

AI 特許紹介(24)

AI 特許を学ぶ！究める！

～ワイドディープニューラルネットワーク～

2021 年 1 月 8 日

河野特許事務所

所長 弁理士 河野英仁

「AI 特許紹介」シリーズは、注目すべき AI 特許のポイントを紹介します。熾烈な競争となっている第 4 次産業革命下では AI 技術がキーとなり、この AI 技術・ソリューションを特許として適切に権利化しておくことが重要であることは言うまでもありません。

AI 技術は Google, Microsoft, Amazon を始めとした IT プラットフォーマ、研究機関及び大学から毎週のように新たな手法が提案されており、また AI 技術を活用した新たなソリューションも次々とリリースされています。

本稿では米国先進 IT 企業を中心に、これらの企業から出願された AI 特許に記載された AI テクノロジー・ソリューションのポイントをわかりやすく解説致します。

1.概要

特許権者 Google

出願日 2016 年 12 月 29 日

登録日 2020 年 9 月 1 日

登録番号 US10762422

発明の名称 ワイドディープ機械学習モデル

422 特許は、ディープラーニングモデルにワイドラーニングモデルを組み合わせたワイドディープニューラルネットワークに関する。

2.特許内容の説明

図 1 は、ディープラーニングモデル 104 と、ワイドラーニングモデル 106 と、組合せ層 134 とを含むワイドディープラーニングモデル 102 を示すブロック図である。

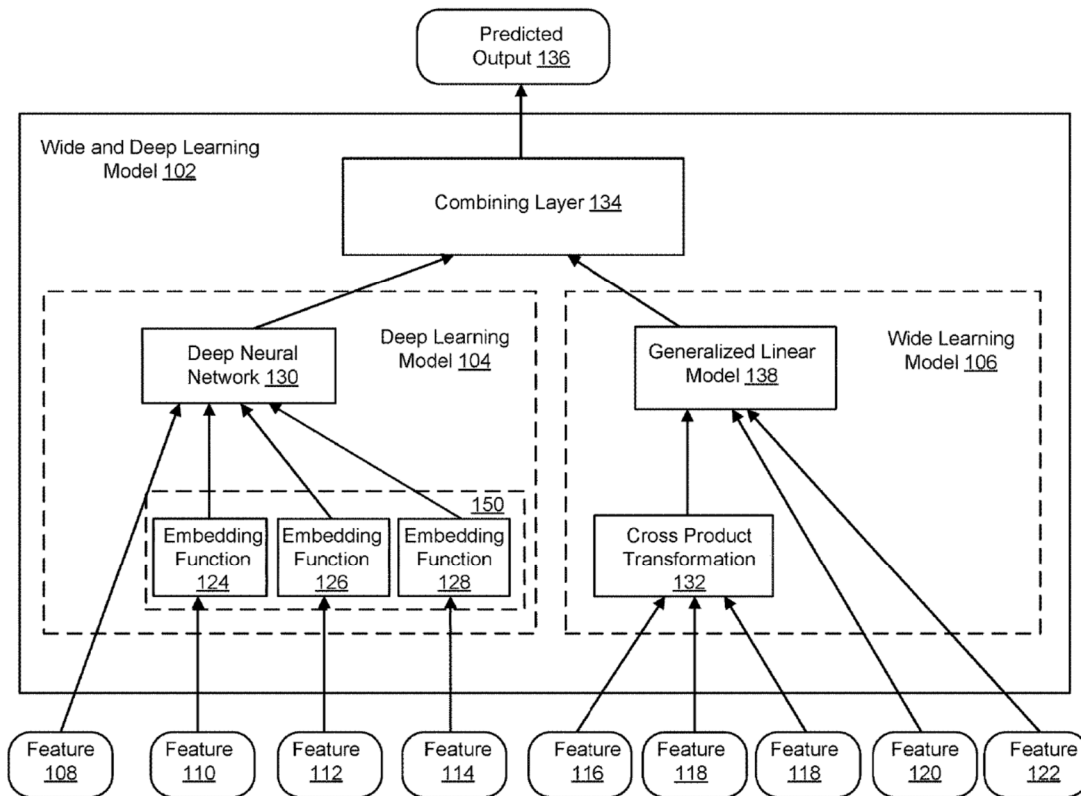


FIG. 1

ワイドディープマシンラーニングモデル 102 は、複数の特徴、たとえば特徴 108～122 を含むモデル入力を受け取り、モデル入力に対する予測出力 136 を生成する。

例えば、モデル入力は、単語のシーケンスであり、特徴 108～122 は、シーケンス内の単語を表すトークンと、シーケンス内の単語を特徴付ける他の特徴とを含む。この例では、予測出力 136 は、特定の単語がシーケンス内の次の単語である尤度、またはシーケンス内の特定の単語に関する品詞もしくは語彙に関する予測である。

その他、特徴 108～122 は、コンテンツ提示設定の特徴を含むことができ、予測出力 136 は、コンテンツ提示設定に関する予測出力とすることができる。

ディープマシンラーニングモデル 104 は、埋め込み層 150 とディープニューラルネットワーク 130 とを含むディープモデルである。埋め込み層 150 は、埋込み関数のセット、たとえば埋込み関数 124～128 を含む。ディープニューラルネットワーク 130 は動作の複数の層を含み、層のうちの少なくとも 1 つは出力を生成するために受信した入力に非線形変換を適用する。

ディープマシンラーニングモデル 104 は、ディープモデル中間予測出力を生成するために、ワイドアンドディープラーニングモデル 102 のモデル入力に含まれる特徴の第 1 のセットを処理する。

埋め込み層 150 は、埋込み関数 124~128 を特徴 110~114 に適用する。埋め込み層によって処理される特徴は、ユーザ特徴(たとえば、国、言語及び人口統計)、コンテキスト特徴(たとえば、デバイス、その日の時刻、曜日)、及び印象特徴(たとえば、アプリの年齢、アプリの履歴統計)などの分類特徴である。

埋込み関数 124~128 の各々は、特徴 110~114 の各々をそれぞれの数値埋め込み、たとえば特徴の浮動小数点数ベクトル表現にマッピングする変換を特徴 110~114 の各々に適用する。

ディープニューラルネットワーク 130 は、入力として、埋め込み層からの数値埋め込みと、オプションで他の入力特徴(たとえば、特徴 108)とを受信する。ディープニューラルネットワーク 130 は、ディープモデル中間予測出力を生成するために、受信した入力に非線形変換をそれぞれ適用する、複数たとえば 3 つ以上の動作層を含む。

ディープニューラルネットワーク 130 は、入力の代替表現、すなわちディープモデル中間予測出力を生成するために、数値埋め込み及び他の入力特徴に複数の非線形変換を適用する。

ワイドマシンラーニングモデル 106 は、ワイドアンドディープラーニングモデル 102 のモデル入力内に含まれる特徴の第 2 のセット(たとえば、特徴 116~122)を処理し、ワイドモデル中間予測出力を生成するように構成されるワイドアンドシャロー(wide and shallow)モデル、たとえば一般化線形モデル 138 である。

ワイドマシンラーニングモデル 106 は、ワイドモデル中間出力を生成するために、たとえばクロス積特徴変換 132 を使用して、特徴の第 2 のセット内の元の入力特徴(たとえば、特徴 120 及び 122)と、他の特徴(たとえば、特徴 116~118)から生成された変換された特徴との両方を処理する。

二値特徴について、クロス積変換は、構成特徴がすべて 1 の場合は 1 であり、そうでない場合は 0 である。この変換は、二値化分類特徴間の相互作用を取り込み、一般化線形モデルに非線形性を加えることができる。

組合せマシンラーニングモデル 102 は、予測出力 136 を生成するために、ディープマシンラーニングモデル 104 によって生成されたディープモデル中間予測出力と、ワイドマシンラーニングモデル 106 によって生成されたワイドモデル中間予測出力とを処理するように構成される組合せ層 134 を含む。予測出力が単一のスコアである実装形態では、組合せ層はロジスティック回帰層である。予測出力が複数のスコアを含む実装形態では、組合せ層はソフトマックス層である。

3.クレーム

422 特許のクレーム 1 は以下の通りである。

1. 1つまたは複数のコンピュータによって実行されると、1つまたは複数のコンピュータに、複数の入力特徴を含む機械学習入力処理するための結合された機械学習モデルを実装させる命令を格納する1つまたは複数のコンピュータ及び1つまたは複数の記憶装置を含むシステムは、機械学習入力の予測出力を生成し、複数の入力特徴はコンテンツ推奨設定の特徴を含み、予測出力は1つまたは複数のコンテンツアイテムのそれぞれのそれぞれのスコアを含み、ここで、それぞれのスコアは、対応するコンテンツアイテムがコンテンツ推奨設定で提示された場合に特定の目的が満たされるそれぞれの可能性を表し、複合機械学習モデルは以下を含む：

ディープモデル中間予測出力を生成するためのコンテンツ推奨設定の特徴を含む、複数の入力特徴の第1のサブセットを処理するように構成されたディープ機械学習モデルと、

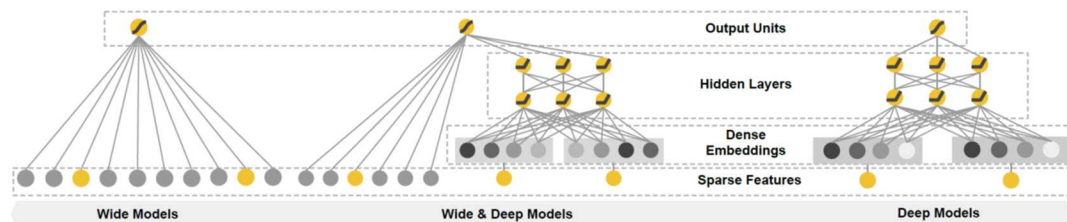
(i) 変換特徴を生成するためのコンテンツ推奨設定の特徴を含む複数の入力特徴の第2のサブセット内の入力特徴の少なくともいくつかに外積変換(cross-product transformation)を適用し、(ii) 一般化線形モデルを使用して、変換特徴と第2のサブセット内の入力特徴を処理して、ワイドモデルの中間予測出力を生成する、よう構成されたワイド機械学習モデルと、

1つ以上のコンテンツアイテム内の各コンテンツアイテムのそれぞれのスコアを含む予測出力を生成すべく、ディープ機械学習モデルによって生成されたディープモデル中間予測出力と、ワイド機械学習モデルによって生成されたワイドモデル中間予測出力を処理するよう構成された結合層と、

ディープモデルの中間予測出力とワイドモデルの中間予測出力とを生成すべく、ディープ機械学習モデルとワイド機械学習モデルは、トレーニングデータで共同でトレーニングされる。

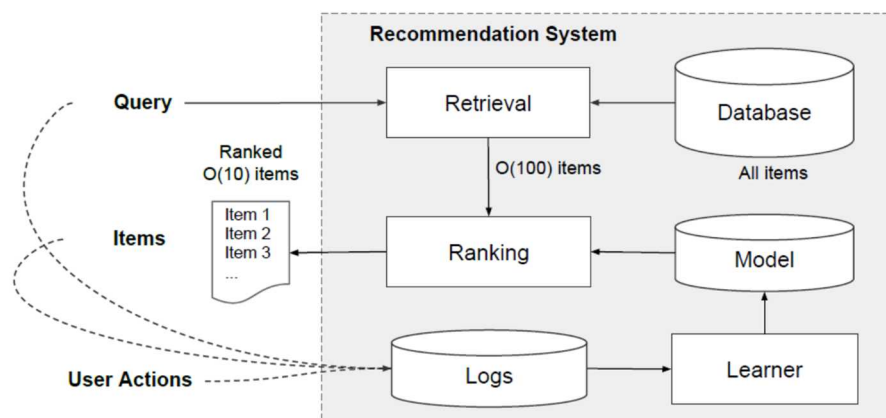
4. ワイドディープニューラルネットワークに関する論文

本特許に関連する論文¹が Heng-Tze Cheng 氏らにより発表されている。



上記図はワイドモデル、ワイドディープモデル、及びディープモデルをそれぞれ示している。

本論文では、10 億人以上のアクティブユーザと 100 万以上のアプリを GooglePlay でシステムを作成し、相互評価している。



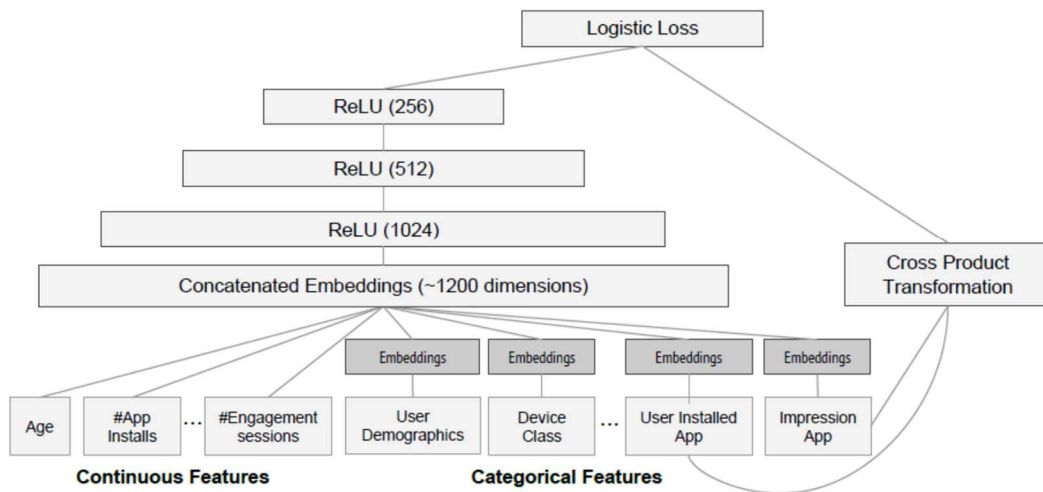
上記図は、アプリレコメンダーシステムの概要である。ユーザがアプリストアにアクセスすると、さまざまなユーザ機能とコンテキスト機能を含むクエリが生成される。レコメンダーシステムは、ユーザがクリックや購入などの特定のアクションを実行できるアプリのリスト（インプレッション）を返す。これらのユーザアクションは、クエリとインプレッションとともに、学習者のトレーニングデータとしてログに記録される。

データベースには 100 万を超えるアプリがあるため、サービングレイテンシ要件内ですべてのクエリに対してすべてのアプリを徹底的にスコアリングするのは困難である。

¹ Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishikesh Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, Rohan Anil, Zakaria Haque, Lichan Hong, Vihan Jain, Xiaobing Liu, Hemal Shah “Wide & Deep Learning for Recommender Systems” arXiv:1606.07792v1 [cs.LG] 24 Jun 2016

クエリを受信したときの最初のステップは検索であり、検索システムは、さまざまなシグナル（通常は機械学習モデルと人間が定義したルールの組み合わせ）を使用して、クエリに最適なアイテムの短いリストを返す。

候補プールを減らした後、ランク付けシステムはすべてのアイテムをスコアでランク付けする。スコアは通常 $P(y|x)$ 、ユーザ特徴（国、言語、人口統計など）、コンテキスト特徴（デバイス、時間帯、曜日など）、及びインプレッション特徴（たとえば、アプリの経過時間、アプリの履歴統計）を含む特徴 x が与えられた場合のユーザアクションラベル y の確率である。



上記図は、実験で使用したモデル構造である。トレーニング中に、入力レイヤーはトレーニングデータと語彙を取り込み、ラベルとともにスパース及びデンスフィーチャを生成する。ワイドコンポーネントは、ユーザがインストールしたアプリとインプレッションアプリの製品間の変換で構成される。モデルのディープ部では、32次元の埋め込みベクトルがカテゴリカル特徴ごとに学習される。

すべての埋め込みをデンス特徴と連結して、約 1200 次元のデンスベクトルを作成する。次に、連結されたベクトルは 3 つの ReLU 層に送られ、最後にロジスティック出力ユニットに送られる。Wide&Deep モデルは、5,000 億を超える例でトレーニングされる。

**Table 1: Offline & online metrics of different models.
Online Acquisition Gain is relative to the control.**

Model	Offline AUC	Online Acquisition Gain
Wide (control)	0.726	0%
Deep	0.722	+2.9%
Wide & Deep	0.728	+3.9%

上記表に示すように、ワイドディープモデルでは、アプリストアのメインランディングページでのアプリ取得率が、コントロールグループと比較して+3.9%向上している。

以上

著者紹介

河野英仁

河野特許事務所、所長弁理士。立命館大学情報システム学博士前期課程修了、米国フランクリンピアースローセンター知的財産権法修士修了、中国清華大学法学院知的財産夏季セミナー修了、MIT(マサチューセッツ工科大学)コンピュータ科学・AI研究所 AI コース修了。

[AI 特許コンサルティング](#)、[医療 AI 特許コンサルティング](#)の他、米国・中国特許の権利化・侵害訴訟を専門としている。著書に「世界のソフトウェア特許(共著)」、「FinTech 特許入門」、「[AI/IoT 特許入門 2.0](#)」、「[ブロックチェーン 3.0\(共著\)](#)」がある。