

AI 特許紹介(8)

～蒸留モデルのトレーニング方法特許～

2019年11月11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 2015年1月4日

登録日 2019年5月14日

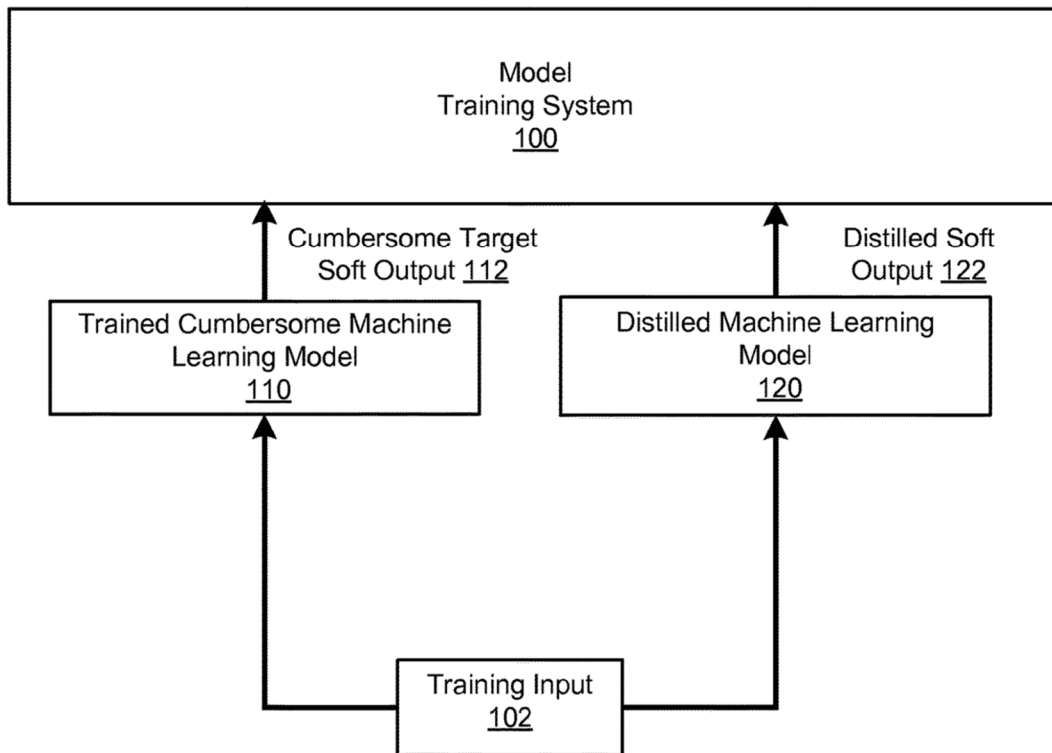
登録番号 US10289962

発明の名称 蒸留機械学習モデルのトレーニング

962 特許は、面倒なニューラルネットワーク(Cumbersome Neural Network)よりもパラメータ数の少ない蒸留ニューラルネットワーク(Distilled Neural Network)を、面倒なニューラルネットワークのソフトスコアを用いてトレーニングする方法である。

2.特許内容の説明

モデルトレーニングシステム 100 は、下記図に示すように面倒な機械学習モデル 110 と、当該面倒な機械学習モデル 110 よりもパラメータ数の少ない蒸留機械学習モデル 120 とにより構成される。



面倒な機械学習モデル 110 は、トレーニング入力 102 を受け取り、各クラスのスコアを生成する。蒸留機械モデル 120 も同じトレーニング入力 102 を受け取り、各クラスのスコアを生成する。

各トレーニング入力 102 を蒸留機械学習モデル 120 に入力した場合に、面倒な機械学習モデル 110 のターゲットソフト出力に合致するソフト出力を生成するよう蒸留機械学習モデル 120 をトレーニングする。ここで蒸留ソフト出力は以下を満たす。

$$q_i^d = \frac{\exp\left(\frac{z_i^d}{T^d}\right)}{\sum_j \exp\left(\frac{z_j^d}{T^d}\right)}$$

q_i^d はクラス i の蒸留ソフトスコア、 z_i^d は蒸留ニューラルネットワークの最後の層により受信された蒸留ニューラルネットワークの前の層の出力の重み付けされた組み合わせ、 j は 1 から合計複数のクラスにおけるクラスの数、 T^d は 1 より大きいそれぞれの温度定数である。

また、蒸留ニューラルネットワークのトレーニング中における温度定数 T^d は、蒸留

ニューラルネットワークがトレーニングされた後にクラスの蒸留スコアを生成するために使用される値よりも高い値に設定される。

3.クレーム

962 特許のクレーム 1 は以下の通りである。

1.1 又は複数のコンピュータにより実行される方法において、

面倒なモデルパラメータのトレーニング値を決定するために、複数のトレーニング例のセットにより、複数の面倒なモデルパラメータを有する面倒なニューラルネットワークをトレーニングし、各トレーニング例は、トレーニング入力および対応するターゲット出力を含み、面倒なニューラルネットワークは入力を受け取り、複数のクラスのそれぞれについてそれぞれのスコアを生成するよう構成されており、

複数のトレーニング例のセットの各トレーニング例の各トレーニング入力に対して、トレーニング入力の対応するターゲット出力に一致する出力を生成するために、面倒なニューラルネットワークをトレーニングし、

面倒なモデルパラメータのトレーニング値を決定するために、面倒なニューラルネットワークを訓練した後、蒸留モデルパラメータのトレーニング値を決定するために、複数の訓練入力のセットについて複数の蒸留モデルパラメータを有する蒸留ニューラルネットワークを訓練し、蒸留ニューラルネットワークは面倒なニューラルネットワークよりもパラメータが少なく、蒸留ニューラルネットワークは入力を受け取り、複数のクラスのスコアを生成するよう構成されており、

トレーニング入力の面倒な目標ソフト出力を生成するために、面倒なモデルパラメータのトレーニング値に従って面倒なニューラルネットワークを使用して、複数のトレーニング入力のセットの各トレーニング入力を処理し、

複数のトレーニング入力のセットの各トレーニング入力について、面倒なニューラルネットワークによって生成されたトレーニング入力の面倒な目標ソフト出力は、面倒な最後の層によって生成された複数のクラスのそれぞれの面倒な目標ソフトスコアを含み、

面倒な各目標ソフトスコアは、以下の条件式を満たし、

$$q_i^c = \frac{\exp\left(\frac{z_i^c}{T^c}\right)}{\sum_j \exp\left(\frac{z_j^c}{T^c}\right)}$$

ここで、 q_i^c はクラス i の面倒なターゲットソフトスコア、 z_i^c は面倒なニューラルネットワークの最後の層によって受信された面倒なニューラルネットワークの前の層の出力の重み付けされた組み合わせ、 j は 1 から a 複数のクラスにおけるクラスの総数、

および T^c は、1 より大きいそれぞれの温度定数であり、

複数のトレーニング入力のセットの各トレーニング入力に対して、該トレーニング入力を処理し、同じトレーニング入力を処理することにより、面倒なニューラルネットワークによって生成される面倒なターゲットソフト出力に一致する蒸留ソフト出力を生成すべく、蒸留ニューラルネットワークをトレーニングし、

複数のトレーニング入力のセットの各トレーニング入力について、蒸留ニューラルネットワークによって生成されたトレーニング入力の蒸留ソフト出力は、蒸留ニューラルネットワークの最後の層によって生成された複数のクラスのそれぞれの蒸留ソフトスコアを含み、各蒸留ソフトスコアは以下を満たし、

$$q_i^d = \frac{\exp\left(\frac{z_i^d}{T^d}\right)}{\sum_j \exp\left(\frac{z_j^d}{T^d}\right)}$$

ここで、 q_i^d はクラス i の蒸留ソフトスコア、 z_i^d は蒸留ニューラルネットワークの最後の層により受信された蒸留ニューラルネットワークの前の層の出力の重み付けされた組み合わせ、 j は 1 から合計複数のクラスにおけるクラスの数、 T^d は 1 より大きいそれぞれの温度定数であり、

ここで、蒸留ニューラルネットワークのトレーニング中、蒸留ニューラルネットワークに対応する温度定数 T^d は、蒸留ニューラルネットワークがトレーニングされた後にクラスの蒸留スコアを生成するために使用される値よりも高い値に設定される。

4. 蒸留モデルに関する論文

本特許に関連する論文「ニューラルネットワークにおける知識の蒸留」が Google の Hinton らにより発表されている¹。

論文では蒸留モデルの性能について検討すべく、Automatic Speech Recognition (ASR) において使用されているモデルの効果について検討を加えている。

面倒なニューラルネットワークにおけるアーキテクチャとして、2560 の Relu ユニットの有する 8 層の隠れ層と、14,000 のラベルを有する最終のソフトマックス層を用いた。全パラメータ数は 85M である。この比較的古いモデル(Baseline)がアンドロイド

¹ Geoffrey Hinton “Distilling the Knowledge in a Neural Network” 2015 年 3 月 9 日

における音声認識に使用された。この DNN 音声モデルをトレーニングするために、700M のトレーニング例を生成する 2000 時間の英語の発音データを使用した。

その結果、下記表 1 に示すように **Baseline** では、フレーム正確性が 58.9%、Word Error Rate (WER) が 10.9%との結果になった。

| System | Test Frame Accuracy | WER |
|------------------------|---------------------|-------|
| Baseline | 58.9% | 10.9% |
| 10xEnsemble | 61.1% | 10.7% |
| Distilled Single model | 60.8% | 10.7% |

表 1

続いて、10 の蒸留モデルを用意し、温度定数を[1, 2, 5, 10]と変化させ、ハードターゲットのクロスエントロピーにおいて 0.5 の重みを使用した。そのうち温度定数 2 の蒸留モデルの結果を表 1 に示す。フレーム正確性が 60.8%、WER は 10.7%となっている。

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](#)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](#)」がある。

以上