

# 機械学習要求工学の現状と課題

北川貴之<sup>†1</sup> 長岡武志<sup>†1</sup>

機械学習要求工学における現状と課題について議論する。

## Challenges for Machine Learning Requirements Engineering

Takayuki Kitagawa<sup>†1</sup> Takeshi Nagaoka<sup>†1</sup>

This article discusses the challenges for Machine Learning Requirements Engineering.

### 1. はじめに

情報サービス産業では、デジタルビジネスへの対応が課題となっている<sup>[1]</sup>。デジタルビジネスの実現において要求工学のアプローチは、問題定義のための課題発見へと転換が求められている。また IoT (Internet of Things) や Web の発展と共に、ヒトや情報機器のデータがリアルタイムに収集可能となっている。今後のソフトウェアは、これらのデータから有用な価値を抽出し、活用することが必要になると考えられる。このような環境の変化の中でデータ駆動要求工学(Data-Driven Requirements Engineering)というアプローチが出てきている<sup>[2]</sup>。また、データから価値を抽出し活用する方法の1つに、機械学習を取り入れたシステム(機械学習応用システム)があると考えられる。しかし、機械学習応用システムの開発方法論(機械学習工学)は確立されていない<sup>[3]</sup>。本稿では、機械学習工学における要求定義の方法論(機械学習要求工学)の現状と課題を議論する。

### 2. 機械学習×要求工学

筆者らは、機械学習の要求工学への適用は、図1のように進むと考える。以降、各ステップにおける現状を示す。

#### (1) データの収集

Webに公開されている要求文書から構築したデータセット<sup>[4]</sup>やOSSプロジェクトの成果物から収集したデータセットの提案がある<sup>[5]</sup>。一方、不特定多数の個人(Crowd)からデータを収集するというアプローチについても報告されている<sup>[6]</sup>。今後、CrowdがSNSやアプリストア等の媒体を通して発信する、ソフトウェア利用に関するフィードバックの取り扱いが必要となると考える。

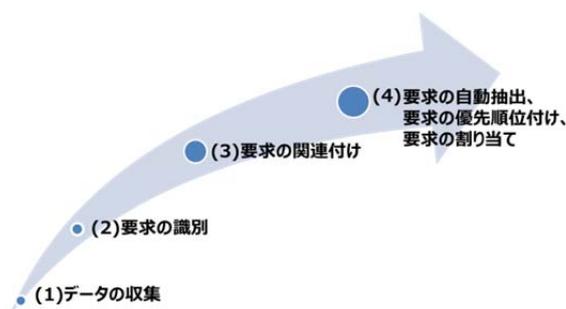


図1 機械学習の要求工学への適用

#### (2) 要求の識別

要求の識別では、蓄積した大量のデータから要求に関する情報を抽出する。Word2Vecと深層学習を組合せた適用<sup>[7]</sup>やSVM(Support Vector Machine)の適用<sup>[8]</sup>により高い精度(precision)と再現度(recall)を得たと報告されている。例えば文献[8]では、SVMによりデータセット内の機能要求(FR)と非機能要求(NFR)をどれくらい正確に自動識別できるかを調査した結果、92%の精度と再現度を得ている。非機能では、セキュリティ要求および性能要求の識別で精度と再現度が高かったとしている。非機能の品質特性により識別精度が異なる可能性が示唆されている。

#### (3) 要求の関連付け

要求の関連付けの1つにトレーサビリティの構築がある。トレースの紐付けは、要求と設計/ソースコード/テスト項目/その他成果物へ関連付けをする。これまでも情報検索(Information Retrieval(IR))や機械学習によりトレースの紐付けを自動化する方法が提案されてきた。しかし、成果物の意味理解やドメイン知識の活用ができず、正しくない結果を導く傾向があった。文献[9]では、深層学習を適用することで成果物の意味構造とドメイン知識を結び付けたトレースの紐付け方法を提案している。今後、要求の関連付けでは、成果物の意味理解やドメイン知識の活用が進むと考える。

<sup>†1</sup> 東芝デジタルソリューションズ(株)  
Toshiba Digital Solutions Corporation.

#### (4) 要求の自動抽出, 優先順位付け, 割り当て

機械学習の要求工学への適用は, 要求の自動抽出, 優先順位付け, 割り当てへと進むと考える. 例えば, 要求の割り当てでは, 要求に合致するパッケージの利用ノウハウを自動抽出する技術が提案されている<sup>[10]</sup>. ただ, これらの活動は, 従来の人手に頼る部分が大きい状況である.

### 3. 機械学習×要求工学の課題

機械学習応用システムでは, 収集・蓄積したデータからモデル(ニューラルネットワーク)を作成し, そのモデルを活用してシステムを構築する. 以降, 機械学習応用システム開発の要求定義において重要と考える課題を以下に示す.

#### 3.1 新しいデータセットの取り扱い

IoT や Web の発展により, Crowd が発信するデータの取得が可能になってきている. 現状の要求工学では, 主にビジネス活動を通して収集・蓄積したデータを取り扱ってきた. 機械学習応用システムの要求定義では, ビジネス活動を通して収集・蓄積したデータに Crowd が発信するデータを加えた, 新たなデータセットを取り扱う必要がある.

新たなデータセットを要求工学で取り扱うには, 課題がある. 要求定義では, ステークホルダが重要であり, どのステークホルダから獲得したのかによって要求の重要度が異なる. これまでは, 重要なステークホルダを特定し, そのステークホルダから要求を獲得してきた. 一方, 新たなデータセットでは, 様々なステークホルダからデータを収集するためデータ量が多くなり, 個々の要求の影響度が低くなると想定される. そのため, 要求の重要度を分析する場合, ステークホルダの情報を加味しなければ, 正しくない結果を導く可能性がある. このように新たなデータセットの活用に必要な課題を洗い出し, 解決をしていく必要がある. 加えて, 画像認識技術, 音声認識技術の進展により身体データや感情データについても取得が可能になると想定される. これらの情報をどのようにデータセットへ取り込み, 活用していくかも課題である.

#### 3.2 推定根拠の提示

機械学習の現実問題への適用は段階が進むと, 問題解決の意思決定に踏み込むことになると考えられる. 図1の「(4) 要求の自動抽出, 優先順位付け, 割り当て」のうち優先順位付けは, 「売上げ拡大を達成するために, どの要求から実現すべきか」という問いかけに答えるものである. しかし, 機械学習応用システムの開発においては, 機械学習を使ったモデルの開発者とモデルを利用する意思決定者が異なることがある. このような場合, 意思決定者は, 要求を実現するためのモデルを採用するための根拠が必要になる. ただ, 一般的に機械学習の結果の根拠の説明は難しいと言わ

れており, 現状では根拠を説明できない可能性が高い. そこで, 機械学習の結果の根拠を説明する技術の研究開発が必要である. 例えば, 要求にモデルを割り当てることで, 要求を実現するためのモデル採用の根拠の説明性を高めることが考えられる. 他にどのような方法があるのかを議論したい.

#### 3.3 その他

機械学習×要求工学の実現には, 上記の課題に加えて, 機械学習を利用することを前提とした, 要求定義プロセスの改善が考えられる. また, 機械学習×要求工学は, 始まったばかりであり, その実現にはさまざまな課題があると考える. これまでに挙げた課題の他に, どのような課題があるのかについて議論したい.

### 4. おわりに

本稿で議論した機械学習×要求工学は, 要求工学にとどまらず, ソフトウェア開発の様々な各局面で共通すると考えられる. 本稿で提起した課題への研究が求められている.

### 参考文献

- [1] 青山 幹雄; デジタルビジネス時代の要求工学, JISA(編), 情報サービス産業白書 2017, pp.135-141, 2017.
- [2] Maalej, H. et al.; Toward Data-Driven Requirements Engineering, IEEE Software, Vol.33, No.1, pp.48-54, Jan.-Feb. 2016.
- [3] 丸山 宏, 機械学習工学に向けて, 日本ソフトウェア科学会第34回大会 講演論文集, 日本ソフトウェア科学会, 2017.
- [4] Michael Rath et al.; The IlmSeven Dataset, Proc. of RE 2017, Sep. 2017.
- [5] A Ferrari et al.; PURE:a Dataset of Public Requirements Documents, Proc. of RE 2017, Sep. 2017.
- [6] Pradeep K. et al.; Toward Automating Crowd RD, Proc. of RE 2017, Sep. 2017.
- [7] Alex Dekhya et al.; RE Data Challenge: Requirements Identification with Word2Vec and TensorFlow, Proc. of RE 2017, Sep. 2017.
- [8] Z Kurtanovic et al.; Automatically Classifying Functional and Non-functional Requirements Using Supervised Machine Learning, Proc. of RE 2017, Sep. 2017.
- [9] Jin Guo et al.; Semantically Enhanced Software Traceability Using Deep Learning Techniques, Proc. of ICSE 2017, May. 2017.
- [10] 中村 覚 他; 顧客要件に合致するパッケージシステムのノウハウ事例抽出支援システムの開発, 人工知能学会第23回知識・技術・技能の伝承支援研究会(SIG-KST), Jan. 2014.