DBSJ Journal Vol. 12, No. 1
June 2013

時期依存性を有するイベント連鎖の獲得

Acquiring Time-specific Event Chains

中島 直哉♥　吉永 直樹♦
鍛治 伸裕♠　豊田 正史♣
喜連川 優†

Naoya NAKASHIMA　Naoki YOSHINAGA
Nobuhiro KAJI　Masashi TOYODA
Masaru KITSUREGAWA

本論文では，時期依存性を有するイベント連鎖を時系列ウェブテキストから獲得する手法を提案する．提案手法は，言語的手がかりとイベントの時系列頻度を素性に用いた分類器により，テキストから獲得したイベント対から時期依存性を有するイベント連鎖を選別する．分類器の学習データは人手で用意した学習データを適合性フィードバックにより増補することで効率的に構築する．実験では，約 7 年分の日本語ブログアーカイブから四季依存のイベント連鎖を獲得し，得られたイベント連鎖を人手で評価した．

This article proposes a method of acquiring time-specific event chains from time-series text. A classifier with features based on language-based clues and time-series event frequency selects time-specific event chains among event pairs acquired from text. The training data for the classifier is effectually built by expanding manually-labeled examples using relevance feedback. We acquired season-specific event chains from our seven-year Japanese blog archive, and manually evaluated the acquired event chains.

1. はじめに

近年，ブログやマイクロブログの普及に伴い，一般の人々が自身の行動や見聞きした出来事をウェブ上に記述し，発信・共有するようになってきている．このような記述には，人間が常識的に持つ知識を前提とした行動や出来事が含まれる．例えば，熱があれば風邪を引いたのでははないかと疑うだろうし，実際風邪を引けば風邪薬を飲むだろう．この種の常識的・背景的な知識は，計算機が自然言語を理解する上で必要不可欠なものであるが，人手で陽に書き尽くすことは困難であり，どのように計算機が処理しやすい形で整備するかが課題となっている．そこで，このような連続する行動や出来事をテキストからイベント連鎖（あるいは推論規則）として獲得し，知識として蓄える研究が行われている [3, 2, 1].

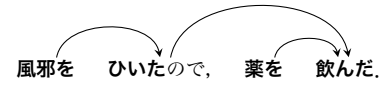

図 1　係り受け解析を用いたイベントの抽出
Fig. 1　Event extraction by using dependency parsing

さて，獲得されたイベント連鎖を行動推薦や発話生成などへ応用する際には，対象ユーザがどのような状況下にいるかを考慮に入れることが望ましい．例えば,「窓を閉める」と「冷房をつける」というイベント連鎖は，夏という季節の下で利用されるべき知識である．しかし，既存のイベント連鎖獲得研究では，文脈によらず成立するイベント連鎖を獲得することに主眼が置かれており，それらの連鎖がどのような状況下（以下，文脈）で成立するかは意識されていない．

本研究では，連鎖が成立する文脈として時期に注目し，ウェブから収集した大規模時系列ウェブテキストから時期依存性を有するイベント連鎖知識を獲得する手法を提案する．本研究では，単純のためイベント連鎖が成立する時期として春夏秋冬の四季を考え，それぞれの季節に成立しうるイベント連鎖を獲得する．具体的にはまず，書かれた季節ごとに分割した時系列ウェブテキストから，係り受け解析を用いてイベント連鎖候補を抽出する．このイベント連鎖候補から，接続表現などの言語的な手がかりや連鎖候補中のイベントの季節別頻度に基づく素性を用いた分類器により，時期依存性を有するイベント連鎖を分類・選別する．

実験では，約 7 年分の日本語ブログアーカイブを知識源として提案手法を適用し，得られたイベント連鎖を人手で評価した．

2. イベント連鎖と時期依存性

本章では，本研究で扱うイベントを定義し，その連鎖と時期依存性について議論する．

2.1 イベント連鎖

本研究では，係り受け関係にある名詞，格助詞，動詞の 3 つ組をイベントと定義する．ここで，格助詞は省略可能とし，動詞は肯定か否定の情報を保持するものとする．例えば図 1 のように,「風邪をひいたので，薬を飲んだ」という文からは，係り受け解析により “風邪をひく” と，“薬を飲む” という 2 つのイベントが獲得される．なお本研究では，動詞に複数の名詞が係る場合には，それぞれを別イベントとして取得する．

♥ 非会員　NTT コミュニケーションズ株式会社
naoya.nakashima@ntt.com
♦ 非会員　東京大学生産技術研究所
ynaga@tkl.iis.u-tokyo.ac.jp
♠ 正会員　東京大学生産技術研究所
kaji@tkl.iis.u-tokyo.ac.jp
♣ 正会員　東京大学生産技術研究所
toyoda@tkl.iis.u-tokyo.ac.jp
† 正会員　国立情報学研究所/東京大学生産技術研究所
director-general@nii.ac.jp/kitsure@tkl.iis.u-tokyo.ac.jp

次に，イベント連鎖を，イベント対 $\langle X, Y \rangle$ のうち，イベント X が起きた際に，それに続いてイベント Y が起きることに不自然性のないようなイベント対，と定義する．以下にイベント連鎖の例を示す．これ以降，イベント連鎖を $X \rightarrow Y$ と記述し，X をイベント連鎖の前項，Y を後項と呼ぶこととする．

(1) a. 風邪をひく → 薬を飲む
b. 外に出る → 公園を散歩する
c. 風呂に入る → 体を洗う
d. 風呂に入る ↛ 山に登る

特に 1b では，外に出た際に毎回公園を散歩するわけではないが，外に出た後の行動として，公園を散歩するというのは不自然ではない．このような関係についても，本研究では獲得の対象とする．ただし，1d のように，連続して起こることが稀で，人間が見て連続して起きるのは不自然と感じるイベント対は獲得の対象としない．連鎖性の判断は個人の主観に影響されうるが，経験的にはある程度の一致が得られることを 4.2.2 節の実験で示す．

2.2 連鎖の時期依存性

イベント連鎖のうち，ある時期においてのみ成立しうるイベント連鎖を時期依存性を有するイベント連鎖とする．本稿では単純のため，時期として季節（春夏秋冬）を考える．以下に時期依存性を有するイベント連鎖の例を示す．以降，イベント連鎖が依存する季節を → の上に記述することで時期依存性を有するイベント連鎖を表現する．

(2) a. 散歩に行く $\overset{\text{春}}{\rightarrow}$ 桜を見る
b. 窓を閉める $\overset{\text{夏}}{\rightarrow}$ 冷房をつける
c. 山に登る $\overset{\text{秋}}{\rightarrow}$ 紅葉を見る
d. 公園に行く $\overset{\text{冬}}{\rightarrow}$ 雪だるまを作る

例えば,「散歩に行っ」て「桜を見る」というイベント連鎖が不自然でないのは春のみである．したがって，2a の例のように春に成立する連鎖として獲得する．

ここで示したようなイベント連鎖は，そもそも時期依存性を考慮しなければ連鎖の蓋然性に乏しく，従って獲得し難い知識であることに注意されたい．すなわち，時期依存性を考慮することで，より多くのイベント連鎖の獲得に繋がるといえる．

3. 提案手法

本章では，テキストから係り受け解析を用いて獲得したイベント対を，季節関係なく生じるイベント連鎖，春夏秋冬それぞれに起こるイベント連鎖，連鎖しないイベント対のいずれかに，機械学習の多クラス分類器を用いて分類する手法について説明する．

分類器の素性としては，イベント間に現れる言語的な手掛かりに加え，時系列ウェブテキストにおけるイベントの頻度情報を用いる．また，分類器の学習に用いる学習データを，適合性フィードバックを応用することで効率的に構築する手法も提案する．以下で，それぞれの手順について順に詳しく説明する．

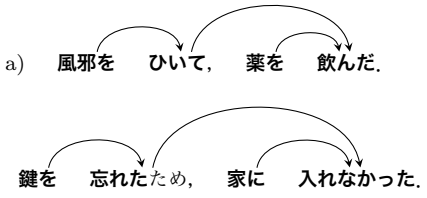

図 2 係り受け関係にあるイベント対
Fig. 2 Event pairs in a dependency relation

表 1 手がかり表現として用いたイベント間の接続表現
Table 1 Conjunctives between events used as clues

ため，が，なら，から，ながら，と，れば，ば，らば，たら，ので，のに

3.1 イベントの抽出とイベント連鎖候補の獲得

まず，係り受け解析を用いてテキストから係り受け関係にある名詞,（格助詞,）動詞をイベント候補として収集する．ここで，時相名詞，形式名詞，数詞などの名詞や “する”, “ある” などの軽動詞を含むイベントは，イベントとしての具体性に乏しいことから収集対象外とした．このようにして収集したイベントを，動詞を全て読みに正規化[1]して集計し，低頻度のイベントを切り捨てることでイベント連鎖候補の要素とするイベント集合を得る．

次に，上記の手順で得られたイベント集合から，各イベントの組み合わせをイベント連鎖候補として列挙する．抽出されたイベントの全組み合わせを考慮すると候補の数が爆発するため，本研究ではテキスト中で図 2a のように連用形で直接係り受け関係にあったイベント対のみをイベント連鎖候補とした．

3.2 素性抽出

本節では，3.1 節で得られた各イベント連鎖候補に対し，言語的な手がかりとイベントの時系列頻度に基づく素性を抽出する方法についてそれぞれ説明する．

■言語的な手がかりに基づく素性 連鎖するイベント対は，図 2 のように文中で連用形または特定の接続表現により直接的または間接的に係り受け関係にあることが多い．そこで，言語的な手がかりに基づく素性として，文中でイベント対が連用形あるいは特定の接続表現を介して係り受け関係にあった頻度を用いる．接続表現には乾らの研究 [3] を参考に，表 1 に示す接続表現を用いた．

次に，イベント連鎖候補中の各イベント（の動詞）が特定のモダリティを伴って文中で出現した頻度を素性として用いる．例えば，後項に “〜らしい” のような推量の表現が多用されるイベント対は，実際には連鎖する蓋然性が低いと考えられる．また，各イベントの時制は，連鎖する順序に影響する．例えば「薬を買うため，薬局に 行った」という文では，実際のイベントは，薬局に行く → 薬を買う，という順で連鎖する．そこで，イベントが過去時制を伴っていた頻度を素性とすることで，このような例に対処する．表 2 に素性として利用した接尾表現を示す．時制につい

[1] 「風邪を（ひく，引く）」のような表記揺れを吸収するため．

表 2　手がかり表現として用いた動詞の時制・モダリティ
Table 2　Tense and modality of verbs used as clues

た，だろう，らしい，れる，られる，せる，させる

表 3　素性の構成
Table 3　Details of features

	言語的手がかり	ウインドウ共起	季節別頻度
種類数	36	3	28

ては，前項のみ過去の場合と後項のみ過去の場合，そして，両項が過去の場合に分けて頻度を数える．以上で取得される素性は，イベント対の連鎖性の判定に寄与すると期待できる．

なお，これら言語的手がかりについては，前項と後項を入れ換えた場合についても別途考慮し，その頻度を素性とする．

■イベントの（時系列）頻度に基づく素性　次にイベントの頻度情報に基づく素性について述べる．本研究では具体的には，イベント対のウインドウ内共起に基づく素性と，イベント中の単語やイベント自身の季節別頻度を利用した素性を用いる．このうち，前者はイベント対の連鎖性の判定に，後者はイベント対の時期依存性の判定に寄与すると期待できる．

イベント対のウィンドウ内共起　連鎖するイベントは，テキスト中で近接して記述されやすい．そこで，イベント対の共起頻度を同一文内，連続する 3 文内，同一のウェブページ内という 3 つのウィンドウごとにそれぞれ数え，素性とする．このとき，イベントの出現順を考慮して共起頻度を数える．ウィンドウを複数用意するのは，イベントが共起した際の位置関係で連鎖の強さが変化するためである．一般的に，より近い位置で共起したイベントの方がより強い連鎖性を有すると考えられる．3 つのウィンドウを設定することで，連鎖性の強さを細かく捉えることができる．

イベント対の季節別頻度　時期依存性を有するイベント対は，特定の季節において共起頻度が高くなるはずである．そこで，頻度計算に用いる時系列ウェブテキストを書かれた季節ごとに分け，各季節のもとでの共起頻度をそれぞれ数える．具体的には，3, 4, 5 月を春，6, 7, 8 月を夏，9, 10, 11 月を秋，12, 1, 2 月を冬とし，それぞれウェブページ内でのイベント対の共起頻度を数える．

また，時期依存性を有するイベント連鎖では，イベントそれ自身が季節性を有していたり，イベントに含まれる単語（名詞や動詞）に季節性があるものも多い．そこで，イベント中の単語やイベントの季節別頻度も素性とする．これに関しても，上記と同様に時系列ウェブテキストを季節で分け，前項・後項の名詞・動詞のそれぞれについて，各季節のもとでの出現頻度を数える．

表 3 に分類器で用いた素性をまとめた．素性値には頻度を 0-1 の値を取るように正規化した実数値を用いた．ただし，季節別のイベント頻度については，頻度の比を保って正規化した．

3.3　適合性フィードバックを用いた学習データの構築

3.1 節で述べたイベント連鎖候補から時期依存性を有するイベント連鎖を選別するため，一部の候補に正解ラベルを付与して得られた学習データから分類器を学習する．ここで，単純に獲得したイベント連鎖候補から無作為に候補を取り出して正解ラベルをつけるというアプローチを取ると，収集したイベント連鎖候補の所属するクラスの分布の偏りから，得られる正解ラベル数に大きな偏りが生じる．そこで我々は，適合性フィードバックを用いて効率的にラベル付けを行う手法を考案した．

まず，候補の大部分がイベント連鎖を構成しないという観察から，素性の連用形接続の頻度が上位のイベント連鎖候補を対象として正解ラベルを付与する．さらに，季節別のイベント連鎖を効率的に増補するため，季節ごとのイベントの共起頻度の偏りを最大頻度と次点との差の大きさに基づいて算出し，各季節ごとに偏りの大きいものから順に正解ラベルを付与した．このようにして得られた学習データセットを，以後，基本学習データ呼ぶ．

次に，この基本学習データを，適合性フィードバックを応用することで拡張し，拡張学習データを得る．具体的には，まず，基本学習データから分類器を学習し，その分類器を用いて連用形接続の頻度が上位 100 万件以内の未分類イベント連鎖候補を分類する．次に，得られた分類結果で各クラスに対する所属確率[2]が高いイベント連鎖候補から順に人手でラベルの修正を行い，基本学習データに加えることで，拡張学習データを得る．

4.　評価実験

本章では，前章で提案した手法を用いて時系列ウェブテキストから時期依存性を有するイベント連鎖を獲得し，得られたイベント連鎖を人手で評価することで提案手法の有効性を検証する．イベント収集のための係り受け解析には J.DepP[3]を，分類器にはロジスティック回帰を実装した LIBLINEAR[4]をそれぞれ用いた．

4.1　実験設定

2006 年 2 月から 2012 年 11 月まで継続的に収集して構築した日本語ブログアーカイブ（約 24 億文）から，3.1 節で述べた方法を用いてイベントの抽出を行った．具体的には，全ブログ記事中で約 1000 回以上出現する約 17 万イベントを連鎖の獲得対象とした．これらのイベントを組み合わせ，最終的にイベント連鎖候補として約 700 万件のイベント対を得た．

次に，3.3 節で説明した手法に従い，学習データを構築した．まず，人手により連用接続の頻度上位 300 件のイベント対に対して正解ラベルを付与した．また，同上位 10,000 件のイベント対から季節性の高いものを 700 件抽出し，それらについてもラベルを付与した．結果として，表 4 中段に示すように 436 イベント対からなる基本学習データを得た．

さらに，この基本学習データから分類器を学習し，適合性フィードバックを利用して学習データを増補した．その結果として，フィードバック前の約 2 倍となる 883 イベント対からなる学習データを得た．最終的に得られた拡張学習データにおける各ラベルの内訳を表 4 下段に示す．人手によるラベル付与作業でほ

[2] 今回分類器として用いたロジスティック回帰モデルでは，出力として各クラスに対する所属確率が得られる．
[3] http://www.tkl.iis.u-tokyo.ac.jp/~ynaga/jdepp/
[4] http://www.csie.ntu.edu.tw/~cjlin/liblinear/

表 4　学習データの構築: 結果
Table 4　Construction of training data: Results

	連鎖しない	季節なし	春	夏	秋	冬	合計
人手のみ	282	52	20	26	9	47	436
提案手法	393	149	72	95	53	121	883

表 5　各クラスに分類されたイベント対の数
Table 5　The number of event pairs classified into each class

	季節なし	春	夏	秋	冬
分類数	78,763	7165	8998	810	7295

とんど学習データを得られなかった秋に特有のイベント連鎖に関しては，適合性フィードバックによりおよそ 6 倍の量の正例を得ることができており，本手法の有用性を確認できた．時期依存性を有する連鎖を見てみると，夏と冬の学習データが多くなっている．これは，気候的な特徴が顕著な季節の方が，その時期特有のイベントが起こりやすいためと考えられる．拡張して得られた学習データの有用性については，4.2.2 節で検証する．

4.2　実験結果

883 イベント対からなる拡張学習データから学習した分類器を用いて，連用接続の頻度が上位 100 万件以内の未分類のイベント対を分類した．表 5 に各クラスに分類されたイベント対の数を示す．季節性のあるイベント連載の総数は，季節性のないイベント連鎖の約 30% ((7165 + 8998 + 810 + 7295) / 78763) であった．

次に，分類されたイベント対の中にどれだけの量の正しいイベント連鎖が存在するかを見積もった．まず，各クラスごとに分類されたイベント対を無作為に 500 件抽出し，以下の基準に従って人手で正誤判定を行った．

1. 前項から後項が連続して起こることが妥当か
2. 時間的な順序関係が適切か
3. 季節性を有する場合，特定の季節のみに依存すると言えるか

その後，クラスに対する所属確率が一定以上のイベント連鎖候補のみを選び，その条件下での適合率を算出する．例えば，所属確率を p とし，$p > 0.7$ における分類結果の適合率は，以下の式により求められる．

$$\text{適合率}_{p>0.7} = \frac{p > 0.7 \text{ のイベント対中の正解数}}{p > 0.7 \text{ のイベント対の数}}$$

季節なしと春夏秋冬の 5 つのクラスについて所属確率の閾値を 0.4 から 0.9 まで 0.1 刻みで変えて分類の適合率を算出した結果を図 3 に示す．結果として，所属確率が 0.7 以上あれば分類された数のおよそ半数以上は正しいイベント連鎖であることがわかった．また，春に特有のイベント連鎖のみ特に適合率が悪かった．これについては，4.2.1 節で誤分類の原因と関連づけて考察する．

さらに，提案手法を利用することで獲得できるイベント連鎖の

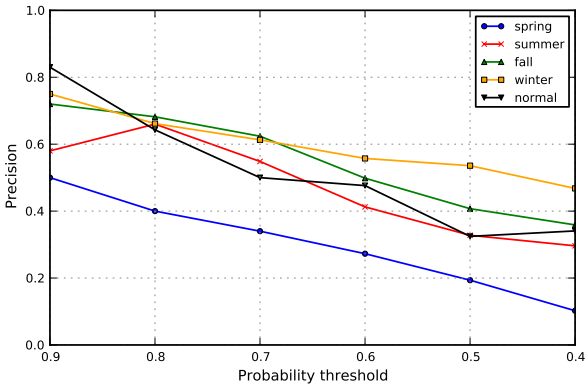

図 3　各クラスへの所属確率の閾値と適合率の関係
Fig. 3　Relation between threshold to probability and precision

表 6　獲得可能なイベント連鎖の概算
Table 6　The estimated number of obtainable event chains

	季節なし	春	夏	秋	冬
$p > 0.7$	2742	20	330	206	1302
$p > 0.5$	8677	507	2162	272	3712

数を以下の式により見積もる．

$$\text{獲得できるイベント連鎖の数} = \text{適合率} \times \text{分類数}$$

まず，表 6 に示すように，所属確率を 0.7 以上とすると，2000 件程度の時期依存性を有するイベント連鎖がおよそ 6 割の適合率で獲得可能だと推定される．また，所属確率を 0.5 以上とすると，夏や冬については数千件単位で時期依存性のあるイベント連鎖が得られる．このことから，本手法により時期依存性のあるイベント連鎖が相当数獲得できることが確認された．

実際に獲得されたイベント連鎖の例を表 7 に示す．それぞれの季節について直感的に妥当なイベント連鎖が獲得されていることが確認できる．また，表中の先頭のイベント連鎖は，どれも前項が「公園を訪れる」という同じ行為を表しているが，連鎖して起きるイベントは，季節ごとに異なっている．このように，前項に季節性は無いが季節性を有するイベント連鎖が獲得できていることは興味深い．

4.2.1　誤分類考察

前節の実験において，評価者により誤分類と判定されたイベント連鎖候補を季節なしと春夏秋冬の各クラスからそれぞれ 100 件ずつ無作為に抽出し，誤分類の原因を分析した．表 8 に誤分類の原因を示す．以下で，誤分類の数が多い原因から順に例を挙げて考察する．

■無関係のイベント対　分類結果には，カメラを持つ → 雨が降る，のような無関係のイベント対を連鎖すると誤分類した例が多くみられた．これは，ブログでは日常的なイベントに加えて偶然起きた印象深いイベントも記述されやすいため，連鎖しにくいイベント対も高頻度で共起することから誤分類したと考えられる．

表 7　獲得された時期依存性を有するイベント連鎖の例
Table 7　Examples of acquired time-specific event chains

春	夏	秋	冬
公園で遊ぶ$\overset{春}{\rightarrow}$桜を見る	公園に行く$\overset{夏}{\rightarrow}$花火を見る	公園へ行く$\overset{秋}{\rightarrow}$どんぐりを拾う	公園に行く$\overset{冬}{\rightarrow}$雪だるまを作る
雨が降る$\overset{春}{\rightarrow}$花粉が飛ばない	浴衣に着替える$\overset{夏}{\rightarrow}$花火大会に行く	落ち葉を集める$\overset{秋}{\rightarrow}$火をつける	雪が残る$\overset{冬}{\rightarrow}$路面が凍結する
桜が終わる$\overset{春}{\rightarrow}$ツツジが咲く	縁側に座る$\overset{夏}{\rightarrow}$スイカを食べる	京都に行く$\overset{秋}{\rightarrow}$紅葉をみる	気温が下がる$\overset{冬}{\rightarrow}$氷がはる
弁当を作る$\overset{春}{\rightarrow}$花見に出かける	窓を閉める$\overset{夏}{\rightarrow}$冷房をいれる	紅葉が終わる$\overset{秋}{\rightarrow}$葉を落とす	手袋を忘れる$\overset{冬}{\rightarrow}$手がかじかむ
風が吹く$\overset{春}{\rightarrow}$桜が散る	甲子園に行く$\overset{夏}{\rightarrow}$高校野球を観る	散歩に行く$\overset{秋}{\rightarrow}$紅葉をみる	気温が下がる$\overset{冬}{\rightarrow}$道路が凍る

表 8　各クラスごとの分類結果の誤り分析
Table 8　Analysis of misclassification for each class

誤分類の原因	季節なし	春	夏	秋	冬	合計
無関係のイベント対	36	26	33	42	37	174
イベントの情報不足	30	19	14	20	27	110
時期依存性の誤認	1	32	25	9	10	77
イベントの連鎖順	1	6	4	24	4	39
特定の出来事に由来	5	8	11	2	2	28
逆説的な連鎖	14	1	4	0	4	23
慣用的な表現	5	1	1	1	0	8
その他	8	7	8	2	16	41

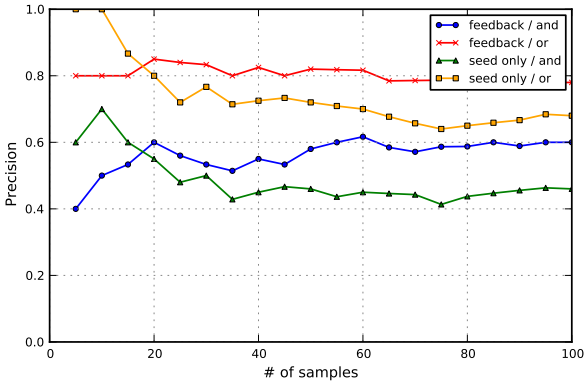

図 4　秋に特有の連鎖として分類されたイベント対の評価
Fig. 4　Evaluation of event pairs classified as fall-specific chains

■イベントの情報不足　本研究で定義した名詞,（格助詞,）動詞の 3 つ組で構成されるイベントには具体性が乏しいイベントが含まれており，結果として適切なイベント連鎖となりえないイベント対が連鎖候補となっていた．例えば，気持ちになる → 手を出す，のような不明瞭なイベント対については，人間にとっても連鎖性の有無を判定することが困難である．

■時期依存性の誤認　時期依存性に関する誤分類として，季節関係なく連鎖するイベント対を特定の季節に分類する例も多かった．例えば，風邪をひく → 病院に行く，というイベント連鎖は季節関係なく成立するが，冬に特有のイベント連鎖と分類された．これは，風邪をひくというイベントそのものが冬に起こりやすいことに起因すると考えられる．また，春に特有のイベント連鎖については，この要因による誤分類が最も多いことがわかった．

■イベントの連鎖順　また，雪が積もる → 雪が降る，のように前項と後項を入れ替えると成立するイベント連鎖を連鎖ありと誤分類する例も多くみられた．これについては，前項と後項を入れ替えたイベント連鎖候補と元のイベント連鎖候補で，抽出される素性に類似性があることが原因であると考えられる．

■特定の出来事に由来する連鎖　ブログ上で話題となった出来事について，記述の頻度が偏ることから誤って時期依存性を有すると誤分類される例もみられた．例えば，東日本大震災が 3 月にあったことから，地震がくる → 津波が発生する，のような地震関連のイベント連鎖が春に特有のイベント連鎖と誤分類された．また，同様にオリンピックなどが原因で，アメリカに勝つ → 金メダルをとる，のように本来は連鎖しないイベント対を連鎖すると誤分類する例もみられた．

■逆説的な連鎖　例えば，お腹が空く → ご飯を食べない，のように一方のイベントが肯定の表現に変わるとイベント連鎖として成り立つイベント対を誤分類する例もみられた．これについては「お腹がすいたけど，ご飯を食べられなかった」のような逆説的な記述がブログに多く書かれていたことが原因と考えられる．

■慣用的に用いられるイベント対　慣用的に用いられるイベント対を連鎖すると誤分類する例も散見された．例えば，背中を見る → 子供が育つ，というイベント対が連鎖性ありと誤分類されたのは,「（親の）背中を見て子供が育つ」という定型表現が，多くのブログで書かれていたためと考えられる．

4.2.2　適合性フィードバックの有効性の検証

最後に，学習データ構築の作成の際に利用した適合性フィードバックの効果を検証する．まず，基本学習データと適合性フィードバックを利用して得た拡張学習データを用いた分類器による分類結果から，所属確率の降順に各 100 件のイベント連鎖候補を抽出し，比較用のイベント対とする．この分類結果の正誤を 2 人の被験者にそれぞれ前述の基準に従って判定してもらった．なお，分類結果の正誤を判定する際には，基本学習データを用いて学習した分類器による分類結果と，適合性フィードバックを利用して得た拡張学習データを用いて学習した分類器による分類結果を合わせて無作為に並び替えて被験者に提示した．人手による評価の被験者間一致度は $\kappa = 0.558$ であり，中程度の一致が得られた．

図 4 に，フィードバックにより最も訓練例数が増加した秋につ

いて，クラスの所属確率上位から順に 5 個ずつサンプルした際の適合率の変化を示す．グラフ中で横軸はイベント連鎖候補数を，縦軸は適合率を示している．凡例で seed only 及び feedback はそれぞれ基本学習データと拡張学習データを用いて学習した分類器による分類結果である．また，and, or はそれぞれ，2 被験者の両方が適切と判定したとき正解とした場合の適合率，どちらか一方でも適切と判定したとき正解とした場合の適合率である．グラフから拡張学習データを用いることで高い適合率でイベント連鎖を分類できていることが確認できる．なお，紙面の都合上省略したが，その他の季節についても同様の傾向がみられた．

5. 関連研究

本稿で取り扱ったイベント連鎖は，推論規則や因果関係と呼ばれる事態間関係知識に相当し，これまで質疑応答や文書要約，機械翻訳，談話理解などへの応用を念頭にテキストから自動獲得する研究が行われている．

推論規則や因果関係は，事態対 $\langle X, Y \rangle$ で，X が成立する際に Y が蓋然的に起きる関係と定義されている．乾ら [3] は，必然性を有する因果関係を表現する “ため” という接続表現を利用して因果関係の獲得を試み，1 年分の新聞記事より 27,000 件以上の因果知識を獲得したと報告している．また Shibata ら [2] はイベントの共起に注目してイベント連鎖を獲得する手法を提案した．彼らの手法では，係り受け構造を利用してイベント連鎖の候補となるイベント対を取得し，事前獲得した格フレームを用いて省略されている項を補うことで，詳細なイベント連鎖を獲得する．結果として，約 1 億ウェブページ上から約 2 万件の推論規則を獲得したと報告している．

これらの研究では普遍的に成立するイベント連鎖を獲得しており，本研究で獲得を試みたような特定の時期でのみ成立するイベント連鎖については，蓋然性なしと判断され獲得できなかったり，獲得できたとしても（時期依存性に関する情報が付与されないことから）適切に利用することが難しいことが予想される．

6. おわりに

本論文では，イベント連鎖知識において，連鎖が成立する状況を考慮することが知識を獲得する上でも運用する上でも重要であることを議論した．その上で，特定の時期（本稿では四季）に成立するイベント連鎖を，言語的手がかり，イベント対の共起頻度，また各季節における単語やイベントの出現頻度を素性として利用した分類器により選別する手法を提案した．約 7 年分の日本語ブログアーカイブに対して提案手法を適用し獲得された季節別のイベント連鎖を人手で評価することで，本手法の有効性を示した．

本稿ではイベント連鎖を排他的に特定の季節に分類したが，実際には複数の季節で成立するイベント連鎖も存在する．このようなイベント連鎖を獲得するため，本稿で取り組んだイベント連鎖の分類を多ラベル分類として定式化することを考えている．さらに，時期依存性に留まらず空間依存性や，知識の利用者の属性情報など，様々な文脈を考慮したイベント連鎖の獲得を行いたい．

[謝辞]

本研究の一部は JSPS 科研費 13372965 の助成を受けたものです．

[文献]

[1] Q. Do, Y. Chan, and D. Roth. Minimally supervised event causality identification. In *Proceedings of EMNLP*, pp. 294–303, 2011.

[2] T. Shibata and S. Kurohashi. Acquiring strongly-related events using predicate-argument co-occurring statistics and case frames. In *Proc. IJCNLP*, pp. 1028–1036, 2011.

[3] 乾孝司, 乾健太郎, 松本裕治．接続標識「ため」に基づく文書集合からの因果関係知識の自動獲得．情報処理学会論文誌, Vol. 45, No. 3, pp. 919–933, 2004.

中島 直哉 Naoya NAKASHIMA
2011 東大・工・電子情報工学卒．2013 同大大学院情報理工学系研究科修士課程了．同年より NTT コミュニケーションズ株式会社勤務．

吉永 直樹 Naoki YOSHINAGA
2000 東大・理・情報科学卒．2002 同大大学院理学系研究科修士課程了．2005 同大大学院情報理工学系研究科博士課程了．博士（情報理工学）．2002 より 2008 まで日本学術振興会特別研究員 (DC1, PD)．2008 東京大学生産技術研究所特任研究員，特任助教を経て現在，同大学生産技術研究所特任准教授．計算言語学・機械学習の研究に従事．

鍛治 伸裕 Nobuhiro KAJI
2005 東京大学大学院情報理工学系研究科博士後期課程了．情報理工学博士．2007 東京大学生産技術研究所特任助教を経て現在，同大学生産技術研究所特任准教授．自然言語処理の研究に従事．

豊田 正史 Masashi TOYODA
東京大学生産技術研究所准教授．1994 東工大・理・情報科学卒．1996 同大大学院情報理工学研究科修士課程了．1999 同大学院情報理工学研究科博士後期課程了．博士（理学）．同年，科学技術振興事業団計算科学技術研究員．ウェブマイニング，ユーザインタフェース，ビジュアルプログラミングに興味をもつ．ACM, IEEE CS，情報処理学会，日本ソフトウェア科学会各会員．

喜連川 優 Masaru KITSUREGAWA
東京大学工学系研究科情報工学専攻博士課程修了 (1983 年)，工学博士．国立情報学研究所所長，東京大学生産技術研究所教授，東京大学地球観測データ統融合連携研究機構機構長，文部科学省「情報爆発」特定研究領域代表 (2005-2010)，経済産業省「情報大航海プロジェクト」戦略会議委員長 (2007-2009)．データベース工学の研究に従事．ACM SIGMOD Edgar F. Codd Innovation Award 受賞，ACM フェロー，IEEE フェロー，情報処理学会フェロー，電子情報通信学会フェロー．現在，内閣府最先端研究開発支援プログラムを中心研究者として推進中．