

システム開発での機械学習の活かし方

株式会社クレスコ
株式会社インテック

伊佐陽介
小田博司

y-isa@cresco.co.jp
oda_hiroshi@intec.co.jp

開発における問題点

システム開発で収集するデータ(工数や成果物の指摘など)は、開発中では状況分析等に活用され、開発終了後には、過去の実績として見積もり等で活用される。しかし、収集されたデータは機密性が高く、狭い範囲の暗黙知になってしまいうケースが多い。そのため、経験の浅い担当者はそのノウハウを活用できない。

手法・ツールの適用による解決

過去のシステム開発で収集されたデータを機械学習の学習データとして用いることで、経験の浅い担当者に対して、開発の状況分析や新たな作業の工数見積もりを補助できることが期待できる。
今回はプロジェクト管理ツールである「Redmine」に蓄積された情報を用いる。

取り組み

各自の社内で収集した実際のシステム開発のデータを利用するため、セキュリティの観点で各自のデータを共有して1つのテーマに取り組むのが困難だった。そのため、独立した2つの小テーマを設けて、各自で本テーマに取り組んだ。

- 小テーマ1: 開発時に作成した成果物の指摘内容から指摘分類を予測する学習モデルの作成
- 小テーマ2: 作業タスクのタイトルから作業工数を予測する学習モデルの作成

小テーマ1

背景

- レビューの結果(指摘分類, 件数, ランク)は品質管理担当者により成果物の品質傾向として分析される
- レビューの結果自体がレビューアの主観や経験に依存しやすい

➡ 品質傾向の分析がレビューアに依存してしまっている

アプローチ

- 指摘内容と正しい指摘分類が設定されたデータの組み合わせ(6558件)を学習データとして用いる
- Pythonの機械学習ライブラリ(scikit-learn), 形態素解析ライブラリ(Janome)を利用し, 機械学習モデルを作成
- 作成した機械学習のモデルはPythonのWebフレームワーク(flask)を利用し, WebAPI化

結果

学習データ20%をテストデータとし, ホールドアウト法を用いてモデルを検証。約80%が正しい指摘分類に予測できた。

小テーマ2

アプローチ

- 作業タスクのタイトルと作業工数が設定されたデータの組み合わせ(8000件)を学習データとして用いる。
- 形態素解析ライブラリ(MeCab), 機械学習ライブラリ(H2O), word2vecのアルゴリズムを用いて, 作業工数を予測するモデルを作成。



結果

タイトル	工数(h)
Azureについて調査する	2
アジャイル開発について調査する	4

上記以外に5つの結果を確認した。

まとめと課題

まとめ

少ない種類のデータの組み合わせでも, 作成した予測モデルで, 一定の結果が得られることがわかった。
暗黙知となっていたノウハウを習熟度が低い担当者でも活用できるようになる点は非常に効果的だったと考える。

課題

実際のシステム開発の現場で今回作成した学習モデルを利用するには, 学習データとして扱うデータの選別, データの前処理(不要文字の除去や単語の正規化)の強化が必要,