言語横断 LDA モデルを用いた統計的機械翻訳システム

西尾 拓　　福富 崇博　　貞光 九月　　山本 幹雄

筑波大学 システム情報工学研究科,
{taku@mibel.cs,fukutomi@mibel.cs,sadamitsu@mibel.cs,myama@cs}.tsukuba.ac.jp

1 はじめに

従来の統計的機械翻訳では言語モデルとして $n$gram モデルを用いるのが一般的であるが、$n$gram モデルは近距離の単語の依存関係をモデル化するものであるため、文書全体の属するトピックのようなより広範囲の単語の依存関係を考慮することができない。本稿では、このようなトピック情報を翻訳システムに取り込む目的で、言語横断 LDA モデルを用いた統計的機械翻訳システムを提案する。言語横断 LDA モデルとは、対訳となっている文書は同じトピックに属すると仮定し、対訳文書を 1 つの文書として学習した LDA モデルである。このモデルでは原言語、目的言語ともに同じトピックの次元に圧縮を行うため、原言語で捉えたトピックをそのまま目的言語の評価に反映させることが可能となる。

本稿ではまず 2 章にて統計的機械翻訳の理論的枠組みを説明し、3 章で言語横断 LDA モデル、4 章で言語横断 LDA モデルと統計的機械翻訳システムの統合手法について述べる。その後 5 章に言語横断 LDA モデル、及び言語横断 LDA モデルを用いた統計的機械翻訳システムの性能を示し、6 章でまとめと今後の課題を述べる。

2 統計的機械翻訳

統計的機械翻訳では、原言語 E から目的言語 J への翻訳を Noisy channel model(雑音のある通信路モデル)でモデル化する (Brown, Pietra, Pietra, and L.Mercer 1993)。このモデルでは目的言語 J は雑音のある通信路を通過したことによって原言語 E に変換されてしまったものと仮定し、翻訳は原言語 E から目的言語 J への復号化 (decode) であるとみなす。復号化の誤りを最小化するような翻訳候補 $\hat{J}$ を求める式は以下のように表現される。

$$\hat{J} = \arg\max_{J} P(J|E) \tag{1}$$

$$= \arg\max_{J} P(J)P(E|J) \tag{2}$$

上述の式は、復号化誤り確率を最小化するものと解釈できる。我々は真の確率分布 $P(J|E)$ を知る事ができないため、実際には $P(J|E)$ を近似するモデル $p(J|E)$ を推定する。$P(J|E)$ を近似する手法としては、log-linear model を用いる手法 (Och 2003) が提案されている。以下に、log-linear model を用いた統計的機械翻訳の理論的枠組みを示す。

$$P(J|E) \approx p_{\lambda_1^M}(J|E) \tag{3}$$

$$= \frac{\exp\left[\sum_{m=1}^{M} \lambda_m h_m(J,E)\right]}{\sum_{J_1'^I} \exp\left[\sum_{m=1}^{M} \lambda_m h_m(J_1'^I,E)\right]} \tag{4}$$

式中 $h_m(J,E)$ は、翻訳精度に貢献する性質を持つ M 個の特徴関数である。それぞれの特徴関数はモデルパラメータ λ_m を持つ。式 (4) の分母は任意の翻訳候補集合に対して一定の値となることから、式 (4) を式 (1) に代入すると以下のように書くことができる。

$$\hat{J} = \arg\max_{J} \sum_{m=1}^{M} \lambda_m h_m(J,E) \tag{5}$$

すなわち、それぞれの特徴関数をモデルパラメータで重み付けし、足し合わせた値が最大となるような翻訳候補の探索問題となる。式 (2) から、従来の統計的機械翻訳の特徴関数は $P(J)$ を近似する言語モデル $p(J)$ と、$P(E|J)$ を近似する翻訳モデル $p(E|J)$ とに大別される。

3 言語横断 LDA モデル

3.1 LDA モデル

トピックをモデル化する手法としては、Latent Dirichlet Allocation (LDA) モデル (M.Blei, Y.Ng, and I.Jordan 2003) が良い性能を示すことが知られている。LDA モデルでは、次に単語 w^* が出現する期待値を以下の式で計算する。

$$P(w^*|\boldsymbol{h}) = \int P(w^*|\boldsymbol{\theta})P(\boldsymbol{\theta}|\boldsymbol{h})d\boldsymbol{\theta} \tag{6}$$

ここで $P(w^*|\boldsymbol{\theta})$ はあるトピックの分布 $\boldsymbol{\theta}$ における単語の出現確率、$P(\boldsymbol{\theta}|\boldsymbol{h})$ はある単語出現履歴 $\boldsymbol{h}$ が与えられたときのトピックの事後分布を表している。

式 (6) は解析的に求めることが困難であるため、一般的には変分ベイズ法、マルコフ連鎖モンテカルロ法等の近似的手法を用いて計算される。本研究では変分ベ

イズ法で $P(\boldsymbol{\theta}|\boldsymbol{h})$ を別の関数 $q(\boldsymbol{\theta};\boldsymbol{\gamma})$ で変分近似することにより、式 (6) の計算を行う。$q(\boldsymbol{\theta};\boldsymbol{\gamma})$ をパラメータ $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \cdots, \gamma_Z)$ を持つディリクレ分布とすると、$P(w^*|\boldsymbol{h})$ は最終的に、以下の式で計算できる。

$$\begin{aligned} P(w^*|\boldsymbol{h}) &= \int \left\{ \sum_{z=1}^{Z} P(w^*|z) \right\} q(\boldsymbol{\theta};\boldsymbol{\gamma}) d\boldsymbol{\theta} \\ &= \sum_{z=1}^{Z} P(w^*|z) \int \theta_z q(\boldsymbol{\theta};\boldsymbol{\gamma}) d\boldsymbol{\theta} \\ &= \sum_{z=1}^{Z} P(w^*|z) \frac{\gamma_z}{\sum_{z'} \gamma_{z'}} \end{aligned} \tag{7}$$

以下、文脈情報 $\boldsymbol{h}$ からトピックの分布を推定する過程を適応と呼ぶ。LDA モデルでは適応によって得られる事後分布 $p(w^*|\boldsymbol{h})$ を用いて、対象文書の評価を行う。

3.2 言語横断 LDA モデル

言語横断情報検索の手法のひとつとして LSI(Latent Semantic Indexing) を言語横断に拡張した、言語横断 LSI が挙げられる (Littman, Dumais, and Landauer 1998)。この手法では 2 つの言語で書かれた同じ内容の対訳文書を混合し、1 つの文書として学習することによって、LSI の言語横断への拡張を実現している。

本研究では言語横断 LSI の考え方を LDA モデルに適用することで、LDA モデルを言語横断に拡張する (言語横断 LDA モデル)。言語横断 LDA モデルは原言語-目的言語間で同じトピックを共有するため、原言語でとらえたトピックをそのまま目的言語の確率評価に反映させることが可能となる。

言語横断 LDA モデルを用いた目的言語の確率評価は次のように行う。まず、原言語側文書を文脈情報 $\boldsymbol{h}$ とし、変分ベイズ推定を行う。その後、変分ベイズ推定によって得られた $\boldsymbol{\gamma}$ を式 (7) に代入することによって、目的言語側における単語出現の期待値 $P(w^*|\boldsymbol{h})$ を計算する。

4 言語横断 LDA モデルを用いた統計的機械翻訳システム

4.1 言語横断 LDA モデルを用いた統計的機械翻訳

統計的機械翻訳の理論的枠組みが log-linear model を用いて表せることは 2 章で述べた。従来の統計的機械翻訳システムでは、翻訳候補の翻訳らしさを以下の式で評価する。

$$\begin{aligned} score(s) =& \lambda_{\mathrm{LM}} L(s) + \lambda_{\mathrm{TM}} T(s) + \lambda_{\mathrm{D}} D(s) \\ & - \lambda_{\mathrm{WP}} |s| - \lambda_{\mathrm{UNK}} \cdot unk(s) \end{aligned} \tag{8}$$

ここで、s は翻訳候補文を表す。$L(s)$ は $n$gram モデルによる尤度、$T(s)$ は翻訳モデルによる尤度、$D(s)$ は歪みモデルによる尤度、$|s|$ は s に含まれる単語数、$unk(s)$ は s に含まれる未知語の数であり、それぞれ翻訳精度に寄与する特徴関数である。特徴関数が持つモデルパラメータ $\boldsymbol{\lambda}$ は、Minimum Error Rate Training(以下、MERT)(Och 2003) によって決定する。本研究ではリランキング法と log-linear model による手法の 2 つによって、それぞれ言語横断 LDA モデルと統計的機械翻訳システムの統合を実現する。言語横断 LDA モデルの捉える大域的な文脈情報を統計的機械翻訳に持ち込むことで、システム全体の翻訳精度の向上が期待できる。

LDA モデルを用いた統計的機械翻訳に関する先行研究としては、トピックベース単語アラインメントによる翻訳モデルの研究 (Zhao and Xing 2008) が挙げられるが、我々の手法は言語横断 LDA モデルを独立した素性として扱う点で異なる。また、原言語側で学習した LDA モデルのパラメータを用いて目的言語側の LDA モデルの学習を行うことによって、2 言語間でのトピックの共有を実現する研究 (Tam, lan Lane, and Schultz 2007) も報告されている。この手法では原言語、目的言語それぞれについて言語モデルの学習を行う必要があるのに対し、我々の手法は 1 度の学習で原言語、目的言語双方のモデル化ができるという点が異なる。

4.2 リランキング法による統合

リランキング法ではあらかじめ、従来の統計的機械翻訳システムによって高い $score(s)$ を付与されたものから順に最大 N 個の翻訳候補を出力しておく。以下では、この N 個の翻訳候補を N-best と呼ぶ。次に文書尤度、または文尤度を考慮した翻訳候補の評価式によって N-best の再評価を行い、このとき最も高い評価を得られた翻訳候補を最終的な翻訳結果として出力する。

文書尤度 Doc(S) を考慮した評価式を式 (9) に示す。

$$R(S) = \sum_{i=1}^{M} score(s_i) + \lambda_{\mathrm{doc}} \mathrm{Doc}(S) + \lambda_{\mathrm{wp}} |S| \tag{9}$$

翻訳候補の単語数が少ないと文書尤度の計算で有利な結果が得られる傾向にあるため、式中で単語数 $|S|$ によるペナルティを課している。

実際の評価では文書 S に含まれる各文の N-best から、最適な組み合わせを探索する必要がある。しかし、例え

ば文書に含まれる文数が M 文のとき、翻訳候補文の組み合わせは N^M と膨大な数にのぼる。従って本研究では、最適な翻訳候補の組み合わせを探索するアルゴリズムとして局所的な探索法を採用した。

次に、文尤度 $Sen(s_i)$ を用いた評価式を (10) に示す。

$$R(s) = score(s) + \lambda_{\mathrm{sen}}\mathrm{Sen}(s) + \lambda_{\mathrm{wp}'}|s| \quad (10)$$

ここでいう文尤度とは、各翻訳候補が文として成り立っているかを評価するものである。文尤度は文書尤度ほどの大域的な情報を利用することはできないが、翻訳先の文のトピックをより鋭く捉えることができ、さらに計算時間が短く済むという利点がある。

4.3 log-linear model による統合

式 (8) に示したように、従来の統計的機械翻訳システムの評価式は log-linear model によって表される。本手法では従来の統計的機械翻訳システムに言語横断 LDA モデルを新たな特徴関数として加え、以下の式で翻訳候補の評価を行う。

$$score(s) = \lambda_{\mathrm{LM}}L(s) + \lambda_{\mathrm{TM}}T(s) + \lambda_{\mathrm{D}}D(s) \quad (11)$$
$$- \lambda_{\mathrm{WP}}|s| - \lambda_{\mathrm{UNK}} \cdot unk(s) + \lambda_{\mathrm{LDA}}LDA(s)$$

ここで、$LDA(s)$ は言語横断 LDA モデルの出力する文 s の対数尤度を表す。リランキング法では再評価式による改善が N-best に制限されるのに対し、この手法は LDA モデルを直接翻訳システムに組み込むため探索空間が広がり、より大幅な性能の向上が期待できる。

5 実験

5.1 実験の概要

はじめに言語横断 LDA モデルのパープレキシティ測定を行い、続いて翻訳実験を行った。翻訳実験ではリランキング法、log-linear model による手法の 2 つについて、それぞれ BLEU 値評価を行った。実験条件の詳細を表 1、表 2 にそれぞれ示す。

言語横断 LDA モデルの学習には、NTCIR-7 特許翻訳タスクで配布された特許文対訳コーパスを用いた。学習データとしてコーパスに含まれる文書全てを用いた場合と、8 文以上を含む文書のみを用いた場合とでパープレキシティの比較をしたところ、8 文以上を含む文書のみで学習したほうが良い性能を示した。この結果を受けて、本実験では 8 文以上を含む文書のみを学習データとして用いることとした。

リランキング法による統合では Moses の出力する N-best を用いたため、標準の Moses の出力結果をベースラインに定めた。また、log-linear model による統合では当研究室で開発された Pharaoh[*1]のコピー (Khafra) を用いたため、標準の Khafra の出力結果をベースラインに定めた。

表 1　言語横断 LDA 性能評価実験条件

学習データ	特許文対訳コーパス[*2]180 万文
トピック学習データ	上記データのうち 8 文以上含む文書 文書数 32,522 総文数 1,753,841
語彙サイズ	英語 139,491 日本語 121,815
develpoment データ	文書数 504 テスト文数 915 文脈データを含む総文数 79,745
テストデータ	文書数 515 テスト文数 899 文脈データを含む総文数 72,549
テスト文語彙サイズ	英語 3,867 日本語 3,696
トピック数	1,2,5,10,20,50,100,200,500
学習時の条件	α の初期値: 乱数による生成 収束条件: パラメータ変化率 1% 以下
適応時の条件	収束条件: パラメータ変化率 1% 以下

表 2　翻訳システム性能評価実験条件

統合手法	リスコアリング法	log-linear model
トピック数	500	2,5,10,20,50,100,200
デコーダ	Moses[*3]	Khafra
reordering table	あり	なし
distortion limit	なし	なし
N-best 数	100	-
言語モデル	SRI Language Modeling Toolkit[*4] により学習した 5gram モデル	
discounting 法	Modified Kneser Ney	
補間法	線形補間法	
単語アラインメント	GIZA++[*5]	
翻訳モデル	フレーズベースモデル	
対称化手法	grow-diag-final	

5.2 言語横断 LDA モデルに関する事前実験

原言語側の文書全体を文脈情報として適応し、文書尤度を用いてリランキングを行った先行研究 (西尾 2008) では、翻訳精度の改善がわずかであった。この理由としては、実際には文書中でトピックが変化するために、文書全体で捉えたトピック情報にノイズが含まれていたからではないかと考えられる。そこで本実験では、文書全体で適応した場合と翻訳元の対応する文のみを用いて適応した場合とのそれぞれについてパープレキシティを測定し、どのような違いが生じるかの考察を行う。それぞれの適応イメージを図 1 に示す。

本研究では言語横断 LDA モデルを翻訳システムと統合し、原言語側で適応を行い、目的言語側で確率評価を

[*1] http://www.isi.edu/licensed-sw/pharaoh/
[*2] NTCIR-7
[*3] http://www.statmt.org/moses/
[*4] http://www.speech.sri.com/projects/srilm/
[*5] http://www.fjoch.com/GIZA++.html

行うことを想定している。従って事前実験では日本語文書で適応し、英語の評価対象文で確率評価したときのパープレキシティ測定を行った。トピック数の増加に伴うパープレキシティの推移を図 2 に示す。翻訳元の対応文のみで適応すると、文書全体で適応した場合に比べてパープレキシティは低い値を示した。すなわち、適応する範囲を翻訳元の対応する文のみに絞ったほうが言語モデルとしての予測性能が高くなるといえる。

言語横断 LDA モデルが識別した各トピックにおいて、単語の出現確率の変化率が高かったものを表 3 に示す。各トピックにクラスタリングされた単語集合中には、類語、及び対訳となっている単語が多く含まれているのが見てとれる。

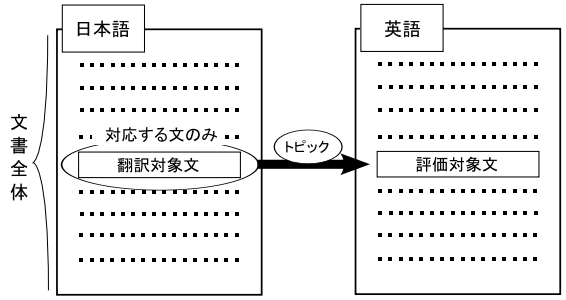

図 1　適応のイメージ

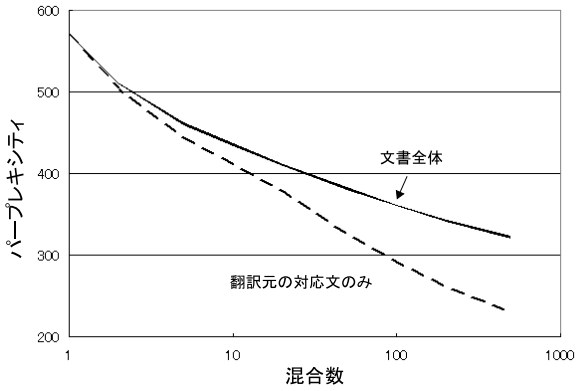

図 2　適応範囲の変化によるパープレキシティの比較

5.3　翻訳システム性能評価実験

5.3.1　リランキング法による統合

まず推定した言語モデル、翻訳モデル及び Moses デコーダを用いて development データによる MERT を行い、翻訳で必要となる特徴関数のモデルパラメータを推定する。次に推定したモデルパラメータを用いて、development データ、及び test データの 100-best を出力する。その後 development データによってリランキングの際に必要となるパラメータ $(\lambda_{\mathrm{sen}}, \lambda_{\mathrm{wp}})$ を推定し、最後にテストデータを用いてリランキングを行う。言語横断 LDA モデルのトピック数は 500 で固定した。

テストデータを用いて英日方向の翻訳を行ったところ、図 3 のような結果が得られた。文尤度を用いたリランキングではベースラインである Moses に対して 0.34% の改善を示したが、文書尤度の場合とではあまり差が見られなかった。

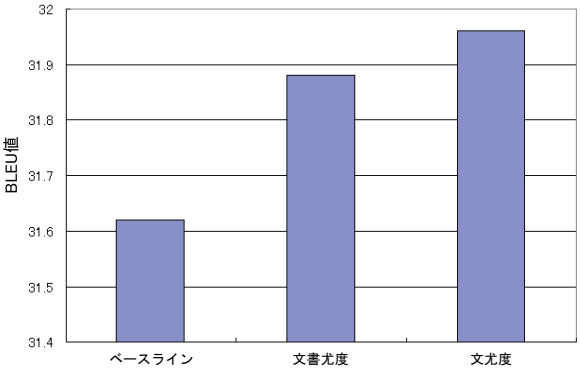

図 3　適応範囲の変化による BLEU の比較

翻訳結果の例を以下に示す。ベースラインや文書尤度の方法ではみられなかった「蒸着」という単語が現れていることが確認できる。これは文尤度を用いた方法がより鋭くトピックを捉えたためと考えられる。

原言語文　widely used among the methods of depositing silicon or silicon compound thin films is an evaporation method using a cvd process .

正解文　シリコン または シリコン 化合 物 薄膜 を 成 膜 する 方法 として 、 CVD 法 を 用い た 蒸着 方法 が 広く 用い られ て いる 。

ベースライン　CVD 法 を 用い て シリコン または シリコン 化合 物 薄膜 の 成 膜 は 、 蒸発 法 は いくつか の 方法 が 広く 用い られ て いる 。

文書尤度を用いたリスコアリング　CVD 法 を 用い て シリコン または シリコン 化合 物 薄膜 の 成 膜 は 、 蒸発 法 は いくつか の 方法 が 広く 用い られ て いる 。

文尤度を用いたリスコアリング　CVD 法 を 用い て シリコン または シリコン 化合 物 薄膜 の 成 膜 は

WEB	依頼, キーワード, 名前, アイコ, 文章, 手続き, サーバ, クライアント, 課金, 辞書
	menu, task, job, messages, password, script, icon, provider, inquiry, sentence
画像	シェーディング, ｄｐｉ, 視線, ポリゴン, 画素, ベクトル, 映像, ビデオカメラ, ＤＣＴ, 補間
	coding, gamma, blanking, interpolation, dct, pictures, gray, predictive, luminance, picture
機械工学	内圧, 弁, 噴射, 点火, 燃焼, 冷凍, ＮＯｘ, アクセル, 内燃, 吸気
	manifold, injector, spark, ignition, purge, exhaust, deceleration, valve, evaporation, engine

表 3　言語横断 LDA モデルが識別したトピックの例

、蒸着 法 は いくつか の 方法 が 広く 用い られ て いる 。

しかし原言語文が短くトピックが捉えにくい場合には、文尤度を用いたリランキング法では翻訳精度の悪化を招いてしまう。以下の例では原言語文に現れる単語からはトピックを推定することが難しく、「outline」を「輪郭」、「issue」を「課題」と翻訳してしまっている。文書尤度を用いた場合では、原言語文書全体からトピックを捉えているため、文尤度のような誤りが起こらなかったと考えられる。

原言語文　fig . 40 shows an outline of the get next issue to be executed in fig . 38 .

正解文　図 ４０ は 、 図 ３８ で 実行 する ｇｅｔ － ｎｅｘｔ 発行 の 概要 を 示し た もの で ある 。

ベースライン　図４０は、図３８の 発行 ｇｅｔ ＿ ｎｅｘｔ の 概要 で 実行 さ れる 。

文書尤度を用いたリスコアリング　図 ４０ は 、 図 ３８ の 発行 ｇｅｔ ＿ ｎｅｘｔ の 概要 で 実行 さ れる 。

文尤度を用いたリスコアリング　図 ４０ は、図 ３８ の 次 に 実行 す べき 課題 の 輪郭 を 得る 。

5.3.2　log-linear model による統合

従来の統計的機械翻訳システムの特徴関数として新たに LDA モデルを加え、日英方向の翻訳精度評価実験を行った。各々特徴関数のモデルパラメータは、考慮するトピックの数ごとにそれぞれ MERT によって決定した。

横軸はトピックの数、縦軸は BLEU 値を示す。トピック数 100 において、BLEU は低い値を示した。この理由としては、MERT を行った際に局所解に陥った可能性が考えられる。

ベースライン、提案手法それぞれの翻訳結果を観察すると、両システムの間には大きな差異がみられた。以下より、いくつかの翻訳例を紹介する。本提案手法はより正確な単語候補の選出に貢献するものであるため、ここでは特に、ベースラインと比べて正しい単語が含まれているかについて検討を行う。

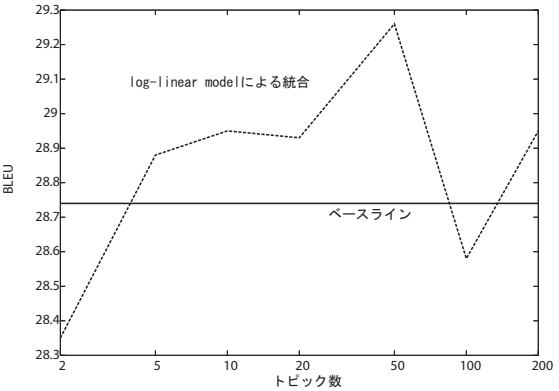

図 4　トピック数の変化による BLEU の比較

原言語文　on the other hand , when the spindle motor 3 is started at one of current values i2 , i3 , and i4 , the control skips step s4 .

正解文　なお、電流値 i2,i3,i4 のいずれかでスピンドルモータ 3 が 起動 した場合には、ステップ s4 はスキップされる。

ベースライン　一方、スピンドルモータ 3 の一方の電流値 i2,i3,i4, スキップ制御が 開始 されると (ステップ s4)、

提案手法　一方、スピンドルモータ 3 が 起動 されると、スキップ s4 で、制御電流値の i2,i3,i4 である。

原言語文の「the spindle motor 3」からトピックを捉えることにより、提案手法では「start」が「開始」よりもふさわしい「起動」に変化したと考えられる。

原言語文　the output terminal of the amplifier 41 is connected to the inverted input terminal of a comparator 46 .

正解文　前記 増幅器 ４１ の 出力 端子 は、コンパレータ ４６ の 反転 入力 端子 に 接続 さ れ て いる 。

ベースライン　増幅器 ４１ の 反転 入力 端子 に は 、

比較 部 ４６ の 出力 端子 が 接続 さ れ て いる 。

提案手法 増幅 回路 の 出力 端子 と コンパレータ ４６ の 反転 入力 端子 に 接続 さ れ て いる 。

これも、LDA モデルが機能した例といえる。原言語文のトピックを捉えることによって、ベースラインでは「比較部」と訳されていた英単語「comparator」が、提案手法では「コンパレータ」と正しく訳された。

一方で、LDA モデルが機能したがために、ベースラインより悪化する例もみられた。

原言語文 the reader unit 1 is further provided with an operation unit 115 for effecting various settings on the composite image input / output apparatus .

正解文 また 、 リーダ 部 １ に は 、 本 複合 画像 入出力 装置 に対して 各種 設定 を 行う ため の 操作 部 １１５ が 設け られ て いる 。

ベースライン また 、 リーダ 部 １ が 設け られ た 操作 部 １１５ の 複合 画像 入出力 装置 の 各種 設定 を 行う 。

提案手法 また 、 リーダ 部 １ の 複合 画像 入出力 装置 の 各種 設定 を 行う ため の 演算 部 １１５ が 設け られ て いる 。

ベースラインでは英単語「operation」が正しく「操作」に訳されていたが、提案手法ではトピックを捉えたことにより、「演算」に変化してしまったものと考えられる。

6 まとめと今後の課題

言語横断 LDA モデルを用いた統計的機械翻訳システムの実装と評価を行った。まず最初に、適応範囲を文書全体とした場合と翻訳元の対応する文のみにした場合とで言語横断 LDA モデルの性能比較を行った。その結果、言語横断 LDA モデルは翻訳元の対応する文に絞って適応した場合により良い性能を示すことがわかった。次にリランキング法、log-linear model による方法の 2 通りで言語横断 LDA モデルを用いた統計的機械翻訳システムを実装し、翻訳精度の評価を行った。その結果、双方ともにベースラインをわずかに上回る性能を示した。

特許文対訳コーパスはドメインが狭いため、トピックモデルによる改善余地は比較的小さくなる傾向にある。本実験では性能の大幅な改善には至らなかったが、今後多様なドメインを含むコーパスを用いることで、より大幅な性能の改善が見込まれる。

参考文献

Brown, P. F., Pietra, V. J. D., Pietra, S. A. D., and L.Mercer, R. (1993). “The mathematics of statistical machine translation:parameter estimation.” In *Computational Linguistics*, Vol. 19,2, pp. 264–311.

Littman, M. L., Dumais, S. T., and Landauer, T. K. (1998). “Automatic Cross-Language Information Retrieval using Latent Semantic Indexing.” In *CROSSLANGUAGE INFORMATION RETRIEVAL*.

M.Blei, D., Y.Ng, A., and I.Jordan, M. (2003). “Latent dirichlet allocation.” In *Journal of Machine Learning Research*, Vol. 3, pp. 933–1022.

Och, F. J. (2003). “Minimum Error Rate Training in Statistical Machine Translation.” In *Proceedings of the 41th Annual Meeting of the Association for Computational Linguistics*, pp. 160–167.

Tam, Y.-C., lan Lane, and Schultz, T. (2007). “Bilingual-LSA Based LM Adaptation for Spoken Language Translation.” In *Proceeding of the 45th Annual Meeting of the Association of Computational Linguistics*, pp. 520–527.

Zhao, B. and Xing, E. P. (2008). “HM-BiTAM: Bilingual Topic Exploration, Word Alignment, and Translation.” In *Advances in Neural Information Processing Systems*, Vol. 20, pp. 1689–1696.

西尾拓 (2008). “トピック言語モデルを用いた統計的機械翻訳システム.” 筑波大学第三学群情報学類卒業研究.