

恒久性と一意性に基づく固有表現の二項関係の分類

高久 陽平[†] 鍛冶 伸裕[‡] 吉永 直樹[‡] 豊田 正史[‡]

[†] 東京大学大学院 情報理工学系研究科 〒113-0033 東京都文京区本郷 7-3-1

[‡] 東京大学 生産技術研究所 〒153-8504 東京都目黒区駒場 4-6-1

E-mail: †{takaku, kaji, ynaga, toyoda}@tkl.iis.u-tokyo.ac.jp

あらまし ウェブテキストなどを知識源として、固有表現間の二項関係を獲得する研究が近年注目されている。本研究では、ある関係において複数の二項関係が獲得された場合、それらは並立しているのか、それとも時間によって変化しているため複数獲得されたのか判断することが困難であることに着目した。そこで、そのような二項関係を区別するため、関係を「時間において恒久的に成立するか否か」、また「一対一で成立するか否か」という二つの観点から分類することを提案する。これを実現するために、時系列ウェブテキストの頻度情報に基づき機械学習を用いて関係を分類する方法を考案した。実験では、5年分のブロッガーアーカイブから獲得された関係を分類し、提案手法の有効性を確認した。

キーワード 関係抽出, 情報抽出, 自然言語処理

Keyword relation extraction, information extraction, natural language processing

1. はじめに

ウェブテキストは、広範囲に渡る莫大な情報が記述されており、世界知識獲得の知識源として注目されている。そのような知識獲得の研究において、特にテキストから二つの固有表現とその間の関係を二項関係として抽出する関係抽出は、質問応答システム[1]や矛盾検出[2]などといったアプリケーションを支える自然言語処理の重要なタスクの一つとして研究されている。二項関係の獲得では、例えば「本田圭佑がCSKAモスクワに所属している」という知識を、二つの固有表現“本田圭佑”と“CSKAモスクワ”をそれぞれ前項X、後項Yに置き換え、(1a)のような三つ組として表現し収集する。

しかし、実際に獲得される二項関係には(1a)に加え、(1b)のように過去に成立していたが、現在は成立していないものが獲得されてしまうことがある。この場合、現在どちらが本田圭佑の所属先なのか、もしくはどちらとも所属先なのか判断することができない。この問題に対する自然な解決策として、ある固有表現(本田圭佑)に対して複数の関係知識が獲得された場合には、最も頻度が高い関係を採用するという方法が考えられる。しかしながら、(1c)と(1d)のように一つの固有表現(東京)に対して複数の関係が同時に成立することもあるため、そのような単純な方法による解決は難しいと考えられる。

- (1) a. (本田圭佑, Xが所属するY, CSKAモスクワ)
- b. (本田圭佑, Xが所属するY, VVVフェンロ)
- c. (東京, Xを流れるY, 荒川)
- d. (東京, Xを流れるY, 多摩川)

そこで本研究では、関係には「時間において恒久的に成立するか否か」、また「一対一で成立するか否か」という二つの性質(以降、前者を恒久性、後者を一意性と呼ぶ)があることを指摘し、その分類手法を提案することでこの問題の解決を図る。このような関係に関する性質を知ることにより、例えば、関係“Xが所属するY”が恒久性がないということと、一意性があるということが分かるので、最近抽出された二項関係であり、かつ頻度が多い二項関係一つを選ぶという方法をとることができる。

提案手法では、時系列ウェブテキストの頻度情報に基づき、恒久性と一意性それぞれに対して関係の分類を行った。頻度情報では、時間移動窓を導入し各移動窓における固有表現の分布から分類に有効な素性を用いた。実験では、5年分のブログ記事から獲得された関係を分類し、提案手法の有効性を確認した。

本稿の構成は以下の通りである。まず、2章で関連研究について述べる。次に3章で本研究が設定する問題についてと二項関係データの作成について述べる。また、4章では提案手法について、頻度情報に基づく素性を用いた手法を述べる。そして、5章では提案手法の評価実験について述べ、最後に6章でまとめと今後の課題について述べる。

2. 関連研究

関係抽出とはテキストから、二つの固有表現とその間の関係を獲得する研究であり、Bunescuら[3]のように文章中の品詞や係り受け構造を基にSVM(Support

Vector Machine)などの機械学習を用いて獲得する手法といった様々な手法が研究されている。また、特に近年では、ウェブテキストを対象とした関係獲得が盛んに研究されており、この場合、獲得対象が大規模になるため、従来の教師あり学習の手法では人手による訓練データの作成に莫大なコストがかかるという問題がある。そのため、訓練データを低コストで生成する研究[4,5,6]やクラスタリングなどの機械学習を行わない手法[7,8]などが研究されている。しかしながら、これらの手法によって抽出された関係の情報だけでは、1節で述べたような二項関係の判別はできない。

二項関係を精緻化する研究は非常に少ない。Linら[9]は関係 R において、「任意の固有表現 X, Y_1, Y_2 に対し、 (X, R, Y_1) かつ (X, R, Y_2) ならば、 $Y_1=Y_2$ である」ものを関数的関係(Functional Relation)と定義した。彼らは、後項 Y に当てはまる固有表現の分布を用いて、類似度を用いた分類を提案し、関係をニクラスに分類した。しかし、彼らの研究では恒久性を考慮していないため、二項関係(1a), (1b)のように、時間において変化する関係の場合に、複数の二項関係が獲得されてしまうという問題に対処することができない。

また、時間情報に着目した研究では、イベント情報に対して、それが行われた期間情報を獲得する研究が行われている。Lingら[10]は文章中の時間表現(例えば since 1997)に着目して期間情報を抽出する手法を提案した。しかし、彼らの研究は獲得対象が主にイベント情報である点や、本研究のように恒久性という性質に着目していないといった点で本研究と異なる。

このように、関係を恒久性と一意性の二つの観点から捉えた研究はなされていない。

3. 恒久性と一意性に基づく分類

本節では、まず、恒久性と一意性の定義について述べる。次に、本研究が対象とする問題の設定について述べる。最後に提案手法で用いる二項関係データの作成法について述べる。

3.1 恒久性と一意性の定義

恒久性の定義 二項関係において、前項 X に固有表現が代入されたとき、後項 Y に当てはまる固有表現の集合が時間的に変化しないとき、その二項関係は恒久性を有するという。例えば、関係“ X が所属する Y ”は、(1a), (1b)のように前項 X の所属先が時間変化するので、恒久性のある関係ではない。

一意性の定義 同様に、前項 X に固有表現が代入されたとき、後項 Y に当てはまる固有表現が任意のある時点において常に唯一に決まるとき、その二項関係は一意性を有するという。例えば、関係“ X が所属する

	一意性あり	一意性なし
恒久性あり	Xの首都であるY Xの父親のY Xの原作者のY Xの出身地Y Xの本名はY Xの出世作Y Xの前進はY Xの主人公はY Xの長男のY XのファーストアルバムY	Xを流れるY Xの両親のY Xの隣町Y Xの支流であるY XのマスコットY Xの姉のY Xと国境を接するY Xの同期のY Xの終点のY Xの代表作Y
恒久性なし	Xが所属するY Xの社長Y XはYと交際する XのニューシングルY XはYが優勝する XがYを離党する Xの監督はY Xの授賞式がYで行う Xの所属事務所Y Xの最新作Y	Xの友達のY XがYに出演する Xで放送中のドラマY XのメンバーのY Xを離党するY XがプロデュースするY Xに連載中のY Xの子のY XではYが流行る Xの幹部Y

表 1: 関係の恒久性と一意性による分類例

Y ”は、(1a)のようにある時点において所属先は一意に決まるので、一意性のある関係である。

3.2 問題設定

本研究では、入力として関係(例えば“ X が所属する Y ”, “ X を流れる Y ”)が与えられたとき、その関係の恒久性と一意性の有無をそれぞれ個別に判定するという二つの問題を扱う。ただし、入力となる関係は2章で述べたような既存の関係抽出手法を用いて与えられることを想定し、関係抽出手法自体については対象外とする。なお、本研究で用いる二項関係データは、3.3節で述べるように簡易な方法を用いて獲得する。表1は恒久性と一意性に基づき、関係を判別した例を示している。

3.3 二項関係データの作成

提案手法に用いる二項関係の獲得法について述べる。2節で述べたように、二項関係の獲得に関しては、様々な手法が研究されている。しかし、本研究ではどの手法を用いても研究の本質は変わらないため、今回は以下の手順により、文章の係り受け構造を用いた簡略な手法を用いた。

(1) 各文章の中で、固有表現(ここでは固有名詞を含む名詞連続)を二つ以上含むものを抽出する。

(2) 任意の二つの固有表現ペアに対し、最初に出現したものを X , 最後に出現したものを Y に置き換える。

(3) 上記の各ペアに対し、 X と Y の間の最短係り受けパス上の文節を抽出し関係とする。

上記によって、 X, Y にそれぞれ置き換えられた固有表現と抽出された関係により二項関係が獲得される。図1はこの二項関係の獲得の概要を示している。なお、態素解析に MeCab^(注1), 係り受け解析に J.DepP^(注2)を用いた。また、MeCabの辞書に NAIST-jdic^(注3)を用いたが、

文章:

昨日、本田圭佑が所属するCSKAモスクワが快勝した。

形態素解析:

昨日、/ 本田圭佑が / 所属する / CSKAモスクワが / 快勝した。
固有名詞 固有名詞

係り受け解析:

昨日、 / Xが / 所属する / Yが / 快勝した。

獲得される二項関係:

X = 本田圭佑、Y = CSKAモスクワ、関係 = Xが所属するY

図 1: 二項関係データの作成例

固有表現をより多く特定するために、Wikipediaの記事タイトルを固有名詞として辞書の拡張を行った。

4. 提案手法

本節では、時系列ウェブテキストの頻度情報を用いた関係の分類について述べる。まず、アプローチについて述べた後に、恒久性と一意性それぞれに対する分類手法について述べる。最後に、提案手法で用いる時間移動窓の期間の設定について述べる。

4.1 アプローチ

まず、恒久性の分類におけるアプローチについて述べる。恒久性がない関係とは時間的に後項 Y の固有表現が変化する関係である。そのため、もし前項 X に固有表現が代入されたときの後項 Y に当てはまる固有表現の時系列データがあれば、ある二時点における分布が変化していると考えられる。

次に、一意性の分類におけるアプローチについて述べる。一意性のある関係は、時間情報を考慮しなかった場合、時間による変化のために一意性がない関係に見えてしまうときがある。しかしながら、先程の時系列データにおいて各時点における分布を見ることができれば、後項 Y の固有表現がある時点においては唯一存在していることを確認することができる。

このように、恒久性と一意性の分類を行うには、後項 Y の固有表現のある時点における分布が必要である。そこで、本研究では、時系列分布を一定期間で切り取る時間移動窓を設け、各時間移動窓における分布を用いることにした。

4.2 頻度情報を用いた関係の分類

図 2 は実際に時系列ウェブテキストから獲得された二項関係において、恒久性がなく、一意性がある関係の代表例として“X が所属する Y”，恒久性があり、一意性がない代表例として“X を流れる Y”を選び、それぞれ前項 X に“本田圭佑”，“東京”を代入したときの後項 Y の固有表現の時系列分布を図 2.1-a と図 2.2-a

に示した。ここでは、時間移動窓の期間を 6 ヶ月とし、2009 年 2 月～7 月における図 2.1-a 上の時間移動窓 $t_{(本田圭佑, 2009.2\sim7)}$ 内の分布を図 2.1-b, 図 2.2-a 上の同期間における時間移動窓 $t_{(東京, 2009.2\sim7)}$ 内の分布を図 2.2-b に示した。また、2010 年 2 月～7 月における図 2.1-a 上の時間移動窓 $t_{(本田圭佑, 2010.2\sim7)}$ 内の分布を図 2.1-c, 同様に図 2.2-a 上の時間移動窓 $t'_{(東京, 2010.2\sim7)}$ 内の分布を図 2.2-c に示した。つまり、図 2.1-b の時間移動窓を 1 年分未来へスライドさせたときの分布が図 2.1-c となっている。

4.2.1 恒久性の分類

図 2 において、“X が所属する Y” は、恒久性のない、“X を流れる Y” は恒久性のある関係の代表例となる。ここで、図 2.1-b では最も頻度の高い固有表現は“VVV フェンロ”であるが、図 2.1-c では“CSKA モスクワ”が最も頻度が高い。これは、本田圭佑の所属先が“VVV フェンロ”から“CSKA モスクワ”へ変わったという事実が分布に現れていると言える。一方で、図 2.2-b と図 2.2-c を比較すると、先程に比べて分布の変化が小さいことが分かる。これは、ある地域を流れる川は時間によって変化しないためである。本研究ではこの違いに着目し、時間移動窓内の分布同士の変化をコサイン類似度を用いて計算する。この際、ベクトルの各成分は時間移動窓内における各固有表現の頻度とする。このようにすることで、例えば、図 2.1-b の分布と図 2.1-c の分布のベクトル同士のコサイン類似度は 0 となる。このように分布が大きく変化している場合、類似度は低くなる。

全時間移動窓間における、隣接する時間移動窓同士のコサイン類似度を計算し、最大値、最小値及び平均値を特徴量として用いた。これらは、式(4-1)、式(4-2)及び式(4-3)で表される。

$$F_{max} = \frac{1}{\sum_{x \in X_n} (|T_x|)} \sum_{x \in X_n} \max_{t_x, t'_x \in T_x, t_x \neq t'_x} \cos(t_x, t'_x) \quad (4-1)$$

$$F_{min} = \frac{1}{\sum_{x \in X_n} (|T_x|)} \sum_{x \in X_n} \min_{t_x, t'_x \in T_x, t_x \neq t'_x} \cos(t_x, t'_x) \quad (4-2)$$

$$F_{ave} = \frac{1}{\sum_{x \in X_n} (|T_x|)} \sum_{x \in X_n} \sum_{t_x, t'_x \in T_x, t_x \neq t'_x} \cos(t_x, t'_x) \quad (4-3)$$

なお、 X_n は X の頻度上位 n 位の固有表現の集合、 T_x は前項 X にある固有表現 x を代入したとき、全時間移動窓の中で後項 Y の固有表現が一つ以上存在している移動窓の集合である。実験では n=5 とした。

4.2.2 一意性の分類

一意性がある関係では、どの時間移動窓上においても後項 Y の固有表現は一つであることが多く、一方で一意性がない関係では Y が複数存在することが多いことに着目し、分類に用いる素性を二種類用いた。

(注 1) : <http://mecab.sourceforge.net/>

(注 2) : <http://www.tkl.iis.u-tokyo.ac.jp/~ynaga/jdepp/>

(注 3) : <http://sourceforge.jp/projects/naist-jdic/>

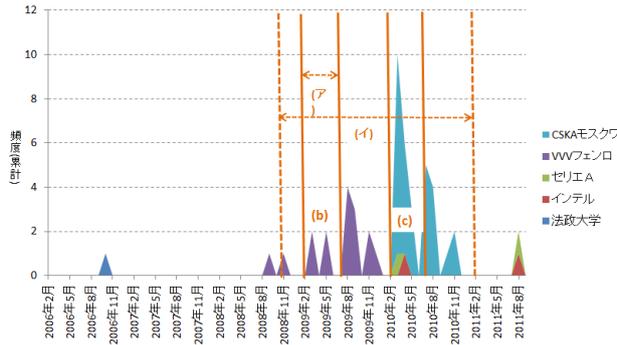


図 2.1-a: 関係“X(=本田圭佑)が所属する Y”の時系列分布

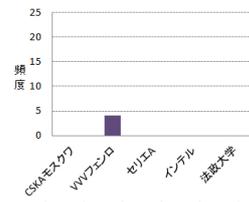


図 2.1-b: 2009年2月～7月における移動窓内の分布

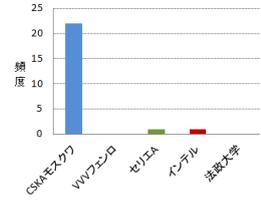


図 2.1-c: 2010年2月～7月における移動窓内の分布

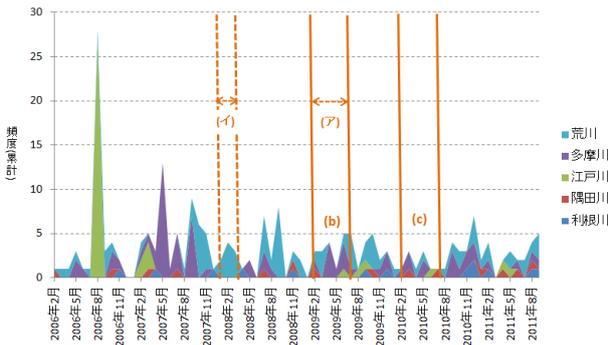


図 2.2-a: 関係“X(=東京)を流れる Y”の時系列分布

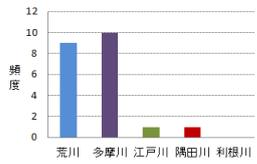


図 2.2-b: 2009年2月～7月における移動窓内の分布

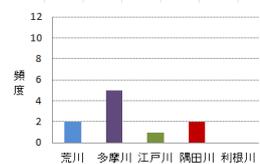


図 2.2-c: 2010年2月～7月における移動窓内の分布

図 2: 二項関係の時系列分布例

(a) 時間移動窓内における固有表現の種類数

一意性がある関係は、ない関係と比べ出現する固有表現の種類数が少ない。例えば、図 2.1-b では 1 種類 (VVV フェンロ) であるのに対し、図 2.2-b では 4 種類 (荒川, 多摩川, 江戸川, 隅田川) である。このような特徴を分類の素性に用いるため、本研究では、各時間移動窓における後項 Y の固有表現の種類数を計算し、全移動窓における最大値, 最小値, 平均値を用いた。これらの式は、それぞれ式(4-4), 式(4-5)及び式(4-6)で表される。

$$F'_{max} = \frac{1}{\sum_{x \in X_n} |T_x|} \sum_{x \in X_n} \max_{t_x \in T_x} \text{type}_Y(t_x) \quad (4-4)$$

$$F'_{min} = \frac{1}{\sum_{x \in X_n} |T_x|} \sum_{x \in X_n} \min_{t_x \in T_x} \text{type}_Y(t_x) \quad (4-5)$$

$$F'_{ave} = \frac{1}{\sum_{x \in X_n} |T_x|} \sum_{x \in X_n} \sum_{t_x \in T_x} \text{type}_Y(t_x) \quad (4-6)$$

なお、 $\text{type}_Y(t)$ は時間移動窓 t 内において出現する Y の種類数である。

(b) 頻度上位一, 二位間の割合

一意性がある関係では、ない関係と比べ頻度上位一位の固有表現 y_1 の頻度 $n(y_1)$ に対する 二位の固有表現 y_2 の頻度 $n(y_2)$ の割合が小さい。例えば、図 2 において、関係“X が所属する Y”は一意性がある、“X を流れる Y”は一意性がない関係の代表例である。図 2.1-b では最も頻度の高い固有表現 y_1 は“VVV フェンロ”であ

り、その次に頻度高い固有表現 y_2 は“CSKA モスクワ”である。よって、これらの頻度の割合は $n(y_2)/n(y_1)=0$ となる。一方で、図 2.2-b では y_1 = “多摩川”, y_2 = “荒川”であり、これらの頻度の割合は $n(y_2)/n(y_1)=0.9$ となり、図 2.1-b の方の割合が小さいことが分かる。

このような特徴を分類の素性に用いるため、本研究では各時間移動窓において頻度上位一, 二位間の頻度割合を計算し、全移動窓における最大値, 最小値, 及び平均値を素性として用いた。これらの式は式(4-4), 式(4-5), 式(4-6)のコサイン類似度を頻度割合に置き換えたものであるためここでは省略する

4.3 時間移動窓の期間の設定

4.2 節では、時間移動窓に基づく素性について、移動窓の期間を 6 ヶ月とした場合を述べた。しかし、実際には、時間移動窓の期間は狭くし過ぎると恒久性のない関係と一意性のない関係との区別がつかなくなる。例えば、図 2.1-a において、時間移動窓(ア)のように適切な期間であればいいが、もし(イ)のように必要以上に期間が広ければ“VVV フェンロ”と“CSKA モスクワ”が同時に成り立ち、一意性のない関係のように見えてしまう。

一方で狭くし過ぎるとデータスパースネスの問題が発生し、一意性のない関係の特徴が捉えにくくなり、一意性のある関係や不変性のない関係との区別がつかなくなる。例えば、図 2.2-a では、時間移動窓(ア)のよ

分類	ベースライン	1 ヶ月	3 ヶ月	6 ヶ月	12 ヶ月	24 ヶ月	提案手法
恒久性	0.572	0.549	0.561	0.590	0.598	0.566	0.608
一意性	0.576	0.646	0.662	0.646	0.663	0.606	0.667

表 2:実験結果(精度)

うに適切な期間であればいいが、もし(イ)のように必要以上に期間が狭ければ固有表現が時系列上において移り変わるように捉えたり、後項 Y の固有表現が唯一しているように捉えてしまう恐れがある。

このように、時間移動窓の期間にはトレードオフが存在するため、本研究では時間移動窓の期間を複数設定し、各期間において 4.2 節の素性をそれぞれ別として同時に学習器に学習させることでこの問題の解決を試みた。なお実験では、この期間を 1, 3, 6, 12 ヶ月の四段階に設定した。

5. 評価実験

5.1 実験設定

時系列ウェブテキストとして、我々が蓄積した日本語のブログ記事(2006年2月~2011年9月)を用いた。これらは、約 23 億文からなっている。このブログ記事から、3.3 節で述べた手法によって二項関係データを作成し、頻度上位 1,000 関係に対して人手によってラベルを付けた。なお、獲得された二項関係の中には“X に行く Y”や“X になる Y”といった、関係を示すものとは言い難いものも含まれているため、これらは人手によって除いた。機械学習には学習器として線形 SVM を用いる。実験では、これが実装された LIBLINER^(注 4)を用い、五分割交差検定を行った。

5.2 実験結果

評価実験の結果は表 2 のようになった。なお、ベースラインについては、恒久性と一意性両者ともマジョリティクラス(それぞれ、恒久性のない関係、一意性のない関係)に全てを分離したときの精度とした。

実験結果から、頻度情報に基づく素性を学習に用いることで、ベースラインに比べ精度が向上したことが分かる。

また、4.3 節において、時間移動窓の期間を複数設定して学習させる手法を述べた。ここでは、その手法の有効性を評価するために、恒久性及び一意性の分類において、時間移動窓の期間を 1, 3, 6, 12, 24 ヶ月のどれか一つのみを設定したときと、提案手法を用いたときの分類結果を表 2 に示した。評価結果より、必ずしも時間移動窓の期間を狭くするほど、もしくは広くすればするほど精度が向上するわけではないことが分かる。そのため、精度を向上させるためには、関係ご

とに時間移動窓の期間を設定しなければならないことが分かる。しかし、恒久性及び一意性の分類ともに提案手法の精度が最も高いことから、そのような関係ごとに適切な期間を設定するためのコストをかける必要はなく、何段階か期間を設定しそれらの素性を全て学習させる提案手法が有効であることが示された。

5.3 考察

評価実験により、提案手法のある程度の有効性を示すことができた。しかし、本研究ではさらなる精度向上の必要があると考えており、今回の実験において分類に誤りがあったものを分析し、誤分類の主な要因を調査した。

表層格による関係の曖昧性 本研究では、関係を語彙統語パターンとして扱っているため、複数の異なる深層の意味として用いられる助詞によって一意にラベル付けできないことが原因として挙げられる。例えば、関係“X が Y でデビューする”では、

(3) a. (サザンオールスターズ, X が Y でデビューする, 勝手にシンドバッド)

b. (東方神起, X が Y でデビューする, 日本)
 という二項関係が獲得された。(3a)の語彙統語パターンで用いられている助詞“で”は、手段を意味しており、この二項関係はミュージシャンとデビュー曲の関係を表している。このとき、関係は恒久性があり、かつ一意性があると考えられる。しかしながら、一方で(3b)の語彙統語パターンで用いられている助詞“で”は、場所を表しており、この二項関係はミュージシャンとデビューした場所の関係を表している。このときは、関係は恒久性がなく、かつ一意性がないと考えられる。このように、語彙統語パターンのみを関係とすると、関係に複数の意味が存在してしまうため一意にラベルを付けると誤分類の原因になる。今後としては、固有表現 X, Y に(LOCATION)や(WORK)といった型の制約を設けることを考えている。

話題性による頻度の偏り ブログ上の話題性によって、書かれない情報があるために分類に十分な方法が得られない。例えば、“X の子の Y”といった関係では、子供が複数いるにも関わらず、話題性のある子にしかブログ上で言及さえないため、子供が一人であると分類してしまう。このような問題を解決するために、今後は言語情報などを用いた素性を学習に加えることを考えている。

(注 4) : <http://www.csie.ntu.edu.tw/~cjlin/liblinear/>

6. おわりに

本稿では、関係抽出によって獲得される関係において、複数の二項関係が獲得された場合、それらは並立しているのか、それとも時間によって変化しているため複数獲得されたのか判断することが困難であるという問題に着目した。そして、関係を「時間において恒久的に成立するか否か」、また「一対一で成立するか否か」という二つの観点から分類することを提案した。さらに、時系列ウェブテキストの頻度情報を基に、恒久性と一意性それぞれに対して分類する手法を提案し、評価実験により提案手法の有効性を示した。

今後の課題としては、さらなる精度向上が挙げられる。そのために、本研究では獲得された二項関係を含む文章を活用することを考えている。文章中の言語情報を素性として加えることによって、分類の精度が向上するかどうか検討していきたい。

参 考 文 献

- [1] Deepak Ravichandran and Eduard Hovy, “Learning Surface text Patterns for a Question Answering System”, In Proceedings of ACL, pp. 41-47, 2002.
- [2] Alan Ritter, Doug Downey, Stephen Soderland and Oren Etzioni, “It’s a Contradiction-No, it’s Not: A Case Study using Functional Relation”, In Proceedings of EMNLP, pp. 11-20, 2008.
- [3] Razvan C. Bunescu and Raymond J. Mooney, “A Shortest Path Dependency Kernel for Relation Extraction”, In proceedings of HLT-EMNLP, pp. 724-731, 2005.
- [4] Michele Banko, Muchel J Cafarella, Stephen Soderland, Matt Broadhead and Oren Etzioni, “Open Information Extraction from the Web”, In Proceedings of IJCAI, pp. 2670-2676, 2007.
- [5] Michele Banko and Oren Etzioni, “The Tradeoffs between Open and Traditional Relation Extraction”, In Proceedings of ACL, pp. 28-36, 2008.
- [6] Fei Wu and Daniel S.Weld, “Open Information Extraction using Wikipedia”, In Proceedings of ACL, pp. 118-127, 2010.
- [7] Yusuke Shinyama and Satoshi Sekine, “Preemptive Information Extraction using Unrestricted Relation Discovery”, In Proceedings of HLT-NAACL, pp. 304-311, 2006
- [8] Danushka Bollegala, Yutaka Matsuo and Mitsuru Ishizuka, “Relational Duality: Unsupervised Extraction of Semantic Relations between Entities on the Web”, In proceedings of WWW, pp. 151-160, 2010.
- [9] Thomas Lin, Mausam and Oren Etzioni, “Identifying Functional Relation in Web Text”, In Proceedings of EMNLP, pp. 1266-1276, 2010.
- [10] Xiao Ling and Daniel S.Weld, “Temporal Information Extraction”, In Proceedings of AAAI, 2010.