

文献紹介ゼミ

林 秀治

紹介する文献

- Comparison, Selection and Use of Sentence Alignment Algorithms for New Language Pairs
- Anil Kumar Singh, Samar Husain
- ACL Workshop 2005. p99-106

概要

- いくつかのアルゴリズムが文のアライメントに利用できるが、異なる条件下での比較ができない
- ほとんどの場合、アライメントアルゴリズムのパフォーマンスに影響を与える要因は、評価に考慮されない
- 文アライメントアルゴリズムのパフォーマンスに関するよりよい評価値を与える評価方法を提案
- 4つの手法の比較を行う

背景

- いくつかの文アライメントの手法が提案されていて、ほとんどのものが精度98から99%と非常に優れている
- これらの数値はあるコーパスに対しての指標となるが、様々な条件下での性能の変化は考慮されていない

背景

- ほとんどの情報は、評価が行われた条件について与えられていて、これはアルゴリズムがすべての条件化でその精度・リコールで動作するような印象を与える
- しかし、これらのアルゴリズムの性能は、条件によって大きく変化する
- 文アライメントアルゴリズムの性能のよりよい評価方法を提案

文アライメント手法

- アライメント手法は文の長さベース、単語の対応、複数手法の組み合わせなどに分類することができる

パフォーマンスの要因

- Source Language(SL)とTarget language(TL)の言語的距離がパフォーマンスに影響を与える
 - 英語-ヒンディー語よりも英語-ドイツ語のほうがよい

評価

- loacl(各々の実行)だけでなく、global(全ての実行)でパフォーマンスを確認した
- 項目は以下の通り
 - 精度(loacl/global)
 - 再現率(loacl/global)
 - F値(loacl/global)
 - F値の95%信頼区間(global)
 - ランタイム(local)

実験方法

- 異なる条件下での性能を見るために、下記の項目を考慮する
 - コーパスタイプ
 - コーパスのサイズ
 - SLとTLのコーパスのサイズの差
 - ノイズ

実験方法

- 4つのアルゴリズムを比較
 - Brn : Brownの文の長さに基づく方法
 - GC : ChurchとGaleの文の長さに基づく方法
 - Mmd : Melamedのジオメトリックな対応に基づく方法
 - Mre : Mooreのtwo-pass法 (語の数と対応)

コーパスの種類

- 3つの違うタイプで言語ペア、サイズが同じコーパスを使用
 - EMILLE
 - ERDC
 - India Today
- 最小のコーパスのサイズと同じ2500文を使用

コーパスサイズ

- コーパスサイズによる違いを見るために、ERDCの500、1,000、5,000、10,000文を使用

ノイズとSL/TLコーパスのサイズ

- ノイズとSLとTLのコーパスサイズの差の影響を確認するために3つの事例を用意
 - 1:ノイズのない同じサイズ
 - 2:ノイズがある同じサイズ
 - 3:ノイズがある異なるサイズ
- ノイズはほかの無関係なコーパスの文
- 2では10%、3ではSLの5%TLの15%がノイズ

評価結果

- 各コーパスタイプの結果

Type		Clean, Same Size				Noisy, Same Size				Noisy, Different Size			
		Brn	GC	Mmd	Mre	Brn	GC	Mmd	Mre	Brn	GC	Mmd	Mre
EMILLE	P	99.3	99.1	85.0	66.8	85.5	87.4	38.2	66.2	87.2	86.5	48.0	65.5
	R	96.0	93.0	80.0	63.2	80.4	80.0	36.2	58.0	81.2	79.1	46.5	57.4
	F	97.6	96.0	82.0	64.9	82.8	83.5	37.2	61.8	84.0	82.6	47.3	61.2
	T	23	23	261	45	47	44	363	64	25	25	413	47
ERDC	P	99.6	99.5	94.2	100.0	85.4	84.4	48.0	96.5	84.6	85.5	50.9	97.7
	R	99.0	99.1	92.7	97.0	81.7	80.6	46.7	78.9	80.5	81.3	49.8	79.1
	F	99.3	99.3	93.4	98.4	83.5	82.4	47.3	86.8	82.5	83.3	50.3	87.1
	T	31	29	1024	85	92	90	2268	124	55	52	3172	101
India Today	P	91.8	93.9	76.4	99.5	71.5	76.7	49.7	94.4	73.6	75.5	51.7	93.4
	R	81.0	83.0	70.6	81.5	61.0	65.5	47.6	67.5	62.4	64.4	50.1	62.6
	F	86.1	88.1	73.4	89.6	65.8	70.7	48.6	78.7	67.6	69.5	50.9	75.0
	T	32	32	755	91	96	101	2120	159	60	68	987	134
English-French	P	100.0	100.0	100.0	100.0	87.4	87.5	77.2	95.2	91.2	93.3	77.7	96.6
	R	100.0	99.3	100.0	99.3	85.5	84.3	81.7	84.6	83.2	83.7	82.6	83.0

P: Precision, *R*: Recall, *F*: F-Measure, *T*: Runtime (seconds)

評価結果

- 各コーパスタイプの結果

Type		Clean, Same Size				Noisy, Same Size				Noisy, Different Size			
		Brn	GC	Mmd	Mre	Brn	GC	Mmd	Mre	Brn	GC	Mmd	Mre
EMILLE	P	99.3	99.1	85.0	66.8	85.5	87.4	38.2	66.2	87.2	86.5	48.0	65.5
	R	96.0	93.0	80.0	63.2	80.4	80.0	36.2	58.0	81.2	79.1	46.5	57.4
	F	97.6	96.0	82.0	64.9	82.8	83.5	37.2	61.8	84.0	82.6	47.3	61.2
	T	23	23	261	45	47	44	363	64	25	25	413	47
ERDC	P	99.6	99.5	94.2	100.0	85.4	84.4	48.0	96.5	84.6	85.5	50.9	97.7
	R	99.0	99.1	92.7	97.0	81.7	80.6	46.7	78.9	80.5	81.3	49.8	79.1
	F	99.3	99.3	93.4	98.4	83.5	82.4	47.3	86.8	82.5	83.3	50.3	87.1
	T	31	29	1024	85	92	90	2268	124	55	52	3172	101
India Today	P	91.8	93.9	76.4	99.5	71.5	76.7	49.7	94.4	73.6	75.5	51.7	93.4
	R	81.0	83.0	70.6	81.5	61.0	65.5	47.6	67.5	62.4	64.4	50.1	62.6
	F	86.1	88.1	73.4	89.6	65.8	70.7	48.6	78.7	67.6	69.5	50.9	75.0
	T	32	32	755	91	96	101	2120	159	60	68	987	134
English-French	P	100.0	100.0	100.0	100.0	87.4	87.5	77.2	95.2	91.2	93.3	77.7	96.6
	R	100.0	99.3	100.0	99.3	85.5	84.3	81.7	84.6	83.2	83.7	82.6	83.0

P: Precision, *R*: Recall, *F*: F-Measure, *T*: Runtime (seconds)

評価結果

- コーパスサイズによる違い

Size		Clean, Same Size				Noisy, Same Size				Noisy, Different Size			
		Brn	GC	Mmd	Mre	Brn	GC	Mmd	Mre	Brn	GC	Mmd	Mre
500	P	99.2	99.2	93.9	99.8	75.4	78.2	57.4	94.3	83.5	87.2	45.4	92.4
	R	98.8	98.8	91.8	95.0	71.0	73.4	56.8	70.0	77.0	80.8	44.8	70.8
	F	99.0	99.0	92.8	97.3	73.1	75.7	57.1	80.4	80.1	83.9	45.1	80.2
	T	9	9	126	14	10	10	148	13	10	10	181	14
1000	P	99.3	99.6	96.4	100.0	84.6	84.6	67.8	96.8	82.2	84.0	47.3	95.1
	R	98.9	99.4	95.1	96.3	81.4	82.2	68.4	73.7	76.3	78.7	46.1	72.7
	F	99.1	99.5	95.7	98.1	83.0	83.4	68.1	83.7	79.1	81.2	46.7	82.4
	T	13	13	278	29	24	23	335	34	15	15	453	30
5000	P	99.8	99.8	93.2	99.9	88.5	88.6	56.1	98.5	85.9	86.6	57.6	97.8
	R	99.4	99.5	91.6	98.2	83.2	83.3	54.9	86.0	81.7	81.3	56.7	86.3
	F	99.6	99.7	92.4	99.1	85.7	85.9	55.4	91.8	83.7	83.9	57.2	91.7
	T	54	53	3481	186	199	185	5248	274	185	174	3639	275
10000	P	99.8	99.9	93.2	100.0	88.0	88.9	59.6	98.5	86.8	88.7	57.2	98.4
	R	99.4	99.6	91.4	98.6	82.9	83.7	58.9	89.9	81.3	82.8	56.2	89.2
	F	99.6	99.7	92.3	99.3	85.4	86.2	59.2	94.0	84.0	85.6	56.6	94.0
	T	102	96	4356	305	370	346	4477	467	345	322	4351	479

評価結果

- コーパスサイズによる違い

Size		Clean, Same Size				Noisy, Same Size				Noisy, Different Size			
		Brn	GC	Mmd	Mre	Brn	GC	Mmd	Mre	Brn	GC	Mmd	Mre
500	P	99.2	99.2	93.9	99.8	75.4	78.2	57.4	94.3	83.5	87.2	45.4	92.4
	R	98.8	98.8	91.8	95.0	71.0	73.4	56.8	70.0	77.0	80.8	44.8	70.8
	F	99.0	99.0	92.8	97.3	73.1	75.7	57.1	80.4	80.1	83.9	45.1	80.2
	T	9	9	126	14	10	10	148	13	10	10	181	14
1000	P	99.3	99.6	96.4	100.0	84.6	84.6	67.8	96.8	82.2	84.0	47.3	95.1
	R	98.9	99.4	95.1	96.3	81.4	82.2	68.4	73.7	76.3	78.7	46.1	72.7
	F	99.1	99.5	95.7	98.1	83.0	83.4	68.1	83.7	79.1	81.2	46.7	82.4
	T	13	13	278	29	24	23	335	34	15	15	453	30
5000	P	99.8	99.8	93.2	99.9	88.5	88.6	56.1	98.5	85.9	86.6	57.6	97.8
	R	99.4	99.5	91.6	98.2	83.2	83.3	54.9	86.0	81.7	81.3	56.7	86.3
	F	99.6	99.7	92.4	99.1	85.7	85.9	55.4	91.8	83.7	83.9	57.2	91.7
	T	54	53	3481	186	199	185	5248	274	185	174	3639	275
10000	P	99.8	99.9	93.2	100.0	88.0	88.9	59.6	98.5	86.8	88.7	57.2	98.4
	R	99.4	99.6	91.4	98.6	82.9	83.7	58.9	89.9	81.3	82.8	56.2	89.2
	F	99.6	99.7	92.3	99.3	85.4	86.2	59.2	94.0	84.0	85.6	56.6	94.0
	T	102	96	4356	305	370	346	4477	467	345	322	4351	479

評価結果

- コーパスサイズによる違い

Size		Clean, Same Size				Noisy, Same Size				Noisy, Different Size			
		Brn	GC	Mmd	Mre	Brn	GC	Mmd	Mre	Brn	GC	Mmd	Mre
500	P	99.2	99.2	93.9	99.8	75.4	78.2	57.4	94.3	83.5	87.2	45.4	92.4
	R	98.8	98.8	91.8	95.0	71.0	73.4	56.8	70.0	77.0	80.8	44.8	70.8
	F	99.0	99.0	92.8	97.3	73.1	75.7	57.1	80.4	80.1	83.9	45.1	80.2
	T	9	9	126	14	10	10	148	13	10	10	181	14
1000	P	99.3	99.6	96.4	100.0	84.6	84.6	67.8	96.8	82.2	84.0	47.3	95.1
	R	98.9	99.4	95.1	96.3	81.4	82.2	68.4	73.7	76.3	78.7	46.1	72.7
	F	99.1	99.5	95.7	98.1	83.0	83.4	68.1	83.7	79.1	81.2	46.7	82.4
	T	13	13	278	29	24	23	335	34	15	15	453	30
5000	P	99.8	99.8	93.2	99.9	88.5	88.6	56.1	98.5	85.9	86.6	57.6	97.8
	R	99.4	99.5	91.6	98.2	83.2	83.3	54.9	86.0	81.7	81.3	56.7	86.3
	F	99.6	99.7	92.4	99.1	85.7	85.9	55.4	91.8	83.7	83.9	57.2	91.7
	T	54	53	3481	186	199	185	5248	274	185	174	3639	275
10000	P	99.8	99.9	93.2	100.0	88.0	88.9	59.6	98.5	86.8	88.7	57.2	98.4
	R	99.4	99.6	91.4	98.6	82.9	83.7	58.9	89.9	81.3	82.8	56.2	89.2
	F	99.6	99.7	92.3	99.3	85.4	86.2	59.2	94.0	84.0	85.6	56.6	94.0
	T	102	96	4356	305	370	346	4477	467	345	322	4351	479

評価結果

- Globalでの結果

		Brn	GC	Mmd	Mre			Brn	GC	Mmd	Mre
Clean, Same Size	L	92.6	93.4	81.4	80.8	Noisy, Different Size	L	74.7	76.4	46.2	71.3
	H	100.0	100.0	96.3	100.0		H	85.6	86.4	55.0	92.0
	P	98.4	98.7	90.3	95.1		P	83.4	84.9	51.2	91.5
	R	96.1	96.1	87.6	90.0		R	77.2	78.3	50.0	74.0
	F	97.2	97.3	88.9	92.4		F	80.1	81.4	50.6	81.6
Noisy, Same Size	L	73.1	75.8	44.1	72.6	Overall	L	81.1	82.4	55.4	80.0
	H	87.5	86.4	62.4	92.3		H	90.4	90.8	73.1	91.0
	P	82.7	84.1	53.8	92.2		P	88.2	89.2	65.1	92.9
	R	77.4	78.4	52.8	74.9		R	83.6	84.3	63.5	79.6
	F	79.8	81.1	53.3	82.5		F	85.7	86.6	64.6	85.5

L and *H*: Lower and higher limits of
95% confidence interval for F-measure

P, *R*, and *F*: Average precision,
recall, and F-measure