

説明可能 AI における上位概念を考慮した ナレッジグラフからの説明変数選択

A Variable Selection Method for Explainable AI using Hyponymy Relations of Knowledge Graph

福田貴三郎¹ 小柳佑介¹ 福田茂紀¹ 大倉清司¹

藤重雄大¹ 岩下洋哲¹ 大堀耕太郎¹

Takasaburo FUKUDA¹, Yusuke KOYANAGI¹, Shigeki FUKUTA¹, Seiji OKURA¹

Yuta FUJISHIGE¹, Hiroaki IWASHITA¹, Kotaro OHORI¹,

¹ 株式会社富士通研究所

¹ Fujitsu Laboratories Ltd.

Abstract: In recent years, there has been increasing interest in explainable AI. When designing a white-box machine learning model, it is important not only to increase the prediction accuracy of the model, but also to design a model with high interpretability. In this paper, we consider the case in which a knowledge graph is used as a resource for learning data, and propose a method for selecting explanatory variables with high interpretability by utilizing the hyponymy relations between entities. We further confirmed with a test using a dataset of professional baseball players that the method can indeed choose the desired explanatory variables.

1. はじめに

AI 技術はここ 10 年来で急速な発展を遂げており、様々な社会システムやビジネス領域に適用されてきている。特に近年ではただ精度向上を目指すだけでなく、AI が想定外の動作を行わないようにする事や、判断結果の根拠を示す事を目的とした説明可能 AI の技術が注目されている。

説明可能 AI では、人が解釈可能な特徴量を説明変数とした適切な機械学習モデルを設計する事が重要となる。一般的に機械学習のモデルを設計する過程において、データサイエンティストは 50%以上の時間をデータ収集及び説明変数の選択作業などに費やしていると言われている[1]。そのため、効率的にデータを収集し説明変数を選択する技術が求められており、そのような方法の一つとして、ナレッジグラフから説明変数の候補を自動で収集する手法が提案されている[2]。

ナレッジグラフから説明変数の候補を自動で収集する場合、エンティティ間の繋がりを基に収集するのが基本的なアプローチとなる。その際、エンティ

ティ間はグラフ形式で繋がっているため、説明変数の候補を選択する範囲を適切に決める必要がある。そのような選択する範囲を決定する方法として、本研究では part-of 関係などの上位概念を考慮する事の重要性に着目し、解釈性の高いモデル設計を実現するための説明変数の選択手法を提案する。

上位概念を考慮する事の重要性を説明するための題材として、本研究では「プロ野球選手がタイトルを獲得するかどうかに影響する特徴量は何か？」という問題を扱う。この問題では、過去に日本野球機構 (NPB) に所属したプロ野球選手の情報を学習データとして扱う。その際、説明変数の候補は主に各プロ野球選手の出身県などの属性情報、目的変数は各プロ野球選手がタイトルを獲得したか否かとなる。ここで架空の例として、図 1 のような関係性から説明変数を選択する場合を考える。図 1 を俯瞰した場合、人であれば直感的に「出身が四国地方かつ入団チームが球団 X である場合はタイトル (ベストナイン) を獲得する」という事が読み取れる。しかし、もし各プロ野球選手に直接繋がっているエンティティのみを説明変数として抽出した場合、各出身県の

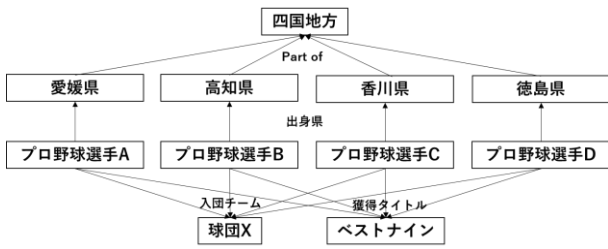


図 1. タイトルを獲得するプロ野球選手をナレッジグラフで表現した例

上位概念にあたる四国地方の情報は読み取る事が出来ない. そのような説明変数から設計された機械学習のモデルを人が理解しようとした場合, 「出身が愛媛県または高知県または香川県または徳島県かつ入団チームが球団 X の場合にタイトルを獲得する」という解釈となってしまう. これは, 先の図 1 全体から人が読み取った結果と比較した場合, 実質的に同じ意味ながら, 条件節が多く, 人が理解しにくい冗長的な解釈であると言える. 条件節の多さと解釈性の関係の例として, 分類木やアンサンブル木モデルでは予測のルールが細分化されると解釈が困難になる事が知られている[3]. そこで本研究では, 学習データのリソースとしてナレッジグラフを使用する場合において, 先述のような上位概念を考慮する事で, 必要に応じて上位概念のエンティティを説明変数として選択し, 解釈性の高いモデル設計を実現する手法を提案する.

2. 関連研究

機械学習モデルに使用する説明変数を n 個の候補 (x_1, \dots, x_n) から決定する問題は, 最良なモデルを実現する説明変数の組み合わせを 2^n 個の部分集合から決定する問題となり, NP 困難である事が知られている. 最良部分集合を決定する方法として, ヒューリスティックに決定する方法や[4], 数理計画法の枠組みで解く方法の研究が行われている[5]. このような変数選択問題ではモデルの情報量基準で行う事があり, 情報量基準として赤池情報量基準 (AIC) などがよく用いられる. ナレッジグラフからこのような変数選択を行う場合, あらかじめ決められた範囲のエンティティやそれらに繋がる上位概念などのエンティティを説明変数の候補とした変数選択問題に置き換える事が出来る. そこで実際にモデルの情報量基準として AIC を用いて変数選択を行う場合を考える.

AIC は推定されたモデルを評価するための基準の一つであり, 次の式で与えられる.

$$AIC = -2(\text{モデルの最大対数尤度}) + 2(\text{モデルの変数の数}) \quad (1)$$

この式で計算された AIC の値が小さい方が良いモデルであり, AIC の値が最小であるモデルを構成する説明変数の集合が最良のものとして選択される. 例えば「(出身県が) 愛媛県かつ (入団チームが) 球団 X」, 「高知県かつ球団 X」, 「香川県かつ球団 X」, 「徳島県かつ球団 X」の 4 つの仮説空間に基づき説明変数を選択した場合と, 「四国地方かつ球団 X」の仮説空間に基づき説明変数を選択した場合とで, モデルの最大対数尤度に差が無い場合は, 必然的に変数が少なくなる後者の方が AIC の値が低くなり, 最良の説明変数として選択される.

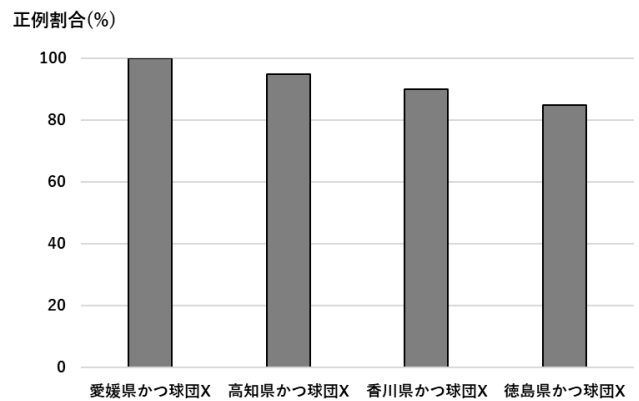


図 2. 仮説空間毎に正例割合にばらつきがある例

しかし, 例えば図 2 のように「愛媛県かつ球団 X」, 「高知県かつ球団 X」, 「香川県かつ球団 X」, 「徳島県かつ球団 X」の 4 つの仮説空間で目的変数への影響度 (正例割合) がそれぞれ高いものの, 各値に差がある場合を考える. このようなケースでは, モデル上で個々の仮説空間毎に重みに差をつけた方が「四国地方かつ球団 X」の仮説空間に基づきモデルを構築するより, モデルの最大対数尤度が高くなる可能性がある. その場合は前者に基づいたモデルの方が AIC の値が低くなる可能性があり, 後者のような上位概念は説明変数として選択されなくなる. 本研究では, このような場合でも, 解釈性の高いモデル設計を行う観点で, 前者のような個々の仮説空間に加えて, 後者のような上位概念も併せて説明変数として選択する手法を提案する.

ナレッジグラフから説明変数の候補を自動で収集する技術として, エンティティ間の繋がりを基に収集する研究が行われている[2]. 文献[2]では任意のエンティティにトリプル表現で繋がっているエンティティを説明変数の候補として抽出し, リスト化してユーザに提示を行い, ユーザが必要となる説明変数を選択する. この手法ではあるエンティティに直接

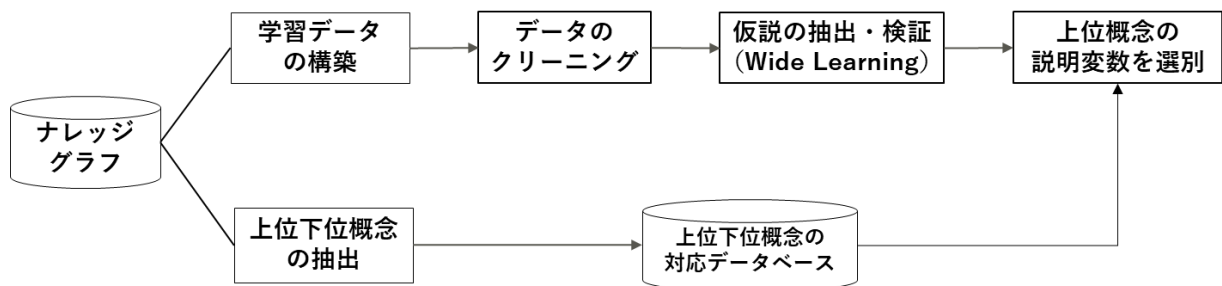


図 3. 提案手法の構成

繋がるエンティティは説明変数として選択出来るが、繋がっているエンティティのさらに上位の概念を参照するような事は行っていない。本研究では、任意のエンティティから直接繋がっているエンティティだけではなく、さらに上位の概念を考慮する事で、解釈性の高い説明変数の選択を行う。

3. 提案

3.1 提案手法の概要

本研究では、「ある上位概念を構成する下位概念すべてにおいて同一条件で成立する仮説は上位概念でも同一条件で成立する」という考えに基づき、上位概念の説明変数を選択するかどうかを決定する。各概念を要素の集合と言い換えた場合、各下位概念の要素の集合それぞれで共通して成立する仮説は、上位概念の包括的な要素の集合でも成立するという事を意味している。具体的に、任意に設定した目的変数に対し、ある上位概念に基づく説明変数 X_{sup} が i である時の要素の集合を $\{X_{sup} = i\}$ 、その下位概念に基づく説明変数 X_{sub} が j である時の要素の集合を $\{X_{sub} = j\}$ とした時、ある上位概念が n 個の下位概念で包括的に構成される状態は以下のように表せる。

$$\{X_{sup} = i\} = \{X_{sub} = J_1\} + \{X_{sub} = J_2\} + \dots + \{X_{sub} = J_n\} \quad (2)$$

この時、「 $X_{sub} = j_k$ ($k = 1, \dots, n$) かつ任意の条件 α の時に目的変数に影響する」という仮説が成立する場合、「 $X_{sup} = i$ かつ α の時に目的変数に影響する」も併せて成立すると考えるのが本提案のコンセプトである。

具体例として、図 4 を用いて先の都道府県の例で説明する。「四国地方」という上位概念は、「愛媛県」などの下位概念となる 4 都道府県との part-of 関係で要素の集合が包括的に構成されており、 X_{sup} を出身地方、 X_{sub} を出身県とした場合に、その集合関係は (2) 式を満たす。この時、下位概念の各都道府県で共通して成立する仮説を探索し、図 4 では「 $X_{sub} = j_k$ ($k = 1, \dots, n$) かつ入団チームが球団 X の時に目的変数 y に

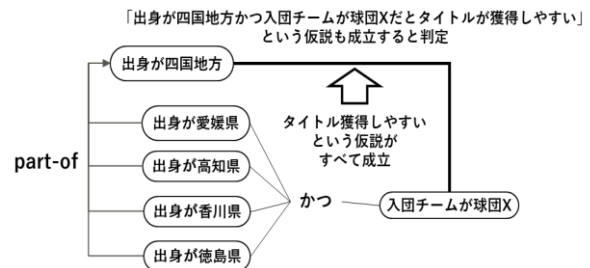


図 4. 上位概念における仮説成立条件の例

影響する」という仮説が成立している事を示している。このような時に、「出身地方が四国地方かつ入団チームが球団 X の時に目的変数に影響する」という仮説が成立すると判定するのが本提案手法の特長である。仮説が成立するという事はすなわち説明変数として使用する価値があるという事であり、四国地方という上位概念を用いた説明変数がモデルに加わる事になる。

本提案手法を実現するための実装の全体構成を図 3 に示す。全体の流れとして、まずナレッジグラフからモデルに使用する学習データを、エンティティやプロパティ情報に基づきルールベースで抽出し、構築する。その際に合わせて、学習データの構築に使用する各エンティティに繋がる上位概念のエンティティの関係性も抽出する。そして構築した学習データのクリーニングを行い、その後、学習データを基に、目的変数に影響する仮説の抽出・検証を行う。そして、その仮説リストおよび前述の上位概念の関係性に基づき、上位概念の説明変数の選別を行う。以下、図 3 の各処理ブロック毎に説明を行う。

3.2 学習データの構築

学習データを構築する方法として、想定する学習データのサンプルに対応する任意のエンティティから直接繋がっているエンティティの情報を基にデータのスクレイピングを行う。例えばプロ野球選手のデータの場合、各プロ野球選手のエンティティに直接繋がっている身長や体重などのエンティティの情

表 1. 学習データの構築ルール

No.	Operator defining fi	type	fi(e)
1	p.exists	boolean	$fi(e) = 1$ if (e, p, o) or $(o, p, e) \in T$, otherwise $fi(e) = 0$
2	p.value	int / double / category	$fi(e) = \{v \mid (e, p, v) \in T\}$
3	p.num	int	$fi(e) = \{v \mid (e, p, v) \in T\} $
4	p.hyper	category	$fi(e) = \{s \mid (s, p = 'part_of', p.value) \in T\}$

報を網羅的に抽出する事になる。抽出後は、エンティティ間のプロパティに基づいて、特徴量へと変換を行う。具体的な方法として、トリプルの集合を T 、学習データの各サンプルにあたるエンティティの集合を E 、各サンプルのエンティティ e ($e \in E$) において付与する特徴量を f_i とした場合に、文献[2]を参考にした表 1 の No.1~3 に示すルールに従い特徴量の抽出を行う。

表 1 における No.1 のルールは、各サンプルにあたるエンティティと任意のプロパティを含んだトリプルが存在しているかどうかを示し、No.2 は存在する場合の具体的な値、No.3 はその数を示す。このようにナレッジグラフから特徴量を抽出する際に、任意のプロパティで繋がった先のエンティティの値だけではなく、プロパティの関係性の有無や同一プロパティで繋がっているエンティティの数も特徴量として網羅的に抽出している。

3.3 上位下位概念の抽出

本提案における上位概念と下位概念の関係性は(2)式で表される集合関係を満たしている事を想定している。要素と集合関係のオントロジーの意味は instance-of 関係と part-of 関係に集約されると考えられており[6]、特に全体と部分のような集合演算に近い関係性は part-of 関係で表現される事が多い。しかし、「どのような時に part-of 関係が成立するか?」という part-of 関係の根本的な定義はまだ明らかになっておらず、様々な議論が行われている[6][7][8]。例えば文献[6]では part-of 関係を機能性や分離可能性などの複数の尺度から 6 種類の関係性に分類しており、文献[7]ではそれをさらに発展させ、集合関係に空間的関係性や時間的関係性などのオントロジー的な制約を加えたうえで 7 種類の定義化を行っている。そのため、本提案を実現するためには、part-of 関係を上記のような分類に照らし合わせた上で、(2)式を満たすような関係性に限定して抽出する必要がある。しかし、現時点ではまだ上記分類を行えるような高機能な part-of 関係が付与されているデータセットは数少く、用意するためには多大なコストを要する。そのため、本研究では表 1 の No.4 に示すルールに基づき part-of で繋がる関係を網羅的に抽出し、その後

は人手で(2)式を満たすような関係性のものを選別するという方式を取った。この点に関しては、将来ナレッジグラフを拡張する事により、自動で選別する方式を実現したいと考えている。

3.4 データのクリーニング

3.2 節のような方法で網羅的にデータを抽出した場合、意味の無い特徴量の抽出を大量に含まれている可能性が高い。この処理過程では、シンプルにすべてのサンプルにおいて値が同じになる特徴量、または値が全て違う特徴量は全て削除し、残った特徴量を説明変数の候補とする。

3.5 仮説の抽出・検証

本提案を実現するためには、上位概念を構成する下位概念すべてにおいて共通して成立する仮説を探索して検証する必要がある。各仮説を統計的に検証するためには、仮説毎に目的変数に対する影響度合いを計算する必要がある。例えば今回扱っている問題のような「プロ野球選手がタイトルを獲得したか否か」という目的変数からなる 2 値分類問題であれば、母集団の正例割合と各仮説の正例割合に基づく t 検定によって検証が成される。このような計算は、特に変数間の交互作用を考慮した場合、説明変数が多くなると変数間の組み合わせパターンが膨大となり、多大な計算時間が必要となる。本研究では、このような膨大な仮説空間の探索および計算を高速に行う技術として、富士通研究所の開発技術「Wide Learning」[9][10][11]を使用する。



図 5. Wide Learning による仮説の抽出・検証のイメージ

Wide Learning は分類問題を対象とし、説明変数のあらゆる組み合わせを網羅的に探索し、目的変数の判定に適した組み合わせを重要な仮説「ナレッジチャンク (knowledge chunk, KC)」として抽出する。Wide Learning ではこのような網羅的な仮説の抽出を、発見科学の知見に基づく独自技術を用いて、高速に実現している。例えば、今回扱っている問題の場合、3.4 節までの処理で構築されたプロ野球選手のデータを入力として、図 4 のようにタイトルを獲得するかどうかに関連する KC を網羅的に抽出する事が出来る。本提案では、この抽出された KC を基に、タスクに応じた任意の仮説検定をそれぞれの KC に対して行い、重要な仮説リストを作成する。ここで、重要な仮説リストに一切出現しない説明変数の候補は、モデルに使用する可能性は無いものと見なし、候補から削除し、残った候補を実際にモデルで使用する説明変数とする。

3.6 上位概念の説明変数を選別

前節で作成した仮説リストを基に、3.3 節において抽出した上位下位概念の対応関係に照らし合わせ、ある上位概念を構成する下位概念すべてにおいて同一条件で成立する仮説があるかどうかを探索する。そして、そのような仮説が存在する場合に、その上位概念を説明変数として新たに選択する。ここで、「下位概念すべてにおいて成立する」というルールがデータセットによっては厳しすぎる事も考えられる。実際に後述の今回実験で扱ったデータにおいて、人から見た限りでは上位概念で解釈した方が良い場合でも、一部の下位概念のデータが全く存在しないために、上記ルールを満たさないというケースが散見された。これは主に下位概念毎にデータのサンプル数が不均衡である場合や、そもそも全体のサンプル数が十分に足りていない場合に起きやすい。そのため、「下位概念すべてにおいて成立する」というルールは、データセットによってはデータが全く存在しない下位概念をある程度許容した上で運用される事も想定される。

4. 実験

4.1 実験の概要

本研究では冒頭で述べた通り、「プロ野球選手がタイトルを獲得するかどうかに影響する特徴量は何か?」という問題を用いて、提案手法の有効性を検証する。具体的に、本提案手法を用いて上位概念に基づく説明変数選択を行い、そのような説明変数を追加した場合に、モデルを設計する際の根拠となる仮説リストがどのように変化するかを分析した。

本実験で扱うプロ野球選手のデータとして、タイトル獲得経験があるかどうかを判定可能な既に引退済みの 1970 年から 1990 年にプロ野球球団に入団した選手を対象とし、学習データのソースとして、Wikipedia ベースのナレッジグラフである DBpedia および Wikidata から各選手の情報を抽出した。目的変数の条件となるタイトルに関しては、投手・野手の主要タイトル（最多勝、本塁打王など）に加えて、ベストナイン賞、ゴールデングラブ賞、シーズン MVP、新人王のいずれかを獲得した事があるかどうかによって定めた。

4.2 データの構築ならび仮説の検証

Wikipedia の「Category:日本の野球選手」の情報を基に、DBpedia および Wikidata から計 1184 選手の情報を抽出した（ただし、Wikipedia 上に情報が無い選手や、データ記述に誤りがある選手は漏れている可能性がある事を補足しておく。）このうち、タイトルを獲得した選手の正例割合は 0.16 である。3.2 節の処理により自動で取得した特徴量は計 163 種類で、そのうち 3.5 節までの処理で計 151 種類の特徴量が自動的に説明変数の候補から削除され、表 2 に示す残りの 12 種類の説明変数が、上位概念を考慮しない状態における説明変数として選択された。仮説の検証は母集団の正例割合と各仮説の正例割合に基づく有意水準 5% の t 検定によって行っている。なお、旧称が混在していた球団名の統一、および学歴情報の統一は DBpedia および Wikidata 上の情報だけでは困難であったため、手動で値の編集を行っている。また、説明変数のラベル名も、識別しやすいように手作業で設定している。

表 2. 上位概念を考慮しない段階における説明変数

説明変数	値の例
入団チーム	カーブ
入団年	1970
ドラフト順位	2
投げ	右
打ち	右
ポジション	投手
学歴	高卒
誕生年	1952
誕生月	10
出身都道府県	大阪
身長	178
体重	78

表 3. 上位概念を考慮しない段階における仮説リスト (カバレッジ上位 5 件)

No.	仮説	カバレッジ	正例割合
1	ドラフト順位 <2	0.37	0.31
2	ポジション_内野手 ∧ 出身都道府県_千葉	0.05	0.47
3	入団チーム_現オリックス ∧ 投げ_右 ∧ ポジション_外野手 ∧ 学歴_社会人	0.03	0.60
4	入団チーム_現オリックス ∧ 打ち_右 ∧ 出身都道府県_兵庫	0.02	0.80
5	学歴_社会人 ∧ 出身都道府県_鳥取	0.02	0.75

表 4. 上位概念を考慮後の仮説リスト (カバレッジ上位 5 件)

No.	仮説	カバレッジ	正例割合
1	ドラフト順位 <2	0.37	0.31
2	入団リーグ_パリーグ ∧ 投げ_右 ∧ ポジション_外野手 ∧ 学歴_社会人	0.06	0.38
3	ポジション_内野手 ∧ 出身都道府県_千葉	0.05	0.47
4	打ち_右 ∧ 出身地域_南関東 ∧ 体重 ≥ 93kg	0.03	1.00
5	入団リーグ_パリーグ ∧ 学歴_社会人 ∧ 体重 ≥ 93kg	0.03	1.00

4.3 上位概念の説明変数の選択

上位概念を考慮しない段階で前述の仮説検定によって棄却されなかった仮説リストのうち、全体の正例データに対するカバレッジ (網羅率) 順にソートした結果の上位 5 件を表 3 に示す。このような仮説リストを基に、3.1 節および 3.6 節の方式で上位概念の説明変数の選択を行ったところ、

- ・「出身都道府県 ⇒ 出身地方」
- ・「入団チーム ⇒ 入団リーグ」
- ・「(捕手, 内野手, 外野手) ⇒ 野手」

の 3 通りの上位概念の説明変数の選択が行われた。例えば、出身地方の選択に関しては、

- ・「打ち_右 ∧ 出身都道府県_東京 ∧ 体重 ≥ 93kg」
- ・「打ち_右 ∧ 出身都道府県_神奈川 ∧ 体重 ≥ 93kg」
- ・「打ち_右 ∧ 出身都道府県_埼玉 ∧ 体重 ≥ 93kg」

の 3 仮説が成立し、その結果上位概念に対して「打ち_右 ∧ 出身地域_南関東 ∧ 体重 ≥ 93kg」でも成立すると判定した。なお、「打ち_右 ∧ 出身都道府県_千葉 ∧ 体重 ≥ 93kg」のデータは存在しなかったが、今回は 3.6 節で述べた通り、データが全く存在しない下位概念を、ある程度の範囲までは許容するルールに基づいて選択を行った。具体的に、データが全く存在しない下位概念がその上位概念を構成する下位概念の総数の過半数を越さない場合は許容するというルール設定を行っている。このように選択された上位概念の説明変数を基に、再度抽出した仮説リストのうち、全体の正例データに対するカバレッジ順にソートした結果の上位 5 件を表 4 に示す。

表 3 の上位概念を考慮しない段階では、「出身県」や「入団チーム」の情報に基づいた仮説が上位に来

ているが、表 4 では No.2 や No.5 の「入団リーグ」や、No.4 の「出身地域」のような上位概念に基づく、よりカバレッジの高い仮説が上位に来ている事が読み取れる。特に、表 3 における No.3 の仮説は表 4 では No.2 の仮説に置き換わっており、より上位の概念で言い換え可能であった事が分かる。また前述の「打ち_右 ∧ 出身地域_南関東 ∧ 体重 ≥ 93kg」の例も、表 4 の No.5 に出現している。このようにカバレッジ順でソートする事で、下位概念毎の仮説だけでなく、それを包括する上位概念に基づく仮説も合わせて読み取れる事が分かる。以上の事から、本実験では提案手法を用いる事で、上位概念を考慮した説明変数を適切に選択し、上位概念に基づく解釈が可能な仮説リストが列挙出来る事を確認した。

最後に各仮説の中身を見てみる。表 3 および表 4 の No.1 「ドラフト順位 <2」は、ドラフト順位が 1 位である事を意味しており、プロ野球のファンであれば直感的に想像可能な説明変数である。一方、表 3 および表 4 の No.2~5 は、プロ野球のファンから見てもあまり直感的な仮説であるとは言えない。これらの仮説はあくまで統計的に棄却されなかった仮説群であり、真に目的変数に影響する仮説かどうか証明するためには、偶然性や疑似相関の可能性などを考慮した厳密な分析が必要となる。しかし、例えば表 3 の No.2、表 4 の No.3 の「ポジション_内野手 ∧ 出身県_千葉」という仮説は、「もしかしたら千葉県出身の名内野手である長嶋茂雄氏が影響しているのでは？」という新たな仮説を検討するための示唆となる。このように、プロ野球のファン、すなわち専門家でも気付くづらい仮説を発見出来るという点は、本研究のようなナレッジグラフから説明変数の候補を自動で選択する手法のメリットであると言える。

さらに本提案の上位概念に基づく解釈を実現する手法は、そのような専門家でも気づきづらい仮説を発見する可能性を高める事にも繋がる技術であると考えている。

5. むすび

本研究では、学習データのリソースとしてナレッジグラフを使用する場合を想定し、エンティティ間の上位概念の関係性を利用した、解釈性の高いモデルを実現するための説明変数を選択する手法を提案した。また、実際にプロ野球選手の実データを利用した実験により、上位概念を考慮した説明変数を必要に応じて選択し、上位概念に基づく解釈が可能な仮説リストが列挙出来る事を確認した。

今後は 3.3 節でも言及した、集合関係の対応付けが可能な上位概念を自動で選別する方式を実現したいと考えている。また、現状では任意のエンティティに繋がるエンティティとその上位概念しか説明変数の候補として見ていないが、例えばあるプロ野球選手の出身校の設立年など、さらに広い繋がりも説明変数の候補として効率的に選別する技術を開発したいと考えている。

参考文献

- [1] CrowdFlower : 2017 Data Scientist Report, https://visualfigure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport.pdf, (2017)
- [2] Michalis, M. Yannis, T.: How linked data can aid machine learning-based tasks, 21st International Conference on Theory and Practice of Digital Libraries, Vol. 10450, pp. 155-168, (2017)
- [3] Satoshi, H. Kohei, H.: Making Tree Ensembles Interpretable: A Bayesian Model Selection Approach, Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics, Vol. 84, pp. 77-85, (2018)
- [4] Efron, M.: Multiple regression analysis, Mathematical Methods for Digital Computers, pp. 191-203, (1960)
- [5] Dimitris, B. Romy, S.: Algorithm for cardinality-constrained quadratic optimization, Computational Optimization and Applications, Vol. 43, No. 1, pp. 1-22, (2009)
- [6] 溝口 理一郎, 池田 満, 來村 徳信, オントロジー工学基礎論 : 意味リンク, クラス, 関係, ロールのオントロジーの意味論 (「オントロジーの基礎と応用」), 人工知能学会誌, Vol. 14, No. 6, pp. 1019-1032, (1999)
- [7] Winston, M. Chaffin, R, and Herrmann, D.: A taxonomy of part-whole relations, Cognitive Science, Vol. 11, No. 4, pp. 417-444, (1987)
- [8] 古崎 晃司, 溝口 理一郎: オントロジー構築における Part-of 記述とその実践-ルール理論に基づく部分構造表現モデル, 人工知能学会論文誌, Vol. 34, No. 1, pp. 1-13, (2019)
- [9] Hiroaki, I. Takuya, T. Hirofumi, S. Keisuke, G. Kotaro, O. Hiroki, A: Efficient Constrained Pattern Mining Using Dynamic Item Ordering for Explainable Classification, <https://arxiv.org/abs/2004.08015>, (2020)
- [10] 大堀 耕太郎, 浅井 達哉, 岩下 洋哲, 後藤 啓介, 重住 淳一, 高木 拓也, 中尾 悠里, 穴井 宏和: 知識発見によって信頼をつなぐ Wide Learning 技術: FUJITSU, Vol. 70, No. 4, pp. 48-54, <https://www.fujitsu.com/jp/documents/about/resources/publications/magazine/backnumber/vol70-4/paper08.pdf>, (2019)
- [11] Fujitsu Laboratories Ltd.: Hello, Wide Learning!, <https://widelearning.labs.fujitsu.com/ja/>