決定木学習を用いた翻訳システム自動選択手法の検討

安田圭志[†‡]　菅谷史昭[†]　竹澤寿幸[†]　山本誠一[†]　柳田益造[‡]　([†]ATR音声言語コミュニケーション研究所,[‡]同志社大)

6M−07

A Study on Automatic Selection Method of Machine Translation System using Decision Tree

Keiji YASUDA[†‡], Fumiaki SUGAYA[†], Toshiyuki TAKEZAWA[†], Seiichi YAMAMOTO[†], Masuzo YANAGIDA[‡]

([†]ATR Spoken Language Translation Research Laboratories,[‡]Doshisha University)

1　はじめに

これまでに様々な機械翻訳システムが研究・開発されているが，すべてのドメインや表現の形式に対応できる万能なシステムは未だ開発されていない．一方，ある特定のドメインや，表現の形式に対して有効なシステムが存在する．このような状況で，入力毎に最も有効な翻訳システムを自動的に選択する技術が確立できれば，それぞれの翻訳システムを相補的に利用でき，ドメインおよび表現の形式に対して頑健な統合システムを構築することが出来る．

我々の研究所では，EBMT (Example Based Machine Translation) [1]と，TDMT (Transfer Driven Machine Translation) [2]と呼ばれる2つの翻訳システムの研究・開発を行っている．これらのシステムは全く異なった手法で翻訳を行っており，それぞれの翻訳システムが持つ特性も異なる．

EBMTでは翻訳の際に，コーパス内の用例の内，入力文と最も類似した用例だけを用いる．そのため，入力文とコーパス内の用例との類似性が高ければ，高品質の翻訳を生成できるが，類似性が低い場合，翻訳品質が著しく低下することがある．EBMTで用いているコーパスは，慣用的な表現を多く含んでいるため，このような入力に対しては，高品質の翻訳が期待できる．一方，TDMTはコーパス全体から学習された部分的な情報である変換知識を用いて翻訳を行うため，入力文とコーパス内の各文との表現の違いに対する頑健性を持っている．

本論文では，決定木学習を用いたこれら2つの翻訳システムの自動選択手法について述べる．

2　EBMT

EBMTにおけるDP距離を以下に定義する．

$$P_{DP} = \frac{I + D + 2\sum D_{semantic}}{L_{input} + L_{example}} \qquad (1)$$

ここで L_{input} と $L_{example}$ はそれぞれ，入力とコーパス内の用例の単語数を表している．また，Iは入力文とコーパス内の用例とをDPマッチングで比較した時の挿入語数を表し，D は同様に比較した場合の脱落語数を表している．$D_{semantic}$は単語間の意味距離である．ここで定義したDP距離は，入力文と用例とがどの程度類似しているかを表しており，0から1までの値をとる．完全一致の場合はDP距離が0となる．

EBMTでは，入力文と対訳コーパス内の各原言語用例とのDP距離を求め，DP距離が最小となった対訳用例のみを翻訳に用いる．入力文と，原言語用例とが完全一致した場合は，目的言語用例をそのまま出力し，完全一致しなかった場合は，目的言語用例の単語を部分的に置き換えて出力する．

3　TDMT

TDMTでは，構文構造の基本単位である構成素境界パターン（以下，パターンと呼ぶ）を単位とした変換により翻訳が行われる．パターンは，変項と構成素境界から成り，パターンごとに照合する翻訳例を収集，編集することにより，原言語表現と目的言語表現を対応づける変換知識を作る．TDMTにおいて，構文解析は以下に定義する構文スコアが最小となるように行われる．

$$P_{TDMT} = \arg\min_{\{p_j\}\in P} \sum_i S(b_i, p_j) \qquad (2)$$

ここで，PはTDMTが持つパターンの集合を表し，b_iは入力文を解析して得られる部分文を表している．またSはパターンと部分文との意味距離を表している．構文スコアは0以上の値をとる．TDMTでは，複数の文からなる長い発話や，話し言葉に見られる文法規範から逸脱した入力等が原因で，変換を適用する部分文の組み合わせで，一つの構文木を作れなかった場合は，最も整合性のとれた部分に分割し，部分毎に翻訳を行う．

4　決定木学習

決定木によるシステム選択の際に用いるパラメータは，2節で述べたEBMTのDP距離と，3節で述べたTDMTの構文スコアである．また，決定木ルール生成ツールとして，機械学習の分野でよく知られるC5.0を用いた．

学習セット，および評価セットについては，各入力発話に対する各システムの出力を，評価者が一対比較により優劣を決定している．一対比較の結果は，表1に示すように，EBMT優位（表1中のEBMT won），TDMT優位（表1中のTDMT won），同等（表1中の

Even）のいずれかである．また，表1中の数字は実際の文数を表している．元々の学習セットは508文，評価セットは510文からなるが，同等と判断されたデータについては，自動選択を行う場合に，どちらが選択されても翻訳品質に影響がでないため，学習および評価には用いていない．実際の学習および評価には用いたのは，一対比較により優劣がついたデータ（表1中の下線部）であり，学習セットは364文，テストセットは375文となっている．

表2に，決定木学習データの構造を示す．

表1 データの内訳

	Learning Set	Test Set
EBMT won	189	189
TDMT won	175	186
Even	144	135

表2 決定木学習データの構造

Input		Teacher
EBMT (DP Distance)	TDMT (Syntax Score)	Result of the Paired Comparison Method
1	4.57	TDMT
0	0	TDMT
0.2	25	EBMT
0.2	0.67	TDMT
0	0.83	EBMT
：	：	：
0	1.33	EBMT

表3に，学習された決定木により自動選択を行った結果を示す．表中の数字は，テスト文の数を表しており，括弧内の数字は，テストセット全体（375文）に対する割合を表している．また，表3の下線部が正しく選択された結果を表しており，正しく選択される割合は，76%となっている．

表3 決定木による選択の結果

		Human Selection	
		EBMT	TDMT
Decision Tree	EBMT	135(36%)	36(9.6%)
	TDMT	54(14.4%)	150(40%)

5 翻訳システムの性能評価

各システム（EBMT単体，TDMT単体，EBMTとTDMTを決定木により選択した場合,人がEBMTとTDMTの選択を行った場合）の翻訳性能の評価には，TOEICスコア685のTOEIC受験者による翻訳と，各システムによる翻訳との一対比較[3]を行っている．図1に評価結果を示す．一対比較の結果は，システムによる翻訳がTOEIC受験者による翻訳より優れている場合（図1中のMT won），システムによる翻訳とTOEIC受験者による翻訳が同等の場合（図1中のEven），TOEIC受験者による翻訳が翻訳システムによる翻訳がより優れている場合（図1中のHuman won）の3種類の結果が得られるが，それらを集計した結果が図1である．図1の横軸において，左からEBMT単体での評価結果，TDMT単体での評価結果，決定木により選択を行った場合の評価結果，人が選択を行った場合（言い換えれば全く誤りのないシステム選択が出来た場合）の評価結果を表している．

図1において，EBMT単体とTDMT単体では，それぞれの評価結果の割合がほぼ同じであり，ほぼ同等の翻訳性能であると言える．決定木により選択を行った場合，人手により選択を行った場合よりは劣るが，それぞれのシステム単体の場合と比較し，システム優位となる文の割合が増加し，TOEIC受験者優位となる文の割合が減少している．このことから，決定木により選択を行いEBMTとTDMTを併用した場合，それぞれのシステム単体より翻訳性能が改善されていることが分かる．

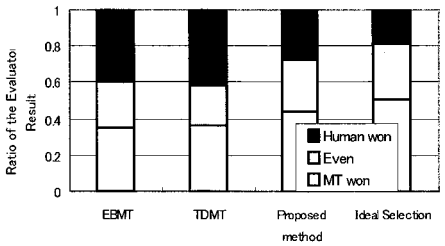

図1 一対比較による評価結果

6 むすび

決定木学習を用いた翻訳システムの自動選択手法について検討した．ATRで研究・開発されたEBMTとTDMTの2つの翻訳システムを用いて，決定木学習による翻訳システム自動選択を行った結果，76%の精度で，正しい選択を行えた．また自動選択を行った場合の翻訳性能は，それぞれの翻訳システム単体よりも改善されることが示された．

謝辞 本研究の一部は，同志社大学学術フロンティア事業の援助を受けている．

文献

[1] E. Sumita, “Example-based machine translation using DP-matching between word sequence”, Proc. ACL-2001 Workshop on Data-Driven Methods in Machine Translation, pp.1-8, 2001.

[2] 古瀬，山田，山本 “頑健な多言語翻訳のための不適格入力の分割処理”，情報処理学会論文誌，Vol. 42 , No. 5, pp.1223-1231, 2001

[3] 菅谷，竹澤，横尾，山本 “音声翻訳システムと人間との比較による音声翻訳能力評価手法の提案と比較実験”, 信学論, Vol. J84-D-Ⅱ, pp.2362-2370, No.11, 2001