

ネットオークションの出品情報文書からの属性抽出の精度向上

塚原 裕常[†] 西村 純^{†*} 宮崎 林太郎[†] 前田 直人^{†*} 森 辰則[†]

小林 寛之[‡] 石川 雄介[‡] 田中 裕也[‡] 翁 松齢[‡]

[†]横浜国立大学大学院環境情報学府 〒240-8501 横浜市保土ヶ谷区常盤台 79-7

[‡]ヤフー株式会社 〒106-6182 東京都港区六本木 6-10-1 六本木ヒルズ森タワー

E-mail: [†]{ht, jun-n, rintaro, n-maeda11, mori}@forest.eis.ynu.ac.jp

[‡]{hkobayas, yuishika, hkobayas, yuutanak, shou}@yahoo-corp.jp

あらまし ネットオークションにおける属性検索を目的として、出品情報文書から出品物の属性・属性値対を抽出する手法が提案されている。本論文では、その抽出精度を向上させる一手法について検討をする。具体的には、前処理として、出品情報文書中の各文について、出品物の説明記述を含むか否かの二値分類を行うことにより、送料や関連商品の紹介など出品物と直接関係の無い記述を削除した文書を抽出処理の対象とする。評価実験によれば、本手法を適用することにより、属性抽出における再現率、適合率のいずれもが数ポイント上昇することが確認された。

キーワード ネットオークション, 属性, 情報抽出

Improvement of the performance of attribute-value extraction from description of exhibits in net auction system

Hironobu TSUKAHARA[†] Jun NISHIMURA[†] Rintaro MIYAZAKI[†] Naoto MAEDA[†] Tatsunori MORI[†]

Hiroyuki KOBAYASHI[‡] Yusuke ISHIKAWA[‡] Yuya TANAKA[‡] Shorei O[‡]

[†]Yokohama National University 79-7 Tokiwadai, Hodagaya-ku, Yokohama-shi, 240-8501 Japan

[‡]Yahoo Japan Corporation Roppongi Hills Mori Tower, 6-10-1 Roppongi, Minato-ku, Tokyo, 106-6182 Japan

E-mail: [†]{ht, jun-n, rintaro, n-maeda11, mori}@forest.eis.ynu.ac.jp

[‡]{hkobayas, yuishika, hkobayas, yuutanak, shou}@yahoo-corp.jp

Abstract In order to achieve faceted search in net auction system, several researchers have dealt with the automated extraction of attributes and their values from descriptions of exhibits. In this paper, we studied a method to improve the performance of the extraction. First, the method selects each sentence in a description that is judged to have attributes and/or values. Then, the extraction of attributes and values are performed on the cleaned text that does not contain parts of description irrelevant to exhibits, like descriptions for the postage, other exhibits, and so on. The experimental result shows that the proposed method improves both the precision and the recall in the attribute-value extraction.

Keyword net auction, attribute, information extraction

1. はじめに

近年の情報抽出技術の利用例のひとつとして、ネットオークションにおける出品物の検索の高度化が挙げられる。現在のネットオークションでは、出品物の検索に全文検索が用いられているため、サイズや色などの属性情報を検索語として入力した場合、ユーザが必要としている属性記述のある出品情報だけでなく、図1のように属性情報以外に現れる文字列に一致する出品情報も検索されてしまうという問題点がある。しか

し、ネットオークションの出品情報文書に多数存在する商品の属性情報に着目し、それらを高い精度で抽出することが可能となれば、属性検索によって利用者が望む柔軟な検索を実現できることだろう。

また、ネットオークションの出品情報文書中には、送料や代金の支払い方法などの記述部分が多く含まれており(図2)、出品物の説明をしている記述部分は総テキスト量の半分以下でしかない。さらに、出品物の属性情報を含む文となると、総テキスト量の1/3にも満たない。そのため、これらの不要な部分をあらかじめ

* 現在、ヤフー株式会社所属。

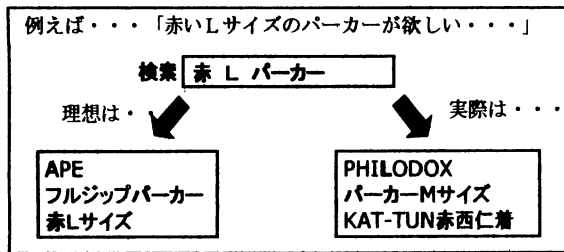


図1：現在のオークション検索の問題点の例

●●●商品説明

【半袖Tシャツ黒 サイズXL】新品未着用
男性用XLサイズ対応（胸囲104-112、身長175-185）。

1枚目の画像は正面胸部分のロゴ。
2枚目は袖（マーク）と背面（水色部分）です。
お勧めですよ♪

●●●発送

- ・定形外郵便（240円）
- ・ポストケット（400円、配達追跡が可能です）

●●●お支払い

- ・ヤフーかんたん決済
- ・ヤフーネットバンキング

図2：出品情報文書の例

め切り分けることにより、属性情報抽出において不要部分に含まれる属性情報に類似した記述の影響が排除できると期待される。また、属性・属性値の抽出が比較的局所的な文脈情報に基づき行われるのに対して、ある文にそれらが含まれているか否かの判断には文に亘るより大域的な文脈情報が利用できる。よって、両者を組み合わせた手法は、すべてのテキストに対して属性抽出を行う場合よりも精度が向上することが期待される。

そこで、本論文では機械学習手法を用いて、出品情報文書中において、出品物の属性情報が含まれている部分とその他の不要な部分を切り分ける手法を検討する。具体的には、出品情報文書を文単位に分け、二値分類を行うことにより、各文が属性情報を含むかどうかを判定する。

そして、西村ら[1]による既存の属性抽出手法と結合し、属性抽出の前処理として本手法が効果的であるかどうかを検討する。

2. 先行研究

テキストから必要な情報を抽出する研究としては、中野ら[2]が行った日本語固有表現の抽出がある。中野

ら[2]は文節区切りを行い、文節内の情報を素性としてチャンカーに与える手法を提案している。

また、固有表現抽出の技術を応用して様々なテキストに対して属性抽出が行われている。属性関係を抽出する先行研究としては、高橋ら[3]によるものがある。高橋ら[3]は対象物・属性名・属性値という三つ組の属性関係をテキスト中から抽出するタスクを考え、抽象化したパターンによりドメインを限定せずに三つ組の候補を抽出し、統計量を使ってそれらをフィルタリングする手法を提案している。他にも、新里ら[4]は固有表現抽出手法を用いて、レストランに関する属性情報を自動的に認識するという研究を行っており、飯田ら[5]は意見抽出を目的として機械学習に基づく手法を用いて属性と評価値の対を同定する方法を提案している。

上記の先行研究に加えて、本研究に最も関連が深いものとして、西村ら[1]のネットオークションにおける出品情報テキストからの属性抽出がある。西村ら[1]は、ネットオークションの出品情報を各種属性により柔軟に検索することを目的として、中野ら[2]の固有表現抽出に基づき、出品情報に多数存在する属性・属性値の情報を機械学習により自動抽出し、抽出された属性・属性値の対の同定を行う手法を提案している。

本研究は、西村ら[1]の属性抽出をより正確かつ迅速に行うことを可能にすることを目的としたものである。次節において、西村らの抽出手法を説明し、続く次章において、提案手法について述べる。

2.1. ネットオークションにおける属性抽出[1]

2.1.1. 抽出対象となる属性情報と出品情報文書

まず、抽出対象となる属性・属性値の定義がなされている。対象とする出品情報文書として、属性情報による検索の需要が多いと思われる、ファッションカテゴリに属するものが用いられていて、表1に示す種類の属性を対象としている。これらの情報はいずれも利用者が検索の対象として必要と考えられるものである。

そして、属性情報の抽出のタスクは、属性、属性値に対応する表層表現にそれぞれ<attr id="...">, <val id="...">タグを付与し部品の抽出すること、ならびに、組になる属性・属性値を見つけ出し、共通のidを付与して部品の組を同定することからなるとされる。組の情報を付与するのは、抽出された属性・属性値が属性検索の際に利用されることを想定しているためである。学習用のコーパスについても、同じ形式の注釈がなされている。なお、本論文では前半のタスクである属性、属性値の抽出を対象とし、後半のタスクである組の同定は対象としない。

表 1 : 抽出対象となる属性情報

属性の種類	属性値の例
色	黄色, イエローなど
素材	ポリエステル 50%, 綿 100%など
サイズ	着丈: 65cm, M サイズなど
形状	半袖, ノースリーブなど
状態	新品, 未使用, 古着など
定価	定価 2000 円など
製造場所	日本製, made in USA など
シーズン/モデル	秋冬モデル, 1970 年代など
デザイン	花柄, ストライプなど
その他	重さなどあまり出てこないもの

2.1.2. 系列ラベリングに基づくチャンキングによる属性情報の抽出

文字を単位とする分類問題として定義された系列ラベリングに基づくチャンキングにより情報抽出を行う。そのために、出品情報文書を文字の単位に分け、各々に6つの素性を与えている。具体的に与えた素性は、表層文字、文字種、品詞、文節内素性、複合名詞主辞素性(以下、主辞素性と記載)、シソーラス上の概念分類番号(以下、分類番号と記載)である。

文節内素性とは文節内に固有表現が存在すれば、最も先頭に近い固有名詞の品詞細分類を、固有名詞がなければ文節の先頭の単語を素性とするものである。

複合名詞主辞素性とは、連続する名詞が存在する場合、その最後の名詞を素性とするものである。

分類番号とは、角川類語新辞典において各単語に付与されている番号のことである。同じ範疇に属する、意味的に近い単語には同じ分類番号が付与されるので、表層表現が異なっても同じ素性を持つ事例として考慮される。この素性は、属性抽出の出品カテゴリに対する依存を弱め、カテゴリ横断的な属性抽出を可能にするために用いられた素性である。

また、チャンキングには IOE2 法を用い、チャンキングを行う文字の前後 2 文字ずつ計 5 文字を文脈長としている。IOE2 法では、チャンクの最終トークンには E という記号が、それ以前のトークンには記号 I が、要素以外のトークンには O が付与される。要素の抽出規則の学習は、図 3 の枠内の素性から対応するチャンクタグを得るような分類器を、学習事例と機械学習手法を用いて構成することに相当する。

これらの素性展開とチャンキング手法により、未知の文における抽出の際には、各文字に枠内の素性集合を導出し、その素性集合を分類器に与えることによりチャンクタグを文末から文頭に向けて順次推定する。

位置	文字素	文字種	品詞	文節内素性	主辞素性	分類番号	タグ
	材	KANJI	B-名詞-普通名詞	素材	素材	805	
#2	材	KANJI	E-名詞-普通名詞	素材	素材	805	
#1	は	HIRAG	S-助詞-副助詞	*	*	*	
i	レ	KATAK	B-名詞-普通名詞	レーヨン	レーヨン	907	I-val
f1	ー	OTHER	I-名詞-普通名詞	レーヨン	レーヨン	907	I-val
f2	ヨ	KATAK	I-名詞-普通名詞	レーヨン	レーヨン	907	I-val
	ン	KATAK	E-名詞-普通名詞	レーヨン	レーヨン	907	E-val
	,	OTHER	S-特殊-句点	*	*	*	O

図 3 : 素性集合に対するチャンクタグの推定

3. 提案手法

Kaynak ら[6]は、複数の分類器を直列に多段接続することにより、既存の独立した分類器を用いるのとは比べて複雑さとコストを過度に増加させることなく、分類精度を上昇させる手法について提案している。初期の段階においては、単純ではあるが汎用性のある分類器を学習するが、段階が進むにつれて、局所的な事例に特化したより複雑な分類器を学習する。Kaynak らはこれを Cascading と名づけ、その有効性について考察している。

本論文で提案する手法は、この Cascading の考え方に類似するものである。すなわち、一段階目より大域的な情報を用いて、大まかな分類を行い、二段目より局所的な情報を用いて、詳細な抽出を行うものである。具体的には、一段階目として、出品情報文書中の各文についてその文が出品物の属性情報を含むか否かの二値分類を行い、そこで含むと判断された文についてのみ、二段階目の処理である、西村ら[1]による属性抽出を行う。

一方で、Cascading が同質の分類タスクを各段で行っているのに対して、本稿での処理では一段目でのタスクが、実際に行うべき二段目のタスクの範囲を限定するために用いられており、質の異なる分類タスクの組み合わせである点が異なる。

第二段階目については、第 2 章で説明したので、本章の残りの部分においては、上記の第一段階目について述べる。なお、以下の記述において、「出品情報文書中の当該出品物の属性情報が含まれる文」を「説明記述文」と呼ぶことにする。

3.1. 出品情報文書中の各文に対する素性の付与

本論文では、Support Vector Machine(以下、SVM と記す)を用いた二値分類を行うことにより、出品情報文書から出品物の説明記述となる部分を抽出する。この時、出品情報文書を文単位に分け、文単位で二値分類を行い、抽出をする。そのため、出品情報文書中の文毎にその特徴量を抽出し、機械学習の際の素性とする。具体的に利用した素性は、以下の 3 つである。

3.1.1. 分類対象となる文における形態素の出現頻度

まず、3つの素性のうち主となるのは、各々の文を形態素解析して得られた形態素原形の出現頻度である。出現した形態素に通し番号をつけ、文中でのそれぞれの出現頻度をその素性番号に対する値として用いた。つまり、これは形態素を次元として各文をベクトルとして表したものであり、いわゆる bag-of-words である。実際の素性の作り方の例を図4に示す。4章の評価実験においては、このような素性を SVM の学習データ・テストデータとして与えた。

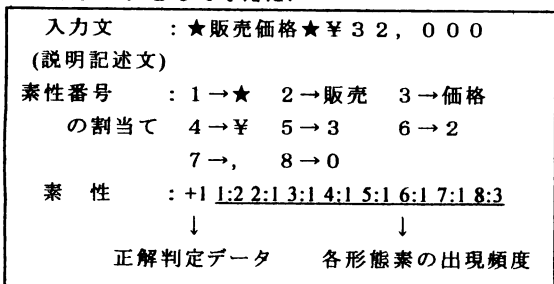


図4：形態素頻度に基づく素性

3.1.2. 分類対象となる文の前後の文の判定結果と形態素の頻度情報の利用

出品情報文書には出品物の説明記述文だけでなく、前述の通り、送料の説明や他の出品物の宣伝や「ノークレーム・ノーリターンでお願いします。」といった決まり文句など、出品物とは直接的に関係のない文も含まれている。しかし、ネットオークションの出品者が出品情報文書を記述する場合、出品物の説明記述を出品物と関係のない文と混ぜ合わせて書くよりは、それぞれまとめて書く方が一般的だと考えられる。

そこで、形態素の出現頻度のみを素性とした場合の分類器の出力を初期分類結果として利用し、分類対象としている文の前後数文についての初期分類の判定結果を新たに2種類目の素性として加え、再度学習を行う手法についても検討する。

また、これに加え、同じ理由から、3種類目の素性として前後数文の形態素の出現頻度を採用した場合についても検討する。

4. 評価実験および考察

機械学習手法である SVM を用いることにより、出品情報文書中の説明記述文を抽出する実験を行った。形態素解析器には ChaSen2.3.3 を、SVM の実装系としては、TinySVM 0.09 を用いた。また、この抽出結果を既存の属性抽出システムの入力とする、連結したシステムについて、属性抽出の精度を検証し、提案手法の

有用性について考察する。

評価に際しては、説明記述文の抽出のみの場合、ならびに、連結したシステムの精度評価の場合のいずれについても、出品情報文書中の文を要素とした集合を5分割して行う交差検定を用いている。また、評価尺度としては、説明記述文の抽出のみの場合は説明記述文が正しく抽出されているかどうかについて、また、連結したシステムの場合は個々の属性・属性値が正しく抽出されているかについて、適合率、再現率、F 値の平均値を求めた。

4.1. 実験データ

実験には、Yahoo!オークションに出品された商品の出品情報のうち、「ファッション-アパレル(男性用)-トップス-シャツ-半袖」(150 ページ・4930 文、属性：総数 1422 個/異なり数 149 個、属性値：総数 1794 個/異なり数 512 個)を用いた。この出品情報文書には、あらかじめ表1で示された属性・属性値について XML タグによる注釈付けが行われている。

また、出品者固有の記述表現により抽出精度が左右されないために、同一出品者による出品情報は用いないようにした。

4.2. 出品情報文書からの説明記述文抽出の場合

本節では、出品情報文書からの説明記述文の抽出実験について述べる。3.1.節で述べたとおり、機械学習に用いる素性を

- ①：対象となる文における形態素の出現頻度
- ②：①+前後数文における判定
- ③：②+前後数文における形態素の出現頻度と変化させた。各結果を表2に、また、素性に②を用いた場合の結果を図5に示す。なお、図5中の横線は、①のそれぞれの指標を示している。

表2：説明記述文抽出結果

	適合率[%]	再現率[%]	F 値
①	<u>90.82</u>	86.24	<u>88.47</u>
②(1文)	89.41	85.94	87.64
②(3文)	89.26	86.38	87.80
②(5文)	88.93	86.46	87.68
②(10文)	89.12	86.97	88.03
②(20文)	89.26	<u>87.04</u>	88.14
②(50文)	89.20	85.87	87.50
②(記事内全文)	88.04	85.65	86.83
③(1文)	86.99	85.65	86.32
③(5文)	87.72	83.67	85.65
③(20文)	87.53	83.75	85.60

*括弧内は、素性として用いた前後文の数

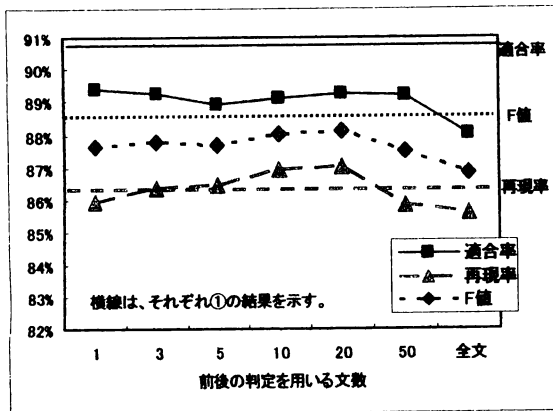


図5：前後数文の判定結果を用いた説明記述文抽出

表2に示す通り、対象となる文における形態素の出現頻度だけを素性として用いた場合でも、適合率、再現率ともにおよそ9割程度に至る示すことがわかった。また、F値については①の場合が最も高い値を示している。

また、②では、①に比べて適合率は1~2ポイント低下するものの、用いる文数によってはおよそ1ポイントながら再現率が向上している。これにより、前後の文が説明記述文であるか否かが、対象文自身が説明記述文であるか否かの判定に影響を与えている可能性があるといえる。しかし、前後各1~20文を用いた場合は、F値が①と比べて1ポイント以内の差となっており、用いるカテゴリやコーパスによって、最も優れた結果を示す文数は変化しうると予想される。

③では、用いる文数に限らず、適合率・再現率ともに①と②を下回る結果となった。従って、対象となる文の前後数文における形態素の出現頻度を素性に用いることの有用性は低いと考えられる。

4.3. 説明記述文の抽出と属性抽出を連結した場合

本節では、説明記述文の抽出手法と、西村ら[1]による既存の属性抽出システムとを連結した場合の属性・属性値の抽出実験について述べる。

属性抽出システムの処理対象としては、

学習時：コーパス中の注釈情報から判断した、説明記述文(属性もしくは属性値を有する文)の集合。注釈情報つき。

評価時：説明記述文抽出手法によって説明記述文であると判断された得られた文。注釈情報なし。

を用いた。結果を表3に示す。なお、表中のBaselineは、属性抽出システム単独での抽出結果である。また、「上限値」は、説明記述文の抽出が完全に正しく行わ

れた状況における、属性・属性値の抽出結果であり、説明記述文抽出を前処理に付加することによる精度向上の上限値である。

表3：連結したシステムによる属性抽出結果

	属性			属性値		
	適合率 [%]	再現率 [%]	F値	適合率 [%]	再現率 [%]	F値
Baseline	88.4	84.2	86.2	83.4	74.4	78.6
上限値	93.2	90.3	91.7	91.4	81.9	86.4
①	89.4	87.3	88.3	87.0	80.7	83.7
②(1文)	89.4	86.5	87.9	86.6	80.7	83.5
②(5文)	89.2	86.1	87.6	86.1	80.9	83.4
②(20文)	89.2	86.3	87.7	86.2	81.4	83.7
③(1文)	89.2	84.5	86.8	86.1	80.1	83.0
③(5文)	89.5	83.4	86.3	86.8	79.3	82.9
③(20文)	89.4	83.1	86.1	86.8	79.0	82.8

表3によれば、説明記述文の抽出過程を連結することにより属性抽出の精度が、Baselineに比べて、属性の適合率が最大で1.1ポイント、属性の再現率が最大で3.1ポイント、属性値の適合率が最大で3.6ポイント、属性値の再現率が最大で7.0ポイント向上しており、素性選択を変更してもいずれの指標もほとんどがBaselineに比べて向上している。Baselineでは属性値の再現率が他の指標に比べて9~14ポイントも低く、属性値の再現率の低さはBaselineの弱点であったが、説明記述文の抽出過程を連結することによってその弱点が幾分克服できたのではないかと考えられる。

また、Baseline、①、②、③の4者を比べると、①はどの指標においても上位の精度を示している。これは、②や③よりもベクトルの次元数が低いので、他に比べて十分な学習ができていているという要因も考えられる。

加えて、属性値の再現率において②(20文)が①と③よりも高い値を示していることも見受けられる。表2によれば、②(20文)は説明記述文抽出に関する再現率が高いので、これが、要因になっていることが考えられる。

4.3.1. 説明記述文抽出を連結した場合の属性抽出結果の変化

表3のように、説明記述文抽出システムと属性抽出システムを連結することによって、属性抽出の適合率と再現率はそれぞれ数ポイントずつ向上した。本小節では、属性抽出システム単独での抽出結果と連結したシステム(4.2節における素性①を使用)を比較し、抽出結果の変化について考察する。

まず、適合率の向上について考察する。適合率が向

上する典型的な状況は、本来属性情報でないにもかかわらず抽出してしまっていた情報が抽出されなくなることである。今回の連結では、このような例がいくつも見られた。属性抽出システム単独では、「ゆうパック60サイズ」の「サイズ」や、「汚れが無いことをチェックしておきました」の「チェック」などが属性情報として抽出されてしまっていたが、これらは説明記述文抽出の段階で無関係の文を削除することにより抽出されなくなった。これらは、出品物の属性情報の中に同じ表層をしている例があるため、属性情報システム単独では抽出誤りを起こしてしまうことが多くあったと思われる。

次に、再現率の向上について考察する。再現率が向上する典型的な状況は、本来属性情報であるにもかかわらず抽出できていなかったものが抽出されることである。説明記述文抽出を連結した場合には、「素材」や「サイズ」、「デザイン」等に対する属性値を中心に抽出精度が向上している傾向にある。具体例としては、属性抽出システム単独では、「豚皮（属性:素材の値）」や「花（属性:デザインの値）」などの属性値の抽出漏れがあったが、説明記述文のみに対して属性抽出の学習を行うことにより、これらの抽出が可能となった。

しかし、連結によって変化したのは必ずしも良い方向にだけではなく、抽出できなくなった属性情報や、誤って抽出してしまうようになった不要な属性情報も存在する。加えて、連結を行ってもなお抽出することができない属性情報も存在することから、説明記述文の抽出過程を前処理として連結する手法は全体的な精度向上をもたらしているものの、局所的には常に精度向上をもたらすものではない。

5. まとめと今後の課題

本論文では、オークションの出品情報文書を対象とした既存の属性抽出手法について、その抽出精度を向上させる手法について検討をした。具体的には、前処理として、出品情報文書の各文に対して出品物の説明記述を含むか否かの二値分類を行うことにより、送料や関連商品の紹介など出品物と直接関係の無い記述を削除するものである。また、この二値分類のための素性として、形態素の出現頻度、前後文の判定、前後文の形態素の出現頻度を用い、前処理に有効な素性について検討した。

4.2 節では、説明記述文抽出に関する実験を行い、素性を変化させた場合の精度を検討した。F 値については、素性として形態素の出現頻度のみを用いた場合(①)が最も良い結果が得られたが、①の素性に、前後

文の判定を加えた場合(②)においては、再現率の向上が見られることを確認した。

4.3 節では、説明記述文抽出手法と既存の属性抽出手法とを連結し、説明記述文抽出が属性抽出の精度向上をもたらすことを確認した。この時、全体的な精度向上としては、素性に形態素の出現頻度のみを用いた場合が最良であったと言えるが、属性値の再現率の向上という観点では、前後文の判定を加えたものが最も良い結果となった。

今後の課題としては、以下の事柄が考えられる。前後文の判定や前後文の形態素の出現頻度を素性に用いた場合に、精度向上が見られなかったので別の素性を検討する必要がある。また、これらの素性を追加した場合、素性の次元数が大きく増加してしまうため、今回使用したコーパスでは学習データ不足であった可能性が考えられる。従って、次元圧縮を行うとともに、学習データを増やすことにより、さらなる精度向上が達成できるかもしれない。また、説明記述文抽出は、人手によるコーパス作成の際に、不要なテキスト部分を見る必要をなくし、作業効率を向上させるためにも用いることができると考えられる。

最後に、今回の検討では属性抽出までを考察したが、その先の段階にある属性・属性値の組の同定手法とも組み合わせることにより、最終的な属性・属性値対の抽出精度向上について検討したい。

参 考 文 献

- [1] 西村純, 宮崎林太郎, 前田直人, 森辰則, 翁松齡, 石川雄介, 小林寛之, 田中裕也, 木戸冬子 "ネットオークションにおける属性検索のための出品情報文書からの属性抽出" 言語処理学会第 14 回年次大会発表論文集, pp392-395(2008)
- [2] 中野桂吾, 平井有三 "日本語固有表現抽出における文節情報の利用" 情報処理学会論文誌, Vol.45, No.3, pp.934-941(2004)
- [3] 高橋哲朗, 乾健太郎, 松本裕治 "テキストから属性関係を抽出する" 情報処理学会研究報告, 2004-NL-164, pp19-24(2004)
- [4] 新里圭司, 関根聡, 吉永直樹, 鳥澤健太郎 "固有表現抽出手法を用いたレストラン属性情報の自動認識" 言語処理学会 第 12 回年次大会 発表論文集(2006)
- [5] 飯田龍, 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一 "意見抽出を目的とした機械学習による属性-評価値対同定" 情報処理学会研究報告, 2005-NL-165(2005)
- [6] Cenk Kaynak and Ethem Alpaydin "Multistage Cascading of Multiple Classifiers: One Man's Noise is Another Man's Data" Proc. of the Seventeenth International Conference on Machine Learning (2000)