

固有表現抽出におけるアノテーション手法の比較

鈴木 雅也^{1,a)} 古宮 嘉那子^{1,b)} 岩倉 友哉^{2,c)} 佐々木 稔^{1,d)} 新納 浩幸^{1,e)}

概要：本稿では、非専門家による固有表現抽出のタスクとしてのアノテーションを題材に、次の2つのアノテーション手法について比較を行った。1つ目は、既存の固有表現抽出器によるアノテーション結果に対し、人手で修正を行う手法である。2つ目は、既存の固有表現抽出器を使用せず、人手でアノテーションを行う手法である。実験では、1テキストに対し、2人のアノテータを割り当て、それぞれの手法でアノテーションを行わせた。そして、アノテーションにかかる時間、タグの一致率、及び、Gold Standard との比較による正解率の各指標について、平均値とどちらか一方でも正解のタグを付与しているならば正解とみなした場合の数値を算出した。本実験の結果から、全体的に見たとき、既存のアノテーション結果を用いた手法は、アノテーションにかかる時間が少なく、一致率や正解率についても高い値になることがわかった。しかし、訓練事例から離れたジャンルのテキストについては、既存の固有表現抽出器を用いない手法の方が良い結果を出しており、このようなテキストについては、既存の固有表現抽出器を用いず、人手のみでアノテーションを行った方が良いと考えられる。

Comparison of Annotating Methods in Named Entity Extraction

MASAYA SUZUKI^{1,a)} KANAKO KOMIYA^{1,b)} TOMOYA IWAKURA^{2,c)} MINORU SASAKI^{1,d)}
HIROYUKI SHINNOU^{1,e)}

1. はじめに

非専門家をアノテータとする、クラウドソーシングによるコーパスへのアノテーションは、安価で速く仕上がることで Snow ら [1] によって明らかとなっている。しかし、アノテーション手法に起因したアノテーションの品質の違いについては、これまで言及されてこなかった。固有表現抽出におけるアノテーションはルールが多く複雑なため、非専門家にとってタグの付け間違いが発生しやすいタスクとなっており、この観点での議論が必要なタスクの1つであると考えられる。そこで、本稿では、固有表現抽出におけるアノテーションを題材として、非専門家の手で高品質なコーパ

スを作成するための手法についての考察を行った。なお、本稿は [2] を元としている。

固有表現抽出におけるアノテーションでのタグの付け間違いを減らすための手法として、既存の固有表現抽出器によるアノテーション結果に対し、人手で修正を行う手法が考えられる。しかし、訓練事例として特定ジャンルのコーパスのみを用いている固有表現抽出器の場合、特にそのジャンル以外のコーパスのアノテーションにおいて、タグの付け間違いが発生することがある。そこで、本研究では、前述の手法と既存の固有表現抽出器を使用せず、人手でアノテーションを行う手法の2つの手法について、アノテーションにかかる時間、タグの一致率、及び、Gold Standard との比較による正解率の各観点から比較することで考察を行った。この際、テキストのジャンルに起因したアノテーションの品質の違いについても考察を行っている。

2. 関連研究

アノテーションに関する先行研究としては、次のようなものが挙げられる。Snow ら [1] は、非専門家によるコーパス

¹ 茨城大学
Ibaraki University

² 富士通研究所
Fujitsu Laboratories Ltd.

a) 13t4038a@vc.ibaraki.ac.jp

b) kanako.komiya.nlp@vc.ibaraki.ac.jp

c) iwakura.tomoya@jp.fujitsu.com

d) minoru.sasaki.01@vc.ibaraki.ac.jp

e) hiroyuki.shinnou.0828@vc.ibaraki.ac.jp

へのアノテーションに関して、アノテーションにかかる時間、アノテーションの品質、及び、コストの観点から、専門家が行った場合と比較することで考察を行った。Alexら [3] は、反復的で agile なアノテーション手法を提案し、既存の線形によるアノテーション手法との比較を行った。van der Plasら [4] は、英語のテンプレートをを用いたフランス語のコーパスへの意味情報の付与を題材に、言語横断的なアノテーションの信頼性について考察を行った。Marcusら [5] は、品詞アノテーションや bracketing といったタスクのための Penn TreeBank を開発するため、既存のアノテーション結果を用いる手法と人手のみで行う手法について比較を行った。しかし、我々が知る限り、非専門家の手で高品質なコーパスを開発するために、既存のアノテーション結果を用いる手法と人手のみで行う手法を比較したという論文は存在しない。

本稿では、固有表現抽出について研究を行った。固有表現抽出とは、固有名詞に数値や時間といった表現を加えた固有表現と呼ばれる要素を自動抽出する技術のことであり、昔から研究が行われてきた。固有表現抽出に関する先行研究としては、次のようなものが挙げられる。橋本ら [6] は現代日本語書き言葉均衡コーパス (BCCWJ)[7]*1 を元にした、拡張固有表現タグ付きコーパスを作成した。徳永ら [8] は、固有表現抽出のタスクにおけるアノテータの視線分析を行った。Sasadaら [9] は、部分的なタグ付きテキストを用いて訓練可能な固有表現抽出器を提案した。また、Information Retrieval and Extraction Exercise (IREX)*2 では、固有表現抽出の共通タスクを行うため、8種類の固有表現タグ (組織名、人名、地名、固有物名、日付表現、時間表現、金額表現、割合表現)、及び、それらと同等に扱われるオプションタグからなる9種類のタグが定義された。しかし、IREX で用いられたのは、新聞コーパスのみであった。

2014年、6領域から構成される Project Next NLP[10], [11], [12]*3 において、前述の拡張固有表現タグ付きコーパスを用いた固有表現抽出のエラー分析が行われた。Ichiyaraら [12] は、既存の固有表現抽出器の性能について調べ、固有表現抽出器の訓練事例から離れたジャンルのテキストにおいて、タグの付け間違いが増加することを示した。本稿では、訓練事例から離れたジャンルのコーパスにおいて、既存のアノテーション結果を用いた手法ではタグの付け間違いが発生する可能性があるということを示す。

本研究では、非専門家の手で高品質なコーパスを作成するため、固有表現抽出のタスクについて、既存のアノテーション結果を用いた手法と人手のみでアノテーションを行う手法の2つの手法による、アノテーションにかかる時間、タグの一致率、及び、Gold Standard との比較による正解率

		手法 X				
		タグ 1	タグ 2	...	タグ n	合計
手法 Y	タグ 1	a_{11}	a_{21}	...	a_{n1}	a_{01}
	タグ 2	a_{12}	a_{22}	...	a_{n2}	a_{02}

	タグ n	a_{1n}	a_{2n}	...	a_{nn}	a_{0n}
	合計	a_{10}	a_{20}	...	a_{n0}	a_{00}

表 1: 2つの手法間で一致したタグの個数

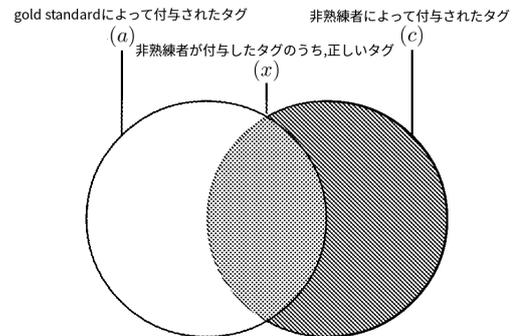


図 1: タグ集合

の評価を行った。

3. アノテーション手法の比較

本稿では、次の2つのアノテーション手法について比較を行った。

- KNP+Manual

既存の固有表現抽出器 KNP[13]*4 によるアノテーション結果に対し、人手で修正を行う。

- Manual

人手のみで固有表現のアノテーションを行う。

また、比較を行うにあたり、それぞれのテキストに対するアノテーションにかかる時間、タグの見かけの一致率とカップ係数、Gold Standard との比較による適合率 (精度)、再現率、及び、F 値を指標として設定した。

2つの手法間で一致したタグの個数が表1で示される時、見かけの一致率とカップ係数はそれぞれ式(1)と式(2)で与えられる。

$$d = \frac{\sum_{i=1}^n a_{ii}}{a_{00}} \quad (1)$$

$$\kappa = \frac{a_{00} \sum_{i=1}^n a_{ii} - \sum_{i=1}^n a_{i0} a_{0i}}{(a_{00})^2 - \sum_{i=1}^n a_{i0} a_{0i}} \quad (2)$$

また、タグ集合が図1のように示されるとき、適合率、再現率、F 値はそれぞれ式(3)、式(4)、式(5)のように与えられる。

$$p = \frac{n(x)}{n(c)} \quad (3)$$

*4 <http://nlp.ist.i.kyoto-u.ac.jp/index.php?KNP>

*1 http://pj.ninjal.ac.jp/corpus_center/bccwj/
*2 <http://nlp.cs.nyu.edu/irex/index-j.html>
*3 <https://sites.google.com/site/projectnextnlp/>

$$r = \frac{n(x)}{n(a)} \quad (4)$$

$$f = \frac{2pr}{p+r} \quad (5)$$

4. 実験

本実験では,ClassA-1^{*5}に分類される136テキストをBC-CWJより抽出して用いた.ClassA-1に分類されるBCCWJのテキストは,Yahoo!知恵袋(OC),白書(OW),Yahoo!ブログ(OY),書籍(PB),雑誌(PM),及び,新聞(PN)の6ジャンルで構成されている.それぞれのジャンルにおけるテキストとそこに含まれるタグの数は表2の通りである.なお,本実験では固有表現抽出器としてKNP Ver.4.16(Linux版)とJUMAN Ver.7.01(Linux版)^{*6}を用いており,前者は訓練事例として新聞コーパスを用いている[13]^{*7}.

被験者は非専門家16人であり,IREXによるアノテーションのルール[14]を読み合わせた後,これに従って9種類の固有表現タグによるアノテーションを行った.この際,全ての被験者のアノテーション結果を集めたときに,それぞれの手法について,2セットのコーパスを構成できるように,被験者は割り当てられた34テキストに対し,それぞれの手法を半分ずつ適用した.また,習熟によるバイアスがかかりにくくするため,被験者を2つのグループに分け,最初に適用する手法をグループごとに変えた.なお,アノテーションの際には,テキストごとのアノテーションにかかる時間の記録も行っており,それを元に手法ごとのアノテーションにかかる平均時間を算出した.また,本実験ではGold standardとして2016年2月1日版のBCCWJコーパス[15]^{*8}を用いている.

模範解答はIREXによるアノテーションのルールに基づき作成した.模範解答にオプションタグが付与されているときはその範囲を超えてタグが付与されていない場合を,それ以外のときはタグとその範囲が模範解答と一致している場合を正解としている.

本実験では,1テキストに対し,2人のアノテータがそれぞれの手法でアノテーションを行ったという条件下で,2人のアノテータの平均正解率と,どちらか一方でも正解のタグを付与しているならば正解とみなした場合の正解率を算出した.後者は,実際にコーパスを作成する際,2人のアノテータによるアノテーション結果を統合して作成することが想定されるため,算出を行った.

これらに加え,機械学習における訓練事例としての品質を確かめるため,それぞれの手法で作成されたコーパスを

訓練事例としたKNPを用いてアノテーションを行った.この際に用いた素性は,形態素,文字種,品詞タグ,分類^{*9},キャッシュ素性,統合素性,及び,格フレーム素性であり,これはオリジナルのKNPと同様である[13].なお,それぞれの手法における2人分のアノテーション結果を結合したものをその手法の訓練事例としており,また,できる限り多くのジャンルのテキストを含むような形で5分割交差検定を行っている.

5. 結果

表3,表4はそれぞれの手法の見かけの一致率とカッパ係数のマイクロ平均とマクロ平均を示しており,表5,表6はそれらをジャンルごとに示したものである.これらにおけるBothは2つの手法を用いた計4人のアノテータによるアノテーション結果の全てのペアを取ったときの一致率の平均を示している.

表7,表8はそれぞれの手法の適合率,再現率,及び,F値のマイクロ平均とマクロ平均を示しており,表9,表10はそれらをジャンルごとに示したものである.これらにおけるKNPはオリジナルのKNPによるアノテーション結果の正解率を,AverageはKNP+ManualとManualの平均を示している.

なお,2つの手法の中でより高い水準を記録した見かけの一致率,カッパ係数,適合率,再現率,及び,F値については太字で示している.また,表11は,それぞれの手法における1テキストあたりのアノテーションにかかる平均時間を示している.

次に,2人のアノテータのうち,どちらか一方でも正解のタグを付与しているならば正解とみなした場合の性能について調べた.表12,表13はそれぞれの手法の適合率,再現率,及び,F値のマイクロ平均とマクロ平均を示しており,表14,表15はそれらをジャンルごとに示したものである.KNPとAverageに関しては,表7~表10と同様である.

これらに加え,それぞれの手法で作成されたコーパスを訓練事例とした固有表現抽出器の性能を調べた.表16,表17はそれぞれの手法で作成されたコーパスを訓練事例としたKNPを用いた場合の適合率,再現率,及び,F値のマイクロ平均とマクロ平均を示しており,表18,表19はそれらをジャンルごとに示したものである.

まず,マイクロ平均について比較する.表7,表16における適合率と再現率,及び,表14における適合率について,有意水準0.05のカイ二乗検定で検定を行った場合,KNPとKNP+Manual,KNPとManual,及び,ManualとKNP+Manualは統計的に有意である.また,正解率におけるジャンルごとのマイクロ平均(表9,表14,表18)の

^{*5} <http://plata.ar.media.kyoto-u.ac.jp/mori/research/NLR/JDC/ClassA-1.list>

^{*6} <http://nlp.ist.i.kyoto-u.ac.jp/?JUMAN>

^{*7} 厳密にはWeb上の記事も訓練事例として用いられているが,本稿では訓練事例としてのウエイトが大きい新聞コーパスをKNPの訓練事例として扱っている.

^{*8} <https://sites.google.com/site/projectnextnlpne/>

^{*9} データとして存在する場合のみ.

ジャンル	テキスト	タグ										合計
		Artifact	Date	Location	Money	Organization	Percent	Person	Time	Optional		
OC	74	44	18	65	9	18	0	6	0	8	168	
OW	8	86	143	147	9	136	33	15	0	26	595	
OY	34	23	61	59	7	64	10	79	3	17	323	
PB	5	32	49	100	0	19	5	174	9	20	408	
PM	2	9	24	36	5	18	1	216	3	1	313	
PN	13	24	166	192	60	123	37	78	22	20	722	
合計	136	218	461	599	90	378	86	568	37	92	2,529	

表 2: ジャンルごとのテキストとそこに含まれるタグの数

手法	見かけの一致率	カッパ係数
KNP+Manual	0.79	0.75
Manual	0.57	0.50
Both	0.64	0.58

表 3: 一致率のマイクロ平均 (全体)

手法	見かけの一致率	カッパ係数
KNP+Manual	0.66	0.48
Manual	0.52	0.29
Both	0.52	0.31

表 4: 一致率のマクロ平均 (全体)

ジャンル	手法	見かけの一致率	カッパ係数
OC	KNP+Manual	0.62	0.54
OC	Manual	0.47	0.34
OC	Both	0.52	0.41
OW	KNP+Manual	0.78	0.73
OW	Manual	0.41	0.28
OW	Both	0.55	0.46
OY	KNP+Manual	0.69	0.63
OY	Manual	0.58	0.50
OY	Both	0.57	0.49
PB	KNP+Manual	0.76	0.68
PB	Manual	0.67	0.56
PB	Both	0.71	0.61
PM	KNP+Manual	0.87	0.84
PM	Manual	0.61	0.55
PM	Both	0.69	0.64
PN	KNP+Manual	0.86	0.75
PN	Manual	0.81	0.65
PN	Both	0.80	0.65

表 5: 一致率のマイクロ平均 (ジャンルごと)

うち、アスタリスクが付与されている箇所においては、適合率、または、再現率について同様に検定を行った場合、Manual と KNP+Manual は統計的に有意である。しかし、表 12 において、再現率について同様に検定を行った場合、KNP と KNP+Manual, 及び、KNP と Manual は統計的に有意であるが、Manual と KNP+Manual は有意ではない。また、正解率のマクロ平均について同様に検定を行った場

ジャンル	手法	見かけの一致率	カッパ係数
OC	KNP+Manual	0.58	0.27
OC	Manual	0.50	0.15
OC	Both	0.47	0.14
OW	KNP+Manual	0.80	0.73
OW	Manual	0.45	0.36
OW	Both	0.59	0.50
OY	KNP+Manual	0.63	0.47
OY	Manual	0.50	0.29
OY	Both	0.47	0.30
PB	KNP+Manual	0.63	0.54
PB	Manual	0.60	0.43
PB	Both	0.62	0.48
PM	KNP+Manual	0.87	0.83
PM	Manual	0.62	0.55
PM	Both	0.69	0.63
PN	KNP+Manual	0.88	0.74
PN	Manual	0.74	0.56
PN	Both	0.77	0.59

表 6: 一致率のマクロ平均 (ジャンルごと)

手法	適合率 (精度)	再現率	F 値
KNP	0.78	0.68	0.73
KNP+Manual	0.84	0.81	0.83
Manual	0.75	0.73	0.74
Average	0.80	0.77	0.78

表 7: 正解率のマイクロ平均 (全体)

手法	適合率 (精度)	再現率	F 値
KNP	0.47	0.40	0.43
KNP+Manual	0.55	0.55	0.55
Manual	0.53	0.51	0.52
Average	0.54	0.53	0.53

表 8: 正解率のマクロ平均 (全体)

合、標本数が少ないという理由から、統計的に有意ではない。

ジャンル	手法	適合率 (精度)	再現率	F 値
OC	KNP	0.72	0.48	0.57
OC	KNP+Manual	*0.78	0.75	0.77
OC	Manual	0.67	0.80	0.73
OC	Average	0.72	0.78	0.75
OW	KNP	0.79	0.79	0.79
OW	KNP+Manual	*0.82	*0.85	0.83
OW	Manual	0.65	0.67	0.66
OW	Average	0.73	0.76	0.74
OY	KNP	0.73	0.57	0.64
OY	KNP+Manual	*0.85	*0.75	0.80
OY	Manual	0.80	0.68	0.74
OY	Average	0.83	0.72	0.77
PB	KNP	0.75	0.60	0.66
PB	KNP+Manual	0.79	0.74	0.76
PB	Manual	0.78	0.73	0.75
PB	Average	0.78	0.73	0.76
PM	KNP	0.61	0.58	0.59
PM	KNP+Manual	0.89	0.86	0.87
PM	Manual	0.90	0.85	0.87
PM	Average	0.89	0.86	0.87
PN	KNP	0.88	0.78	0.83
PN	KNP+Manual	*0.88	*0.85	0.86
PN	Manual	0.77	0.72	0.75
PN	Average	0.83	0.79	0.81

表 9: 正解率のマイクロ平均 (ジャンルごと)

ジャンル	手法	適合率 (精度)	再現率	F 値
OC	KNP	0.31	0.26	0.28
OC	KNP+Manual	0.39	0.41	0.40
OC	Manual	0.42	0.44	0.43
OC	Average	0.40	0.42	0.41
OW	KNP	0.77	0.80	0.79
OW	KNP+Manual	0.83	0.86	0.84
OW	Manual	0.70	0.73	0.71
OW	Average	0.76	0.79	0.78
OY	KNP	0.58	0.44	0.50
OY	KNP+Manual	0.68	0.63	0.66
OY	Manual	0.56	0.49	0.52
OY	Average	0.62	0.56	0.59
PB	KNP	0.66	0.46	0.54
PB	KNP+Manual	0.71	0.65	0.68
PB	Manual	0.81	0.67	0.74
PB	Average	0.76	0.66	0.71
PM	KNP	0.60	0.66	0.63
PM	KNP+Manual	0.82	0.87	0.85
PM	Manual	0.86	0.84	0.85
PM	Average	0.84	0.85	0.85
PN	KNP	0.88	0.78	0.82
PN	KNP+Manual	0.88	0.85	0.86
PN	Manual	0.78	0.72	0.75
PN	Average	0.83	0.78	0.81

表 10: 正解率のマクロ平均 (ジャンルごと)

6. 考察

6.1 一致率とアノテーションにかかる時間

表 3, 表 4 より, KNP+Manual の一致率は, Manual の一致率よりもマイクロ平均, マクロ平均ともに高い数値となっていることがわかる。また, 表 5, 表 6 より, 全てのジャンルについて同様の傾向が見られることがわかる。これは, KNP+Manual の人手により修正される前のコーパスが共に同じ固有表現抽出器によってアノテーションされたものであることが影響していると考えられる。さらに, 表 11 より, KNP+Manual における 1 テキストあたりのアノテーションにかかる時間は, Manual より平均約 2 分程度短いということがわかる。これは有意水準 0.01 の F 検定で検定を行った場合, 統計的に有意である。これらのことから, KNP+Manual は Manual よりもアノテーションにかかる時間が短く, 一致率が高いということがいえる。

また, 表 5, 表 6 より, Both の一致率は多くの場合, Manual と同等以上の数値となっているが, OC における一致率のマクロ平均は, Both が 0.01 ポイント以上 Manual を下回っていることがわかる。このことから, OC には新聞コーパスから生成したルールだけでは抽出できないような固有表現が多く含まれているということがわかる。

手法	時間
KNP+Manual	3:19
Manual	5:23

表 11: アノテーションにかかる平均時間 (手法ごと)

手法	適合率 (精度)	再現率	F 値
KNP	0.78	0.68	0.73
KNP+Manual	0.91	0.89	0.90
Manual	0.87	0.88	0.88
Average	0.89	0.89	0.89

表 12: 2 人のアノテータのうち, どちらか一方でも正解のタグを付与しているならば正解とみなした場合の正解率のマイクロ平均 (全体)

手法	適合率 (精度)	再現率	F 値
KNP	0.47	0.40	0.43
KNP+Manual	0.63	0.62	0.63
Manual	0.62	0.62	0.62
Average	0.63	0.62	0.63

表 13: 2 人のアノテータのうち, どちらか一方でも正解のタグを付与しているならば正解とみなした場合の正解率のマクロ平均 (全体)

さらに, 表 3, 表 4 より, Manual のカッパ係数に関して, マイクロ平均では適度な値だったのに対し, マクロ平均で

ジャンル	手法	適合率 (精度)	再現率	F 値
OC	KNP	0.72	0.48	0.57
OC	KNP+Manual	0.87	0.86	0.87
OC	Manual	0.86	0.91	0.88
OC	Average	0.87	0.89	0.88
OW	KNP	0.79	0.79	0.79
OW	KNP+Manual	*0.91	0.91	0.91
OW	Manual	0.76	0.89	0.82
OW	Average	0.84	0.90	0.87
OY	KNP	0.73	0.57	0.64
OY	KNP+Manual	0.94	0.87	0.90
OY	Manual	0.93	0.86	0.89
OY	Average	0.94	0.87	0.90
PB	KNP	0.75	0.60	0.66
PB	KNP+Manual	0.87	0.82	0.84
PB	Manual	0.90	0.86	0.88
PB	Average	0.89	0.84	0.86
PM	KNP	0.61	58	0.59
PM	KNP+Manual	0.93	0.94	0.93
PM	Manual	*0.97	0.93	0.95
PM	Average	0.95	0.94	0.94
PN	KNP	0.88	0.78	0.83
PN	KNP+Manual	*0.93	0.90	0.92
PN	Manual	0.89	0.87	0.88
PN	Average	0.91	0.89	0.90

表 14: 2 人のアノテータのうち, どちらか一方でも正解のタグを付与しているならば正解とみなした場合の正解率のマイクロ平均 (ジャンルごと)

ジャンル	手法	適合率 (精度)	再現率	F 値
OC	KNP	0.31	0.26	0.28
OC	KNP+Manual	0.46	0.47	0.47
OC	Manual	0.49	0.51	0.50
OC	Average	0.48	0.49	0.49
OW	KNP	0.77	0.80	0.79
OW	KNP+Manual	0.91	0.91	0.91
OW	Manual	0.83	0.91	0.87
OW	Average	0.87	0.91	0.89
OY	KNP	0.58	0.44	0.50
OY	KNP+Manual	0.79	0.74	0.76
OY	Manual	0.68	0.65	0.67
OY	Average	0.74	0.70	0.72
PB	KNP	0.66	0.46	0.54
PB	KNP+Manual	0.84	0.78	0.81
PB	Manual	0.94	0.86	0.90
PB	Average	0.89	0.82	0.86
PM	KNP	0.60	0.66	0.63
PM	KNP+Manual	0.86	0.93	0.89
PM	Manual	0.98	0.93	0.95
PM	Average	0.92	0.80	0.92
PN	KNP	0.88	0.78	0.82
PN	KNP+Manual	0.93	0.90	0.92
PN	Manual	0.89	0.86	0.88
PN	Average	0.91	0.88	0.90

表 15: 2 人のアノテータのうち, どちらか一方でも正解のタグを付与しているならば正解とみなした場合の正解率のマクロ平均 (ジャンルごと)

は低い値となっていることがわかる. マイクロ平均は固有表現ごとの平均であり, マクロ平均はテキストごとの平均であるということから, Manual ではテキストごとに見たときに, 一致率の偏りが大きいということがいえる.

6.2 正解率

表 7, 表 8 より, KNP+Manual の正解率は, Manual の正解率よりもマイクロ平均, マクロ平均ともに高い数値となっていることがわかる. しかし, 表 9 より, OC における再現率と PM における適合率のマイクロ平均についてはこの傾向が見られず, また, KNP におけるこれらの指標は, 他のジャンルよりもかなり低い値となっていることがわかる. このことから, KNP+Manual の正解率は KNP の正解率に依存しているということがいえる.

また, 表 10 より, KNP+Manual の正解率は, OY, OW, 及び, PN については Manual の正解率より高い値となっている一方, OC, PB, 及び, PM については, PM の再現率を除き Manual の正解率より低い値となっていることがわかる. さらに, KNP の訓練事例である新聞コーパスに近く, KNP による正解率が高くなることが示されている [12] OW と PN において, KNP の正解率は Manual の正解率より高

手法	適合率 (精度)	再現率	F 値
KNP	0.78	0.68	0.73
KNP+Manual	0.74	0.38	0.50
Manual	0.67	0.29	0.40
Average	0.71	0.33	0.45

表 16: それぞれの手法で作成されたコーパスを訓練事例とした KNP を用いた場合の正解率のマイクロ平均 (全体)

手法	適合率 (精度)	再現率	F 値
KNP	0.47	0.40	0.43
KNP+Manual	0.40	0.24	0.30
Manual	0.31	0.16	0.21
Average	0.36	0.20	0.26

表 17: それぞれの手法で作成されたコーパスを訓練事例とした KNP を用いた場合の正解率のマクロ平均 (全体)

い値となっている. これらのことから, 非専門家がアノテーションを行う場合, KNP の訓練事例に近いジャンルのテキストについては KNP+Manual の方が良い結果を得られ, KNP の訓練事例から離れたジャンルのテキストについては Manual の方が良い結果を得られるということがい

ジャンル	手法	適合率 (精度)	再現率	F 値
OC	KNP	0.72	0.48	0.57
OC	KNP+Manual	0.88	0.29	0.43
OC	Manual	0.84	0.20	0.32
OC	Average	0.87	0.24	0.38
OW	KNP	0.79	0.79	0.79
OW	KNP+Manual	*0.74	*0.53	0.62
OW	Manual	0.55	0.36	0.43
OW	Average	0.65	0.45	0.53
OY	KNP	0.73	0.57	0.64
OY	KNP+Manual	0.84	*0.32	0.46
OY	Manual	0.80	0.18	0.30
OY	Average	0.82	0.25	0.38
PB	KNP	0.75	0.60	0.66
PB	KNP+Manual	0.70	0.31	0.43
PB	Manual	0.73	0.28	0.40
PB	Average	0.72	0.29	0.41
PM	KNP	0.61	0.58	0.59
PM	KNP+Manual	0.55	0.19	0.29
PM	Manual	0.52	0.14	0.22
PM	Average	0.54	0.17	0.25
PN	KNP	0.88	0.78	0.83
PN	KNP+Manual	0.76	*0.43	0.55
PN	Manual	0.78	0.36	0.49
PN	Average	0.77	0.40	0.52

表 18: それぞれの手法で作成されたコーパスを訓練事例とした KNP を用いた場合の正解率のマイクロ平均 (ジャンルごと)

ジャンル	手法	適合率 (精度)	再現率	F 値
OC	KNP	0.31	0.26	0.28
OC	KNP+Manual	0.24	0.16	0.19
OC	Manual	0.17	0.12	0.14
OC	Average	0.21	0.14	0.17
OW	KNP	0.77	0.80	0.79
OW	KNP+Manual	0.72	0.57	0.63
OW	Manual	0.63	0.43	0.51
OW	Average	0.67	0.50	0.57
OY	KNP	0.58	0.44	0.50
OY	KNP+Manual	0.52	0.24	0.33
OY	Manual	0.31	0.09	0.14
OY	Average	0.42	0.17	0.24
PB	KNP	0.66	0.46	0.54
PB	KNP+Manual	0.51	0.24	0.32
PB	Manual	0.65	0.22	0.32
PB	Average	0.58	0.23	0.33
PM	KNP	0.60	0.66	0.63
PM	KNP+Manual	0.55	0.29	0.38
PM	Manual	0.53	0.25	0.34
PM	Average	0.54	0.27	0.36
PN	KNP	0.88	0.78	0.82
PN	KNP+Manual	0.75	0.44	0.55
PN	Manual	0.78	0.37	0.50
PN	Average	0.77	0.40	0.53

表 19: それぞれの手法で作成されたコーパスを訓練事例とした KNP を用いた場合の正解率のマクロ平均 (ジャンルごと)

える。

6.3 2人のアノテータによる結果を統合したときの正解率

表 12, 表 13 より, どちらか一方でも正解のタグを付与しているならば正解とみなした場合, KNP+Manual の正解率は Manual の正解率よりも高い値となっているが, 2人のアノテータの平均正解率 (表 7, 表 8) に比べると, その差はかなり小さいということがわかる。これは, 少なくとも 2人のアノテータのうち, どちらか一方は正しいタグを付与していることが多いためであると考えられる。

また, 表 7, 表 8 は 2人のアノテータの正解率の平均であることから 1人のアノテータの正解率, 表 12, 表 13 は 2人のアノテータの正解率とみなすことができる。すると, 表 7, 表 8, 表 12, 表 13 より, 2人のアノテータによる正解率は常に 1人のアノテータによる正解率よりも高い値となっていることがわかる。さらに, 2人のアノテータによる Manual の正解率は, 常に 1人のアノテータによる KNP+Manual の正解率よりも高い値となっていることがわかる。このことから, 非専門家をアノテータとする場合, 既存の固有表現抽出器を使用すること以上に, アノテータの人数を増やすことが良い結果を得る上で重要であるといえる。

6.4 訓練事例としてのアノテーション結果

表 16, 表 17 より, それぞれの手法で作成されたコーパスを訓練事例とした KNP を用いた場合, KNP+Manual を訓練事例とした場合の正解率は Manual を訓練事例とした場合の正解率よりも高い値となっていることがわかる。しかし, 表 18, 表 19 より, PB と PN における適合率のマイクロ平均とマクロ平均, 及び, PB における F 値のマクロ平均についてはこの傾向が見られないことがわかる。このことから, KNP+Manual よりも Manual を訓練事例とした方が良いアノテーションとなる場合もあるということがわかる。

また, 表 16, 表 17 より, オリジナルの KNP の正解率は KNP+Manual を訓練事例とした場合の正解率や Manual を訓練事例とした場合の正解率よりも高い値となっていることがわかる。これは, KNP+Manual や Manual で作成されたコーパスが, オリジナルの KNP の訓練事例に比べ, とても少ないためであると考えられる。一方で, 表 16, 表 17 より, KNP+Manual を訓練事例とした場合と Manual を訓練事例とした場合, 及び, オリジナルの KNP において, 適合率は再現率に比べて大きな差がなく, また, 表 18 より, OC と OY の適合率のマイクロ平均において, オリジナルの KNP よりも KNP+Manual や Manual を

訓練事例とした場合の方が高い値となっていることがわかる。このことから、訓練事例が少ないとしても、適合率はオリジナルの KNP と同等以上になるといえる。

7. まとめ

本稿では、非専門家の手で高品質なコーパスを作成する方法について調べるため、固有表現抽出におけるアノテーションを題材として、既存の固有表現抽出器によるアノテーション結果に対し人手で修正を行う手法 (KNP+Manual) と既存の固有表現抽出器を使用せず人手でアノテーションを行う手法 (Manual) について比較を行った。この際、具体的にはアノテーションにかかる時間、タグの一致率、及び、Gold Standard との比較による正解率の各観点から比較を行っている。また、これに加え、機械学習における訓練事例としての品質を確かめるため、それぞれの手法で作成されたコーパスを訓練事例とした KNP を用いたアノテーションも行った。これらの実験を通して、全体的に見たとき、KNP+Manual は Manual よりもアノテーションにかかる時間が少なく、一致率や正解率についても高い値になることがわかった。一方で新聞から離れたジャンルのテキストについては、Manual の方が良い結果となっており、Manual を使った方が良いということが明らかになった。

謝辞 本研究は文部科学省科学研究費補助金 [若手B(No.15K16046)] と富士通研究所の助成により行われました。ここに謹んで御礼申し上げます。

また、KNP についての有益な情報を提供して下さった東京工業大学の笹野遼平先生に、この場を借りて御礼申し上げます。

参考文献

- [1] Snow, R., O’Conner, B., Jurafsky, D. and Ng, A. Y.: Cheap and Fast – But is it Good? Evaluation Non-Expert Annotation for Natural Language Tasks, *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 254–263 (2008).
- [2] Komiya, K., Suzuki, M., Iwakura, T., Sasaki, M. and Shinnou, H.: Comparison of Annotating Methods for Named Entity Corpora, *Proceedings of the 10th Linguistic Annotation Workshop, ACL 2016*, pp. 59–67 (2016).
- [3] Alex, B., Grover, C., Shen, R. and Kabadjov, M.: Agile Corpus Annotation in Practice: An Overview of Manual and Automatic Annotation of CVs, *Proceedings of Fourth Linguistic Annotation Workshop, ACL 2010*, pp. 29–37 (2010).
- [4] van der Plas, L., Samardžić, T. and Merlo, P.: Cross-lingual Validity of PropBank in the Manual Annotation of French, *Proceedings of Fourth Linguistic Annotation Workshop, ACL 2010*, pp. 113–117 (2010).
- [5] Marcus, M. P., Marcinkiewicz, M. A. and Santorini, B.: Building a large annotated corpus of English: the penn treebank, *Computational Linguistics - Special issue on using large corpora: II*, Vol. 19, pp. 313–330 (1993).
- [6] 橋本泰一, 乾 孝司, 村上浩司: 拡張固有表現タグ付きコーパスの構築, 情報処理学会研究報告, Vol. 2008-NL-188, pp. 113–120 (2008).
- [7] Maekawa, K.: Balanced Corpus of Contemporary Written Japanese, *Proceedings of the 6th Workshop on Asian Language Resources (ALR)*, pp. 101–102 (2008).
- [8] 徳永健伸, 西川 仁, 岩倉友哉, 湯上伸弘: 固有表現認識課題におけるアノータの視線分析, 一般社団法人情報処理学会 研究報告, Vol. 2015-NL-223, pp. 1–8 (2015).
- [9] Sasada, T., Mori, S., Kawahara, T. and Yamakata, Y.: Named Entity Recognizer Trainable from Partially Annotated Data, *Proceedings of the PACLING 2015*, pp. 10–17 (2015).
- [10] 岩倉友哉: 固有表現抽出におけるエラー分析, 言語処理学会第 21 回年次大会 (NLP2015) ワークショップ: 自然言語処理におけるエラー分析 (2015). http://www.anlp.jp/proceedings/annual_meeting/2015/html/paper/WS_PNN04_ne.pdf.
- [11] 平田亜衣, 小町 守: 様々なジャンルのテキストに対する固有表現認識の分析, 言語処理学会第 21 回年次大会 (NLP2015) ワークショップ: 自然言語処理におけるエラー分析 (2015). http://www.anlp.jp/proceedings/annual_meeting/2015/html/paper/WS_PNN06_02.pdf.
- [12] Ichihara, M., Komiya, K., Iwakura, T. and Yamazaki, M.: Error Analysis of Named Entity Recognition in BCCWJ, 言語処理学会第 21 回年次大会 (NLP2015) ワークショップ: 自然言語処理におけるエラー分析 (2015). http://www.anlp.jp/proceedings/annual_meeting/2015/html/paper/WS_PNN05_01.pdf.
- [13] Sasano, R. and Kurohashi, S.: Japanese Named Entity Recognition Using Structural Natural Language Processing, *Proceedings of IJCNLP 2008*, pp. 607–612 (2008).
- [14] Information Retrieval and Extraction Exercise: ルール、定義, <http://nlp.cs.nyu.edu/irex/NE/df990214.txt> (1999).
- [15] Iwakura, T., Tachibana, R. and Komiya, K.: Constructing a Japanese Basic Named Entity Corpus of Various Genres, *Proceedings of NEWS 2016* (2016).