

文献紹介ゼミ

林 秀治

紹介する文献

- System Combination for Grammatical Error Correction
- Raymond Hendy Susanto, Peter Phandi, Hwee Tou Ng
- EMNLP 2014. p951-962

概要

- 文法の誤り訂正の手法がいくつか提案されたが、それらは長所と短所を持つ
- それらの多くはclassificationとSMTに基づく
- 訂正の質を改善するために、classificationベースとSMTベースのシステムの出力を組み合わせる

背景

- 文法誤り訂正(GEC)は、第二言語学習者が書いた文の文法誤りを検出・訂正するタスク
- 最も有力な手法は、特定の種類の誤りを扱う分類器を多数組み合わせるclassificationベースと”悪い”文章から”良い”文章への翻訳タスクとして扱うSMTベース

背景

- それぞれの手法に長所と短所

classification

- 独立した分類器を使うので、個々の誤りに焦点を当てることができ、誤りの種類に合わせた分類器を構築できる
- 誤りの種類ごとに分類器を構築しなければならないので、全体のシステムとして複雑になる

SMT

- 文全体として、最も良い文に訂正できる
- より良いシステムを構築するには、十分に大きな訓練データが必要

システムの構成

- classificationベースのシステムとSMTベースのシステムを2つずつ、計4つのシステムを使用する

•classificationベースシステム

- 異なるパイプラインシステムP1,P2を構築
- マルチクラス信頼重み付け分類器を訂正に使用

Step	Pipeline 1 (<i>P1</i>)	Pipeline 2 (<i>P2</i>)
1	Spelling	Spelling
2	<u>Noun number</u>	<u>Article</u>
3	Preposition	Preposition
4	Punctuation	Punctuation
5	<u>Article</u>	<u>Noun number</u>
6	Verb form, SVA	Verb form, SVA

SMTベースシステム

- フレーズベースのSMTを基にしている
- 翻訳モデルを訓練するために二つの異なるパラレルコーパスを使用
 - NUCLE: 英語講師によって校正された、シンガポールの大学生によって書かれた文
 - Lang-8
- 1つのフレーズテーブル(S1)と2つのフレーズテーブル(S2)をこの2つのコーパスで訓練した、2つのSMTを使用

システムの統合

- システムの出力を統合するために、MEMTを使用
- MEMTは各システムの出力のペアを統合するためにMETEORを使用
 - METEORは完全一致、同一語幹、同義語、unigramの言い換えを識別可能

実験データ

- システムを訓練するためにCoNLLのofficial training dataのNUCLEver.3.2を使用
 - このコーパスの文法誤りは28種に分類される
- "Lang-8 Corpus of Learner English v1.0"も使用
- 言語モデルとn-gramを集めるためにEnglish Wikipediaを使用
- 全てのシステムはCoNLL2013のテストデータでチューニングし、2014のテストデータでテスト

実験データ

Data set	# sentences	# source tokens
NUCLE	57,151	1,161,567
Lang-8	1,114,139	12,945,666
CoNLL-2013	1,381	29,207
CoNLL-2014	1,312	30,144
English Wikipedia	86,992,889	1,778,849,655

評価

- システムの性能は精度、再現率、 $F_{0.5}$ (精度が再現率の2倍の重み)

パイプラインシステム

- 品詞タグ付けと係り受け解析にClearNLPを使用し、チャンキングにOpenNLPを使用
- 単数形と複数形の語を作るのにWordNetを使用
- 各分類器はNUCELとLang-8にマルチクラス重み付け学習を使用して学習
- 分類器の閾値は、開発データセット上で単純なグリッド検索を各分類器に使用して調整

SMTシステム

- Mosesを使用
- NUCLEとLang-8の平行コーパスを使用して学習
- システム内の並び替えモデルは使用しない

複合システム

- システムの出力を結合するために、HeafieldとLavieのオープンソースのMEMTを使用
- パラメータはHeafieldとLavieが推奨する値に設定
- 言語モデルにWikipediaの5-gramを使用

複合システムの調整

- 開発データセットを使って調整を行う
- テストデータをパイプラインとSMTに入力し、各出力をMETEORを使って統合する
- BLUEに基づいた特徴の重みは、ZMERTを使って調整する
- このプロセス5回繰り返し、最終的なシステムで最高のスコアを達成する重みを使用

結果

- CoNLL-2014のテストデータを使った実験の結果

System	P	R	$F_{0.5}$	System	P	R	$F_{0.5}$
Pipeline				Combined			
P1	40.24	23.99	35.44	P1+S1	53.85	17.65	38.19
P2	39.93	22.77	34.70	P2+S2	56.92	16.22	37.90
SMT				P1+P2+S1+S2	53.55	19.14	39.39
S1	57.90	14.16	35.80	Top 4 Systems in CoNLL-2014			
S2	62.11	12.54	34.69	CAMB	39.71	30.10	37.33
				CUUI	41.78	24.88	36.79
				AMU	41.62	21.40	35.01
				POST	34.51	21.73	30.88

結果

- パイプラインはSMTよりも低い精度だが、再現率は高い

System	P	R	$F_{0.5}$	System	P	R	$F_{0.5}$
Pipeline				Combined			
P1	40.24	23.99	35.44	P1+S1	53.85	17.65	38.19
P2	39.93	22.77	34.70	P2+S2	56.92	16.22	37.90
SMT				P1+P2+S1+S2	53.55	19.14	39.39
S1	57.90	14.16	35.80	Top 4 Systems in CoNLL-2014			
S2	62.11	12.54	34.69	CAMB	39.71	30.10	37.33
				CUUI	41.78	24.88	36.79
				AMU	41.62	21.40	35.01
				POST	34.51	21.73	30.88

結果

- パイプラインは、順序による影響が大きく、冠詞訂正の前に名詞を訂正したほうが結果が良い(P1が名詞が先)
 - 例: a houses → a house/houses

System	<i>P</i>	<i>R</i>	<i>F</i> _{0.5}	System	<i>P</i>	<i>R</i>	<i>F</i> _{0.5}
Pipeline				Combined			
P1	<u>40.24</u>	<u>23.99</u>	<u>35.44</u>	P1+S1	53.85	17.65	38.19
P2	<u>39.93</u>	<u>22.77</u>	<u>34.70</u>	P2+S2	56.92	16.22	37.90
SMT				P1+P2+S1+S2	53.55	19.14	39.39
S1	57.90	14.16	35.80	Top 4 Systems in CoNLL-2014			
S2	62.11	12.54	34.69	CAMB	39.71	30.10	37.33
				CUUI	41.78	24.88	36.79
				AMU	41.62	21.40	35.01
				POST	34.51	21.73	30.88

結果

- P1とS1、P2とS2の性能は同等だが、それらの結果を統合した違いは統計的に優位ではない

System	P	R	$F_{0.5}$	System	P	R	$F_{0.5}$
Pipeline				Combined			
P1	40.24	23.99	<u>35.44</u>	P1+S1	53.85	17.65	<u>38.19</u>
P2	39.93	22.77	<u>34.70</u>	P2+S2	56.92	16.22	<u>37.90</u>
SMT				P1+P2+S1+S2	53.55	19.14	39.39
S1	57.90	14.16	<u>35.80</u>	Top 4 Systems in CoNLL-2014			
S2	62.11	12.54	<u>34.69</u>	CAMB	39.71	30.10	37.33
				CUUI	41.78	24.88	36.79
				AMU	41.62	21.40	35.01
				POST	34.51	21.73	30.88

結果

- より良い結果の組み合わせ(P1+S1)は、他の組み合わせ(P2+P2)よりも統計的に優位に優れている

System	<i>P</i>	<i>R</i>	<i>F</i> _{0.5}	System	<i>P</i>	<i>R</i>	<i>F</i> _{0.5}
Pipeline				Combined			
P1	40.24	23.99	35.44	P1+S1	53.85	17.65	38.19
P2	39.93	22.77	34.70	P2+S2	56.92	16.22	37.90
SMT				P1+P2+S1+S2	53.55	19.14	39.39
S1	57.90	14.16	35.80	Top 4 Systems in CoNLL-2014			
S2	62.11	12.54	34.69	CAMB	39.71	30.10	37.33
				CUUI	41.78	24.88	36.79
				AMU	41.62	21.40	35.01
				POST	34.51	21.73	30.88

結果

- 全ての組み合わせはCoNLLのTOPよりも良い
- 個々の性能は3番目と同等ぐらいなので、この結果は重要である

System	P	R	$F_{0.5}$	System	P	R	$F_{0.5}$
Pipeline				Combined			
P1	40.24	23.99	35.44	P1+S1	53.85	17.65	38.19
P2	39.93	22.77	34.70	P2+S2	56.92	16.22	37.90
SMT				P1+P2+S1+S2	53.55	19.14	<u>39.39</u>
S1	57.90	14.16	35.80	Top 4 Systems in CoNLL-2014			
S2	62.11	12.54	34.69	CAMB	39.71	30.10	<u>37.33</u>
				CUUI	41.78	24.88	36.79
				AMU	41.62	21.40	35.01
				POST	34.51	21.73	30.88