

報道関係者各位

2022年5月26日
フォルシア株式会社

データクレンジングツール「Masstery」のカテゴリ自動分類精度が大幅向上 ～商品情報のカテゴリ分類を自動化、分類精度は最大96%～

フォルシア株式会社（本社：東京都新宿区、代表取締役社長：屋代浩子、以下フォルシア）はこのほど、当社が提供するデータクレンジングツール「Masstery」に付随するオプション機能「カテゴリ自動分類」の改良を重ね、商品データのカテゴリを自動で分類する際の精度が最大96%を記録したことをお知らせします。

データクレンジングツール「Masstery」は、不揃いなデータのフォーマット統一、更新情報の取得、不正データの検知などの70以上の機能を自由に組み合わせることで、人手でしか対応できなかったデータ整備を自動化するSaaS型のデータクレンジングツールです。2020年のサービスリリース以来、多くの企業様にご導入・ご好評頂いております。

この度、主に大量の商品を取り扱われるECサイト運営者様より強くご要望頂いておりました、カテゴリ自動分類機能の精度が「ルールベース」と「機械学習」の2つのロジックを組み合わせることで大幅に向上しました。

「カテゴリ自動分類」は、用途や形状に共通点のある商品群をカテゴリごとに自動で分類する機能です。カテゴリ別に分類された商品情報は、ECサイトや社内の在庫管理画面上での検索性が飛躍的に向上します。従来の「カテゴリ自動分類」ツールでは一般的に、分類根拠は明示できないものの、全ての商品情報に対して結果を出力することで高い検出率を実現させる「機械学習」ロジックが用いられていましたが、Massteryではこのたび、分類根拠のある商品情報に対してのみ分類を行う「ルールベース」を基本ロジックとして採用し、教師データから抽出した分類ルールや弊社独自の汎用性の高いルールを適用して分類することで高い精度を実現しました。また、「ルールベース」では分類できなかった一部の商品情報に対しては「機械学習」による分類を行い、全ての商品情報に対して結果を出力することが可能です。

今回の精度の向上においては、これら2つのロジックを組み合わせることで両方のメリットを享受し、総合的な検出率は100%、精度は最大96%にまで到達しました。これにより、大量の商品の更新作業を行うECサイト運営者様の業務効率化や、サイトの検索性向上による売上げ拡大に寄与できるだけでなく、現在はWEB上での販売を行っていないユーザー様のネット販売事業の立ち上げにもご活用いただけます。

▼「ルールベース」と「機械学習」を組み合わせることで分類した例(96%記録当時)

カテゴリ自動分類

Massteryでは「ルールベース」と「機械学習」の2つのロジックを組み合わせることで、各ロジックのメリットを活かし、分類の**正答率(精度)**を**約96%**まで高めています。

[STEP1] ルールベース処理

教師データを元にして予測が可能な商品のみ(約96%)について、カテゴリを分類します。

△ 検出率: 予測した数/予測する対象 = 96/100 = **約96%** ○ 精度: 正解した数/予測した数 = 93/96 = **約97%**

[STEP2] 機械学習処理

ルールベースでは分類できなかったデータ(約4%)について、機械学習による予測を元にカテゴリを分類します。

○ 検出率: 予測した数/予測する対象 = 4/4 = **100%** △ 精度: 正解した数/予測した数 = 3/4 = **約88.0%**

ルールベース処理のメリットを活かす!

ルールベース処理は教師データ(過去の分類実績)に基づいてカテゴリを分類するため、**機械学習に比べて分類精度が非常に高い**です。

このメリットを最大限に活かすため、Massteryではルールベース処理を基本としています。

2つのメリットを良いとご取り!

機械学習処理のメリットを活かす!

ルールベースでは分類できなかったデータについては**未知の分類規則を発見してくれる可能性がある**機械学習の予測に基づいてカテゴリ分類を行います。

2つのロジックを組み合わせる結果の精度: $(93+3)/(96+4) =$ **約96%**

※数値は実際にユーザー様にご使用頂いた際に得られた結果です。
※小数点以下を除いて記載しています。

Copyright © FORCIA, Inc. All rights reserved.

【データクレンジングツール「Masstery」の特徴】

特徴1：手作業で行っていたデータクレンジング作業をノーコードで自動化

これまで手作業で行っていたデータの整備・更新・統合作業を簡単な3ステップの操作で自動化することができます。データの変換ルールも難しいコードを用いずに日本語的な表現を用いて登録することが可能です。

<操作手順>

1. データの変換ルールを登録
2. データをアップロード
3. データをエクスポート

特徴2：データ整備に必要な70種類以上の機能を組み合わせてご希望の変換を実現

豊富に取り揃えた70種類以上の機能を組み合わせて多様化するデータ整備のニーズにお応えします。VLOOKUP や別シートから引用しての置換など、通常の Excel を用いた作業では複数のシートを横断する変換も自動化が可能です。

▼各機能の詳細については下記製品ホームページを御覧ください。

<https://mstr.forcia.com/>

特徴3：複数のフォーマットへの出力が何度でも可能

複数のデータをアップロードするだけで不揃いなフォーマットのデータを、ご希望のフォーマットに揃えて統合・更新します。また、出力する際のフォーマットも複数登録でき、何度でも出力し直せます。どのご担当様にとっても活用しやすい形でのデータ整備が可能です。

フォルシアは創業以来、膨大かつ複雑なデータの活用を強みとし、旅行業界を皮切りに商社やメーカー、試薬業界など業界を問わずデータを扱われる全ての企業様の課題解決を進めてきました。当社は今後も、国内・海外の多種多様な企業様の業務の効率化の一助となれますよう、製品の開発・改良を進めてまいります。

【操作画面イメージ】

▼ホーム画面(データアップロード用画面)



対象ファイル	変換完了日時	完了日時	件数
20220301_1500_M2210.csv	04/22 10:19	04/22 10:23	996

▼変換ルール設定画面

Mastery ホーム 変換設定 参照マスタ 基幹データ取込 よくある質問 user-test-b00

変換ルール 出力項目 手動カテゴリ変換 仕入先リスト 表示形式 一括 表

+ 変換ルールを追加

キーワードで絞り込み
キーワードを入力

仕入先で絞り込み
 「全て」を命める
以下の候補を絞り込み

メーカーA
 メーカーB
 メーカーC

1	出力内容	表示用項目 の値を "仕入先_包装数" にする	+ 追加
	適用条件	メーカーA	
2	出力内容	百値値 の値を 0 にする	+ 追加
	適用条件	メーカーA	
3	出力内容	希望納入価格 の値を item_japanese["希望小売価格"] にする	+ 追加
	適用条件	メーカーA	
4	出力内容	希望納入価格 の値を 0 にする	+ 追加
	適用条件	メーカーA	
5	出力内容	製品名_和名_全角 の値を Strade 削除にする	+ 追加
	適用条件	メーカーA	
6	出力内容	商品コード の値を "I_仕入先コード_顧客名_基幹システムコード番号" にする	+ 追加
	適用条件	メーカーA	
7	出力内容	製品元商品コードNo の値を "仕入先_基幹システムコード番号" にする	+ 追加
	適用条件	メーカーA	
8	出力内容	プライスグループ の値を "仕入先_プライスグループ" にする	+ 追加
	適用条件	メーカーA	
9	出力内容	保存条件区分 の値を lookup["保存/輸送温度"]M_輸送温度条件CODE"保存条件区分" にする	+ 追加
	適用条件	メーカーA	

0行目の次に新規作成 状態 On

出力内容

出力項目 の値を

値またはロジック にする

適用条件

仕入先 のとき

全ての仕入先に適用する場合は、「全て」を選択してください。

適用条件1 が に一致する とき

適用条件がない場合は、設定の必要はありません。

[適用条件を追加](#) [キャンセル](#) [登録](#)