

Deep Learner

バージョン 1.3 新機能のご案内

株式会社 NTT データ数理システム

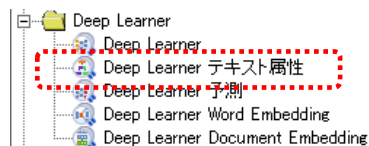
2020.08

1. はじめに

Deep Learner をご利用いただき誠にありがとうございます。本文書では、**Deep Learner** バージョン **1.3** で搭載された新機能および修正点を解説します。

2. テキスト属性 アイコンの新設

テキストデータと属性データの両方を同時に利用して予測モデルを構築することが可能な「テキスト属性」アイコンを新設いたしました。



既存機能である「Deep Learner」アイコンでは、設定項目”データ”を「テキスト」、「用途」を「予測」とすることでテキストデータを利用した予測モデル構築が可能でしたが、本機能を用いると、テキストに紐づく属性情報が存在している場合、その情報もモデルの学習に組み入れることができます。



例えば、不具合報告の文書が蓄積されているが、報告の本文に対して製品種別や部門などの様々な属性情報が付随している、また特許文書の要約や請求項といったテキスト部分に対して分類コードなどの属性情報が存在している、というようなケースにおいて、テキスト部分と属性部分とを同時に用いて分類モデルを作成することで、互いの情報を手掛かりにしてより望ましいモデルが構築できる可能性があります。

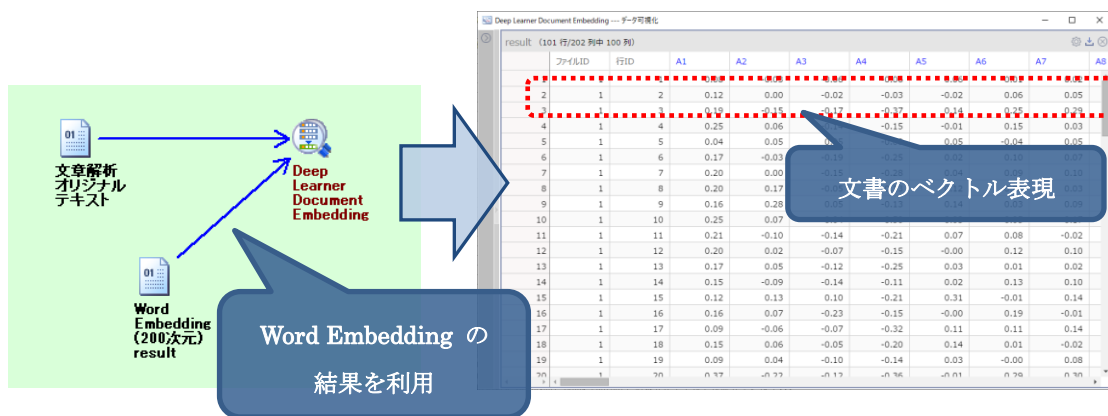
テキストデータは、テキストマイニングツール **Text Mining Studio** で分かち書き処理を行って単語への分割を事前に行っておく想定です。

3. Document Embedding アイコンの新設

文書のベクトル表現を作成する「Document Embedding」アイコンを新設いたしました。

本アイコンでは、次の2通りの方法で文書のベクトル表現を作成することが可能です。

- Word Embedding アイコンで算出した単語の分散表現を利用して、Simple Word Embedding-based Model (SWEM) を用いて文書のベクトル表現を作成する。
- テキストデータ（分かち書き結果）から、頻度ベースの手法（Bag of Words, tf-idf）を利用して、文書のベクトル表現を作成する。



このベクトルで表現された文書の情報を用いて、文書間の類似度計算やクラスタリングを行う、また予測モデルの変数にするなど、**Visual Mining Studio** で提供されている様々なマイニングの機能に適用させることができます。

4. その他の修正点

- 「Deep Learner」アイコンで、データの自動的に正規化して扱うことが可能に

なりました。正規化の方法は、列毎の最小値を 0, 最大値を 1 とする「最大-最小」と平均を 0, 分散を 1 とする「平均-分散」から選択することができます。

正規化
 最大-最小
 平均-分散


本機能を有効にした場合、データの全ての数値列に対して影響を及ぼします。

正規化を行った方が、精度向上が見込める場合も多いため、本バージョンにて「Deep Learner」アイコンのデフォルト動作は「正規化を行う」になりますのでご注意ください。(旧バージョンのプロジェクトを読み込んだ場合は、結果の互換性を確保するため「正規化を行わない」としてパラメータが表示されます。)

- 「Word Embedding」アイコンの動作速度が大幅に向上しました。
内部の計算エンジンの大幅な換装によりこれを実現いたしましたので、以前のバージョンと全く同一の結果が得られるとは限らない点をご留意ください

上記修正に伴い、「Word Embedding」アイコンにおいて以前のバージョンと次の機能差分が生じます。

- 本バージョン **Deep Learner 1.3** で廃止した機能・パラメータ
 - ✧ 近似手法の「なし」
Negative Sampling か Hierarchical Softmax のいずれかを必ず用います。
 - ✧ Smooth Exponent (内部的に固定の値が適用されます)
 - ✧ ミニバッチサイズ (内部的に固定の値が適用されます)
 - ✧ GPU の利用
 - ✧ 追加学習機能
 - ✧ 学習の中断機能
- 本バージョン **Deep Learner 1.3** で新設した機能・パラメータ
 - ✧ 計算対象単語の頻度下限値 を指定可能に
 - ✧ 初期値から学習率減少させていく際に、終了時の値を指定していたのを最小値に変更

- Model Optimizer 利用時に計算が発散するなど処理が異常終了した場合、Model Optimizer の繰り返し処理全体が停止していたのを、停止せず次の計算を続行するよう修正いたしました。
- 「Word Embedding」アイコンのアイコン画像を変更いたしました。
これはプロジェクトボード上に新規にアイコンを置いた場合に反映されます。Deep Learner 1.2 以前のプロジェクトを読み込んだ場合は旧アイコンが表示されます。