

知識発見によって信頼をつなぐ Wide Learning技術

Wide Learning Technology to Provide Trust Through Knowledge Discovery

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

あらまし

近年、AI（人工知能）のブラックボックス問題が生じ、AIの公平性、説明責任および透明性（Fairness, Accountability and Transparency）の議論が活発化しており、企業や学術機関が様々な説明可能なAIを開発している。富士通研究所では、ブラックボックス問題への対応のみならず、知識発見を実現する新たな説明性を有するWide Learning技術を開発した。Wide Learningは、ディスカバリーサイエンス（発見科学）の主要技術である「列挙」により、人が理解できる形式で記述されたナレッジチャンクと呼ばれる知識を網羅的に数え上げる。この知識を用いて判断することで起こり得る可能性の見逃しを大幅に減らし、高精度の予測・分類を実現する。見逃しのない知識を継続的に発見していくことができれば、サービスシステムの信頼（Trust）を強化し、更に知識転用によりサービスシステム間で信頼をつなげることが期待できる。

本稿では、Wide Learningが生まれた歴史・経緯を説明した上で、Wide Learningが有する技術的特徴を説明する。また、社内におけるWide Learningを用いた知識発見の実証実験について報告し、今後の展望について述べる。

Abstract

With the recent emergence of the AI black box issue, the fairness, accountability, and transparency (FAT) of AI have been vigorously debated. Companies and academic institutions have developed various types of explainable AI. Fujitsu Laboratories has developed a new technology, Wide Learning, which is not only capable of dealing with the black box issue but which is also provided with explainability that realizes knowledge discovery. Wide Learning uses “enumeration,” which is a major technology of discovery science, to exhaustively enumerate pieces of knowledge called knowledge chunks described in a human-understandable format. By using this knowledge for judgment, the overlooking of potential candidates is significantly reduced and high-accuracy prediction and classification are achieved. If knowledge without omission can be discovered continuously, the trust of service systems can be strengthened and knowledge diversion in turn is expected to provide trust between service systems. This paper first outlines the history of how Wide Learning came into being and describes its technical characteristics, then reports on a demonstration of knowledge discovery using Wide Learning in Fujitsu and presents future prospects.

1. まえがき

近年AI（人工知能）の分野では、Deep Learningをはじめとする機械学習において、その判断のプロセスが説明できないというブラックボックス問題が生じている。2018年12月27日には、内閣府の「人間中心のAI社会原則検討会議」から、社会が考慮すべき七つの基本原則「人間中心のAI社会原則」の原案⁽¹⁾が公開された。その中では、セキュリティやプライバシーに加えて「公平性、説明責任及び透明性（FAT：Fairness, Accountability and Transparency）」の原則が記載されている。

また、2016年8月に米国国防高等研究計画局が説明可能なAI（Explainable AI：XAI）⁽²⁾への研究プログラムを発表し、それ以降、企業や学術機関が様々なXAIの開発を進めている。例えば、データの偏りを検知し公平性を担保するAI⁽³⁾や、分類の根拠を説明するAI⁽⁴⁾などがある。

このような状況の中、富士通研究所は2018年9月に新たな機械学習技術Wide Learningを発表した。Wide Learningと従来のXAIとの大きな違いは、Wide Learningが現時点のデータから読み取れる知識を漏れなく抽出することで、判断したい対象の見逃しを大幅に減らす点にある。これは、本特集のテーマである信頼の観点から考えると、サービスシステムの膨大なデータの中から新たな知識を発見・獲得できるという点で、信頼を生み出す役割であると考えられる。

また、データが追加された場合には、知識の変化が一目で分かり、継続的な知識発見を行うことができる。更に、得られた知識は、システム間の同型性を利用してほかの問題に転用することで、システムの複雑性を縮減するとともに、新たな知識発見を促進する可能性をも持ち得る。

以上のように、Wide Learningは知識発見を核として、一つのサービスシステムの中で信頼を強化し、そこで得た信頼をほかのサービスシステムへつなぐための技術となり得る。これは、複数のアプリケーションがつながるSociety 5.0⁽⁵⁾のシステム観とも重なる。

本稿では、Wide Learningを開発した歴史・経緯、

およびその技術的特徴を紹介する。また、Wide Learningの社内適用事例をとおして、サービスシステムにおける効果を説明する。

2. Wide Learningが生まれた歴史・経緯

AI（人工知能）、特に機械学習の進化は主として、脳の神経回路網を数学的にモデル化したニューラルネットワークの系列と、論理的な理解を志向する記号処理的AIの系列に大別される。近年、目覚ましい発展を遂げ、高い予測・分類精度を実現しているDeep Learningは、前者のニューラルネットワークと同じ系列である。一方、本稿で紹介するWide Learningは、記号処理的AIの最新技術である。

Wide Learningが生まれた背景には、1990年代に有川氏が提唱したディスカバリーサイエンス（発見科学、以下DS）⁽⁶⁾に関する研究の歴史がある。富士通研究所は、DSが提唱された当時から、データマイニングを含む知識発見に関する研究を推進してきた。また、数々の国家プロジェクトに参画し、技術を蓄積していった。^{(7), (8)}そして、2015年12月に有川氏を招聘して富士通研究所に「有川ディスカバリーサイエンスセンター」を設立し、社外の専門家とともにオープンイノベーションをとおしてDSの理論研究と実証研究に取り組んできた。

Wide Learningは、これらの研究成果の積み重ねによって重要性が明らかになり、アルゴリズムが劇的に進化した「列挙」をキーテクノロジーとする、新たな機械学習技術である。

3. Wide Learningの概要と技術的特徴

本章ではWide Learningの概要、および技術的特徴について述べる。

3.1 Wide Learningの概要

Wide Learningは、学習データの中から、データ項目の組み合わせで記述された人間にとって直接理解可能な知識（ナレッジチャンク）を高速に「列挙」し、その知識を用いて予測・分類を行う技術である（図-1）。Wide Learningでは、高速な列挙技術を用いてデータ項目のあらゆる組み合わせを網羅的に探

大量のモデルで可能性を網羅

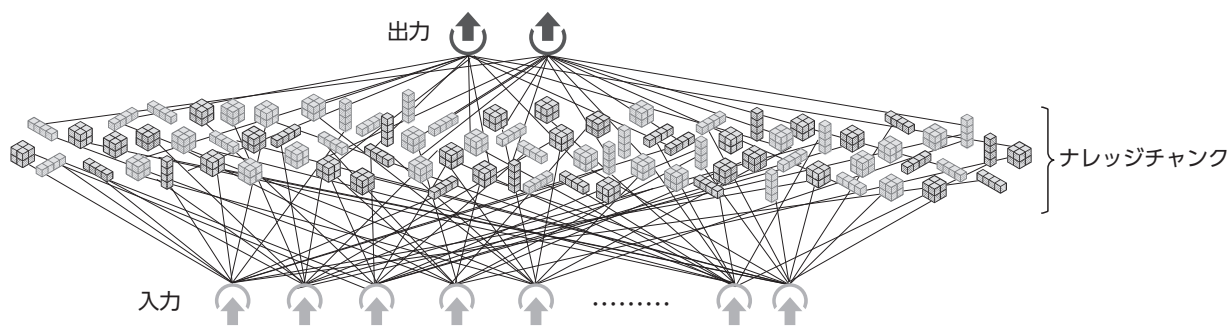



図-1 Wide Learningの概要

問題例

鶴と亀がいます。
頭は10個あります。
足は全部で28本です。
鶴と亀はそれぞれ何匹いますか？



正解

列挙の例		
鶴	亀	足
1	9	38
2	8	36
3	7	34
4	6	32
5	5	30
6	4	28
7	3	26
8	2	24
9	1	22

図-2 鶴亀算と列挙

索し、正例と負例における出現率が有意に異なるもの全てをナレッジチャンクとして抽出している。

Wide Learningの重要な要素技術は、可能性を網羅的に数え上げることで問題の全体像を理解しようとする「列挙」である。条件が複雑に絡み合い、数式化することが困難な問題であっても、問題を構成する全ての具体的事例を数え上げることができるなら、人間はその答えに対して信頼 (Trust) を持つことができる。例えば、図-2に示す「鶴亀算」を考えてみよう。この問題は、連立方程式を知っていれば簡単に解くことができる。しかし、そのような抽象的な概念を理解できない低学年の子供でも、可能性を全て並べてみれば、答えを確実に理解できるであろう。

Wide Learningは、この列挙による知識発見を繰

り返すことで、サービスシステムの特徴を深く理解できるようになり、システム的设计者や受益者に対する信頼を強化する。また、特定のサービスシステムの特徴の理解は、関連システムに対するデザイン指針を与える可能性があり、結果として、信頼の連鎖が期待できる (図-3)。

以降の節では、Wide Learningの技術的特徴である、知識の発見と転用について説明する。

3.2 知識発見のための高速な列挙技術

DSの中心的なテクノロジーの一つである列挙は、明らかに人間よりも計算機の方が得意な領域である。しかし、組み合わせ爆発という現象のため、計算機であっても単純な方法で解ける問題には限りがある。

今、男性/女性、運転免許のあり/なし、既婚/未婚などの顧客属性に関する条件を組み合わせること (例えば「男性かつ既婚」) によって、ある商品の購買層を特徴付けたいとしよう。性別に関しては、(男性に限定する) (女性に限定する) (限定しない) という3通りの条件がある。仮に同様な属性が10種類ある場合には、考えられる組み合わせは3の10乗、すなわち59,049通りとなる。この程度の規模であれば、計算機を用いて、それぞれが購買層をうまく表現しているかを漏れなく確認できるであろう。しかし、属性が20種類になると約35億通り、30種類では約200兆通りとなり、確認すべき組み合わせの数は急激に増えていく。すなわち、そのような単純な列挙では、すぐに組み合わせ爆発の問題に直面するため、効率的に列挙する必要がある。

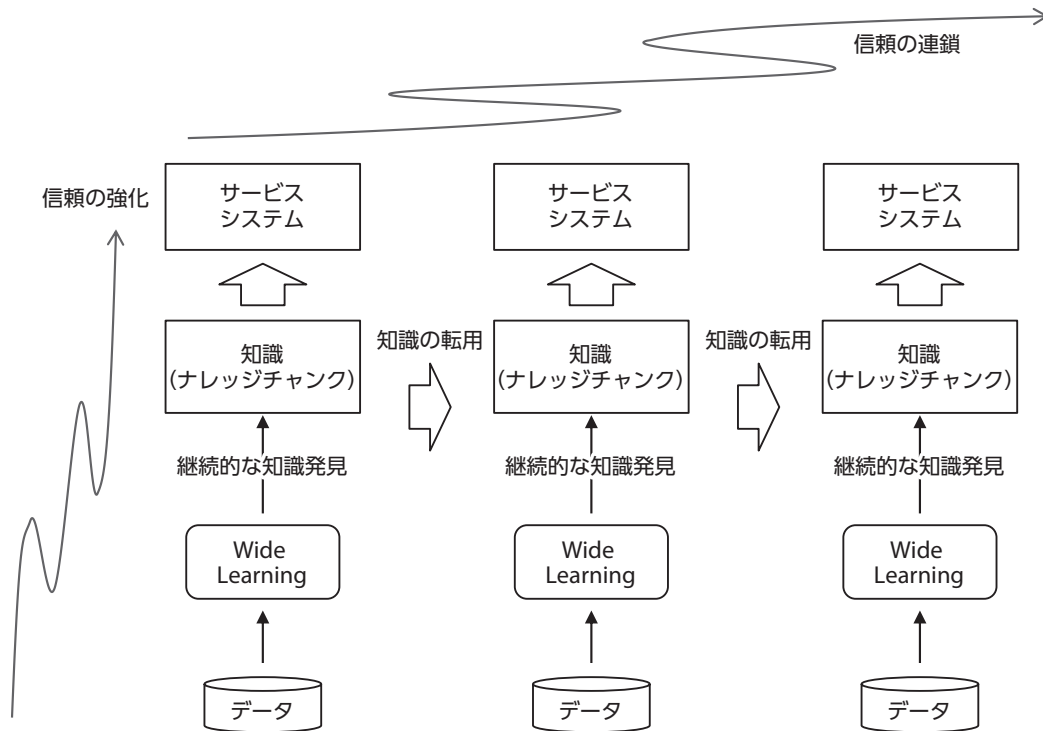


図-3 信頼をつなぐWide Learning

効率的な列挙の鍵となる技術に「枝刈り」と「メモ化」がある。将棋などのゲームのプログラムで例えると、明らかに不利な指し手を解析対象から外す技術が枝刈りである。一方、各局面からの解析結果を記憶しておき、同局面からの重複計算を避ける技術がメモ化である。メモ化と同様に重複計算を避ける技術として、同じ局面にたどり着く手順を束ねながらその先の解析を行う手法⁽⁷⁾、⁽⁸⁾も有効である。Wide Learningでは、それらの技術の積み重ねとそれらの効果を最大限に発揮させるための処理手順の最適化によって、現実規模の問題に対するナレッジチャンクの高速な列挙を実現している。

Wide Learningでは、この列挙したナレッジチャンクの重み・重要度を学習することで、分類・予測などを実現する。実際に、UC Irvine Machine Learning Repository⁽⁹⁾のHeart Disease（医療）とBank Marketing（金融）のデータを用いて実験を行った。その結果、Wide LearningはDeep Learningと比べて正解を当てる精度が約10%～20%向上し、正解を見逃す確率を約20%～50%低減することを確認している。⁽¹⁰⁾ 前述のとおり、Wide LearningはDeep Learningとは異なる系統の

技術であり、直接の比較は困難である。しかし、この結果は、知識を網羅的に列挙して起こり得る可能性の見逃しを減らすWide Learningのアプローチについて、有効性を示唆していると考えられる。

3.3 知識転用への期待

Deep Learningや既存の記号処理的なAIに基づく多くの機械学習手法は、「最適化」の考え方に基づいている。ほぼ同じ予測精度を達成するモデルが複数ある場合でも、最適化ではその中の一つしか採用されない。そのため、入力が少し変わるだけで、あるいは単に再実行しただけでも、予測モデルが大きく変わってしまう可能性がある。この現象は、学習結果を比較したり転用したりすることを困難にしている。

これに対して、列挙を基盤とするWide Learningでは、全ての有用なナレッジチャンクが漏れなく採用される。そのため、同じ入力からは必ず同じナレッジチャンク集合が得られるだけでなく、共通点の多い入力からは共通点の多いナレッジチャンク集合が得られるという特徴がある。

多くの現場では、AI適用に向けて環境整備が加

表-1 抽出したナレッジチャンクに対する専門家の考察

Wide Learningが重要だと判断したナレッジチャンク	専門家による考察
[役職：係長 / 主任クラス]	「部長」や「課長」などの“管理職”だけではなく，“現場リーダー”の影響度が大きい
[サービスへの興味：大] かつ [同じ企業のおユーザーのアクセス]	個人だけではなく、企業単位でサービスに対する行動を起こしている
[他サービスへの興味：大] かつ [12週間以内のアクセス]	商談が近づくと、対象サービス以外にも富士通の様々なサービスを確認する

速度的に進んでいる。一方で、AI適用のためのデータ蓄積が不十分な現場もいまだに多い。もしデータ量が不十分なままAIを適用した場合、性能が思うように上がらなかったり、未知のデータに対して想定外の挙動を示すなど信頼性に乏しい結果になってしまう。しかし、知識を厳密に抽出するWide Learning の特徴を利用することで、データ量が十分な類似システムから発見した普遍的な知識を、データ量が不十分なモデルに転用することが期待できる。

4. Wide Learningの活用事例

本章では、マーケティングの事例を用いて、Wide Learningを用いた知識発見の事例を紹介する。また、そこから得られたWide Learningによる知識転用の可能性について述べる。

本事例では、富士通のグローバルマーケティング部門（以下、専門家）が保有する顧客データにWide Learningを適用し、商談Xの見込み顧客の特徴（ナレッジチャンク）抽出と知識発見を試みた。顧客データは業種や役職などの属性情報のほかに、セミナー参加やウェブサイトなどの行動ログを含む。抽出するナレッジチャンクは、説明性の観点から構成する項目数を2以下に限定した。300の特徴量からなる顧客データにWide Learningを適用し、項目数2以下の約18万個のデータ項目の組み合わせの中から、重要度の重み付きナレッジチャンク1,300個を抽出した。更に、この重み付きナレッジチャンクを専門家に提示し、その正当性や背景を考察した。Wide Learningが重要であると判断したナレッジチャンク、およびその考察の一例を表-1に示す。

専門家による考察では、「経験的にも正しい」と

いった納得性に加え、「今まで気付かなかったが、あり得る特徴だ」といったコメントが得られた。Wide Learningの一連の処理はデータに基づいており、人間の思い込みや経験が入らない。そのため、Wide Learningが獲得したナレッジチャンクには、人間が思いもしなかった知識が含まれている可能性が高く、暗黙知の顕在化を促進させ、知識発見につながると考えられる。

また、表-1のナレッジチャンクは、ほかの商談でもあり得る特徴から構成されており、商談Xに限らず、ほかの商談においても適用可能な普遍的な知識の一例であると現場からのコメントが得られた。すなわち、ほかの商談Yにおける予測において、十分なデータセットがない場合には、商談Xで得られたナレッジチャンクを転用することで、いち早く商談の確度を高める手助けとなると考えている。

5. むすび

本稿では、ナレッジチャンクの列挙により説明可能なAIとして高精度な分類を実現するだけでなく、知識発見と知識転用を可能とするWide Learning技術について述べた。

現在は、デジタルマーケティング分野を中心に実証実験を行っているが、今後はものづくり分野や医療分野など様々な領域に適用範囲を広げる予定である。また、知識の転用を加速するための技術開発を加速し、Wide Learningとナレッジチャンクを介したサービスシステム間の連携を実現していく予定である。

本稿に掲載されている会社名・製品名は、各社所有の商標もしくは登録商標を含みます。

参考文献

- (1) 内閣府：人間中心のAI社会原則検討会議。
<https://www8.cao.go.jp/cstp/tyousakai/humanai/index.html>
- (2) D. Gunning：Explainable Artificial Intelligence (XAI). DARPA.
<https://www.darpa.mil/program/explainable-artificial-intelligence>
- (3) A. Mojsilovic：Trust and Transparency for AI on the IBM Cloud. IBM.
<https://www.ibm.com/blogs/research/2018/09/trust-transparency/>
- (4) E. Webber：Similarity Cracks the Code Of Explainable AI. simMachines.
<https://simmachines.com/similarity-cracks-code-explainable-ai/>
- (5) 内閣府：Society 5.0.
https://www8.cao.go.jp/cstp/society5_0/index.html
- (6) 有川 節夫：ディスカバリーサイエンス再考. FUJITSU, Vol.68, No.5, p.3-7 (2017).
<https://www.fujitsu.com/jp/documents/about/resources/publications/magazine/backnumber/vol68-5/paper01.pdf>
- (7) H. Iwashita et al.：Efficient Computation of the Number of Paths in a Grid Graph with Minimal Perfect Hash Functions. Technical Report, Graduate School of Information Science and Technology, Hokkaido University, 2013.
<http://www-alg.ist.hokudai.ac.jp/tra.html>
- (8) ERATO湊離散構造処理系プロジェクト：超高速グラフ列挙アルゴリズム. 森北出版, 2015.
- (9) Murphy, P. & Aha, D.：UCI repository of machine learning databases. Department of Information and Computer Science, University of California, Irvine, California, 1992.
- (10) 富士通研究所：正解が少ないデータでも高精度に学習するAIの新技术「Wide Learning」を開発。
<https://pr.fujitsu.com/jp/news/2018/09/19-2.html>

著者紹介



大堀 耕太郎 (おおほり こうたろう)
 (株) 富士通研究所
 人工知能研究所
 人工知能・数理技術による社会システムデザインの研究に従事。



浅井 達哉 (あさい たつや)
 (株) 富士通研究所
 人工知能研究所
 発見科学に基づく人工知能の研究に従事。



岩下 洋哲 (いわした ひろあき)
 (株) 富士通研究所
 人工知能研究所
 離散数理に基づく知識発見技術の研究・開発に従事。



後藤 啓介 (ごとう けいすけ)
 (株) 富士通研究所
 人工知能研究所
 文字列処理と人工知能の研究・開発に従事。



重住 淳一 (しげずみ じゅんいち)
 (株) 富士通研究所
 人工知能研究所
 知識処理, 情報検索, 推論などに関する人工知能の研究に従事。



高木 拓也 (たかぎ たくや)
 (株) 富士通研究所
 人工知能研究所
 文字列処理とデータマイニング, 人工知能の研究・開発に従事。



中尾 悠里 (なかお ゆうり)

(株) 富士通研究所
人工知能研究所
推薦システム、公平性配慮型機械学習、
アクセシビリティのユーザー観察の研究に従事。



穴井 宏和 (あない ひろかず)

(株) 富士通研究所
人工知能研究所
数式処理、最適化、人工知能の理論と
応用の研究・開発に従事。