	29	38	55	62
	Baseline + unannotated data	ASR	MT	9602	29	38	53	61
	Baseline + annotated data 1	Manual	MT	10009	31	39	51	62
	Baseline + annotated data 2 (Upper bound)	Manual	Manual	10335	34	44	61	68
Kyushu	Baseline	N/A	N/A	0	50	62	72	76
	Baseline + unannotated data	ASR	MT	9722	50	60	71	76
	Baseline + annotated data 1	Manual	MT	10337	49	62	72	77
	Baseline + annotated data 2 (Upper bound)	Manual	Manual	14138	55	64	74	79

Step2: Step1 で得られたコーパスを用いて, モデル ($h_{field}(e,f)$)の学習を行う.

Step3: BTEC コーパスを用いて学習したモデル $h_{btec}(e,f)$と $h_{field}(e,f)$を，式(2)により線形結合し，アダプテーションモデル($h_{adapted}(e,f)$)とする.

$$h_{adapted}(e,f) = 0.9 \cdot h_{btec}(e,f) + 0.1 \cdot h_{field}(e,f) \quad (2)$$

4. 実験

4.1 実験条件

実験では，5 つの実証実験実施地方の内，音声翻訳性能が最も低かった北海道と，最も高かった九州のデータを用いた．翻訳方向は日英とした．モデル学習に用いた日英 BTEC コーパスは，691,829 文からなる．また，北海道と九州地区の固有表現はそれぞれ，3000 文と 5,095 文からなる.

機械翻訳の評価では, それぞれの地方のデータに対してランダムに抽出した 100 文を評価セットとして用いた．訳質の評価として 5 段階の主観評価 S,A,B,C,D(S>A>B>C>D)を実施した．また， 4.2 で示す全ての評価では，機械翻訳への入力は音声認識を含まないテキスト入力としている.

4.2 実験結果

Table1 は，実利用データの取捨選択を行わなかった場合の結果である．各地方の結果において，1 行目はベースライン，2 行目は利用可能な実利用データ全てを用いて教師無し学習を行った結果(完全教師無し学習)，3 行目は入力音声の書き起しは人手で行い，アダプテーションに用いる目的言語側の情報として機械翻訳の出力を用いた場合の結果(一部教師無し学習)を表す. 最後の 4 行目は書き起しも対訳作成も全て人手で行った結果（教師有り学習）で，アダプテーションによる性能改善の上限を表す．同地域においても，条件により実利用データのサイズが異なるのは，音声認識や機械翻訳の過程で出力が得られなかったデータはアダプテーションに利用していないためである.

表中の白いセルは，ベースラインの性能を上回った場合，ライトグレーのセルはベースラインと同じ性能の場合，ダークグレーのセルは，ベースラインの性能を下回った場合をそれぞれ表す．Table1 を見ると，完全教師無し学習では全く改善が得られていない．一部教師無し学習においては，一部の条件で性能の改善が得られているものの，性能が劣化することもある．一方，教師有り学習では，全ての場合において，性能の改善が得られている.

Table2 は，完全教師無し学習の条件で，提案手法により実利用データの取捨選択を行った結果を表している．北海道のデータセットでは閾値を 0.1 とした場合の一部で性能が劣化しているものの，ほぼ全ての場合において, 性能が向上している. その反面, 九州のデータセットでは, ほぼ全ての条件において, 性能の劣化が生じている.

Table3 は, Table2 の結果に加え, 言語モデルのみ, または，翻訳モデルのみのアダプテーションに実利用データを用いた場合の結果を示している．Table3 を見ると，北海道のデータセットでは，両モデルのアダプテーションを行った場合に最も大きな改善が得られている．一方，九州のデータセットにおいては，言語モデルに利用した場合に著しい劣化が生じており，反面，翻訳モデルに用いた場合には，改善が得られている.

5. まとめ

音声翻訳システムの実利用データを用いた教師無しアダプテーション手法を提案した. 提案手法では, 実利用データを逆翻訳し，入力文と逆翻訳結果が近いデータのみをアダプテーションに用いる.

実験では、2009 年度に全国で実施された音声翻訳実証実験のデータを用いた．実験の結果，ベースラインの性能が低い北海道地区のデータを用いた場合, 提案手法により翻訳性能の改善が得られた．一方，ベースラインの性能が高い九州地区のデータでは,

Copyright(C) 2011 The Association for Natural Language Processing.
All Rights Reserved.

Table2 Evaluation results unsupervised adaptation-1(with data filtering)

Project Area	System type	Additional field data		Ratio (%)			
		Threshold	Size (# of sentences)	S	S, A	S, A, B	S, A, B, C
Hokkaido	Baseline	N/A	0	29	38	55	62
	SRC_PER_1	*PER* <= 0.1	1244	31	40	55	66
	SRCPER_2	*PER* <= 0.2	1861	32	41	56	69
	SRC_PER_3	*PER* <= 0.4	3565	32	41	56	66
Kyushu	Baseline	N/A	0	50	62	72	76
	SRC_PER_1	*PER* <= 0.1	4560	49	60	70	74
	SRC_PER_2	*PER* <= 0.2	5274	49	61	71	75
	SRC_PER_3	*PERS* <= 0.4	6699	51	61	71	74

Table3 Evaluation results unsupervised adaptation-2 (with data filtering)

Project Area	System type	Additional field data		Ratio (%)			
		Used for LM training	Used for TM training	S	S, A	S, A, B	S, A, B, C
Hokkaido	Baseline	No	No	29	38	55	62
	SRC_PER_1	Yes	Yes	31	40	55	66
	SRC_PER_2	Yes	Yes	32	41	56	69
	SRC_PER_3	Yes	Yes	32	41	56	66
	SRC_PER_1_L	Yes	No	32	40	54	63
	SRC_PER_2_L	Yes	No	32	41	56	64
	SRC_PER_3_L	Yes	No	31	41	55	64
	SRC_PER_1_T	No	Yes	31	40	54	64
	SRC_PER_2_T	No	Yes	32	40	55	64
	SRC_PER_3_T	No	Yes	32	40	54	63
Kyushu	Baseline	No	No	50	62	72	76
	SRC_PER_1	Yes	Yes	49	60	70	74
	SRC_PER_2	Yes	Yes	49	61	71	75
	SRC_PER_3	Yes	Yes	51	61	71	74
	SRC_PER_1_L	Yes	No	47	57	68	73
	SRC_PER_2_L	Yes	No	47	57	68	73
	SRC_PER_3_L	Yes	No	47	57	68	73
	SRC_PER_1_T	No	Yes	51	62	73	76
	SRC_PER_2_T	No	Yes	50	61	73	76
	SRC_PER_3_T	No	Yes	52	63	73	76

性能の劣化が見られた．しかしながら，九州のデータセットにおいては，言語モデルのアダプテーションを行わず，翻訳モデルのみのアダプテーションを行うことによりある程度の性能改善がえられることが示された．

実運用時においては，あらかじめ開発セット等を用意しておき，データをフィルタリングする際の閾値や，アダプテーションを適用するモデルを，適宜決めて行く必要があるものの，人手による書き起しや対訳作成無しに，システム性能の改善が得られることが示された．

文献

[1] Nguyen Bach et al. 2009. Incremental adaptation of speech-to-speech translation. In Proceedings of NAACL HLT 2009, pages 149–152.

[2] 河井恒他，H21 年度全国音声翻訳実証実験の概要，日本音響学会 2010 年秋季研究発表会, pages 99–102.

[3] Frank Wessel et al. 2005. Unsupervised training of acoustic models for large vocabulary continuous speech recognition. IEEE Transactions on Speech and Audio Processing, 13:23–31.

[4] Nicola Ueffing et al. 2007. Semi-supervised model adaptation for statistical machine translation. Machine Translation, 21(2):77–94.

[5] Keiji Yasuda et al. 2008. Method of selecting training data to build a compact and efficient translation model. In Proceedings of the Third International Joint Conference on Natural Language Processing, pages 655–660.

[6] Philipp Koehn et al., 2007. Moses: Open source toolkit for statistical machine translation. In Proceedings of the 45th Annual M eeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions, pages 177–180. Association for Computational Linguistics, June.

[7] Genichiro Kikui et al. 2006. Comparative study on corpora for speech translation. In IEEE Transactions on Audio, Speech and Language Processing, volume 14(5), pages 1674–1682.

Copyright(C) 2011 The Association for Natural Language Processing.
All Rights Reserved.