

Assessing the Use of AutoML for Data-Driven Software Engineering

Fabio Calefato, Luigi Quaranta, Filippo Lanubile
University of Bari
Bari, Italy
{fabio.calefato, luigi.quaranta, filippo.lanubile}@uniba.it

Marcos Kalinowski
PUC-Rio
Rio de Janeiro, Brazil
kalinowski@inf.puc-rio.br

2023 ACM/IEEE International Symposium on Empirical Software
Engineering and Measurement (ESEM), New Orleans, LA, USA, 2023
pp. 1-12.

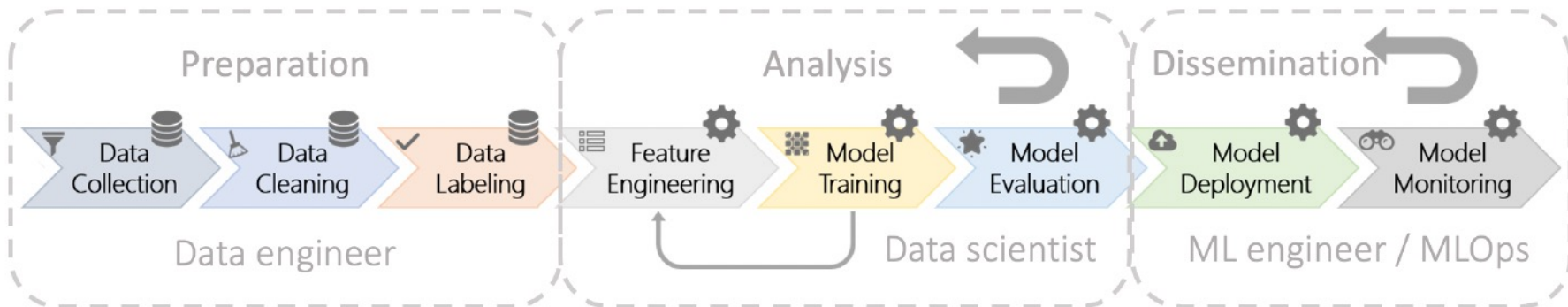
発表：山村

AI/ML(Machine Learning)の普及・課題

技術の進歩により，AI/MLが様々な分野で採用される
—自動車，ビジネス，ヘルスケア etc.

AI/ML採用には，工程ごとに様々な専門家が必要

図：AI/ML導入の典型的なフロー



例) データエンジニア，データサイエンティスト，ソフトウェア・機械学習エンジニア，ドメイン専門家

→専門家の**人材確保**が困難

AutoMLの台頭

AutoML (Auto Machine Learning)

目標：AI/ML導入の一連のプロセス全体の**自動化**

- ・ 現状，プロセスの一部の自動化をAutoMLツールにより実現
例) 特徴量モデリング，モデルの選定，ハイパーパラメータチューニング
- ・ 企業における人材不足の解消，生産性の向上などに期待

SE(Software Engineering)分野においても活用される

例) 田中ら

AutoMLを用いた，ソフトウェア欠陥予測モデルの開発

AutoMLの認知

一方，AutoMLの採用実態や，開発者の認知・期待についてはほとんど知見がない

関連研究：

- ・ AutoMLはOpenMLに登録されたワークフローのうち2%
- ・ AI/ML開発チームメンバーの内，2,30%はAutoMLを**使ったことがない**
- ・ データサイエンティスト視点は分析されているが，**データエンジニア**や**機械学習エンジニア**からの視点は明らかにされていない

SE分野におけるAutoMLツールの性能分析はほとんど行われていない

田中らの研究以外，SE分野からのデータセットを利用した研究はない

目的・アプローチ

目的

SE分野における，AutoMLの活用実態や，開発者のAutoMLに対する認知・期待を明らかにする

RQ

RQ1：AutoMLツールのSEタスクにおける性能は？

RQ2：AutoMLを使用するエンジニアの認識は？

アプローチ

- ・ 実際のAutoMLツールを用いた，SEタスクの比較実験
- ・ AutoMLのユーザー調査

RQ1 (AutoML for SE) | 使用データ

AutoMLのSEタスクにおける性能を，テキスト分類タスク(感情分析)で評価

—AutoMLツールと既存手法の性能を比較

①GitHubデータセット

プルリクエストやコミットメッセージから，約7000件のメッセージに手動で感情ラベルをつけたもの

ラベル：positive, negative, neutral

②jiraデータセット

人気OSSプロジェクトのissueコメント文約6000件に手動で感情ラベルをつけたもの

ラベル：love, joy, anger, fear, surprise, sadness

RQ1 | 使用AutoMLツール

商用ツールとオープンソースツールを合計12種用意

- ・ 有名企業による商用AutoML：4個
—Amazon, Google, IBM, Microsoft
- ・ 「automl」でGoogle検索して見つかった商用AutoML：4個
- ・ Github検索で見つけてきたオープンソースAutoML：4個

ツールによって、できること・できないことが異なる

- ・ ラベル付けや、デプロイ、監視、XAIに対応していないモデルもある
—オープンソースに比べ、商用はよりできることが多い(End to End)
- ・ 分析段階(特徴量, 学習, 評価)はどのツールも
できるっぽいからそこで比較？

XAI : Explainable AI
AIがなぜその出力をしたのかを
人間が理解できるように説明できる
機能

結果 (RQ1)

分類性能をPrecision, Recall, F1, そのMacro-F1, Micro-F1で評価

結果 (一部抜粋)

Tools	Datasets	Macro-F1	Micro-F1
DataRobot AI Cloud	Github	0.92	0.92
	jira	0.82	0.87
Auto Keras	Git Hub	0.91	0.91
	jira	0.83	0.87
IBM Watson AutoAI	Git Hub	0.55	0.58
	jira	0.61	0.76
RoBERTa	Git Hub	0.92	0.92
Senti4SD	jira	0.92	0.92

赤：AutoMLで性能が良かったもの、青：AutoMLで性能が悪かったもの、緑：既存手法(notAutoML)

既存手法に十分匹敵するAutoMLツールが存在

→SE分野でもAutoMLは適用可能

RQ2 | AutoMLに関する認知の調査

Webアンケートで、43の質問を通じて、エンジニアのAutoMLに対する認識を調査

主な調査内容

①AI/MLプロジェクトに関する経験について

②for AutoML使用経験者

AutoMLを工程のどの部分に利用・適合しているか
またAutoML利用時に課題・不便を感じるか

③for 全員

AutoMLをどのように利用したいか、
AutoMLによりサポートしてほしい工程
AutoMLに欠けている機能はあるか

聞いた人

n = 45, コンピュータサイエンス経験者, ML開発経験者

RQ2結果（回答内容の要約）

AutoMLは、主に分析段階(特徴量, 学習, 評価)で利用

- ・ 前処理段階・普及段階の工程では, AutoMLの支援は不十分
- ・ **オープンソース**のツールは, データラベリングなど前処理に対応していないものが多い

ML/MLOpsエンジニアは柔軟性や拡張性を重視

- ・ **商用ツール**は, End to Endな性質なため, 工程の一部分を自動化したい際に不便（柔軟な**オープンソース**を好む傾向）
- ・ 一方, 非エキスパートは使いやすさを重視

透明性(XAI)が不足している

- ・ AutoMLの回答(出力)は基本的に信頼しているが, その解釈性・説明性については不十分

まとめ

AutoMLのSE分野における性能と， AutoMLの認知について調査

AutoMLは， 感情分類タスクにおいて， 既存手法と同等の性能を発揮

アンケート調査により， AutoMLに対する認知が明らかになった

- ・ 前処理・展開段階で支援が不十分
- ・ MLエンジニアは柔軟性を求める
- ・ オープンソース・商用ツールの性質の違い
- ・ XAIの重要性

所感

AutoMLツール自体の概念や仕組みに関する記述はなかったもので、もっと知るなら別途勉強が必要

—特に、前処理の自動化や展開の自動化はイメージしづらい

データエンジニア・データサイエンティスト・機械学習エンジニアの違いについて知れた

—実はあんまりよく分かってなかったので、知れてうれしかった

—あんまりそれぞれの視点の違いについて語られていなかった気が？