

時刻属性付きデータフロー図の自動処理によるシステム解析時間の短縮 Reduction of system analysis time by data flow diagram with time attribute

高橋 加寿子[†] 塚本 良太[†] 佐藤 隆佑[†] 田村 孝之[†]
Kazuko Takahashi Ryota Tsukamoto Ryusuke Sato Takayuki Tamura

1. はじめに

IoT(Internet of Things)の普及に伴い、機器毎に生成するセンサデータ等の時系列データから予測モデルを生成し、機器の制御をするシステム(以下、IoT 機器予測制御システム)の増加が見込まれる。IoT 機器予測制御システムは、予測部と制御部からなり、予測部は、時系列データを基に将来の動作を予測するモデルを作成する。制御部は生成された予測モデルを用いて機器の適切な制御を実施する。

このような IoT 機器予測制御システムにおいて入力となる時系列データが機器の設置場所・条件により変化する場合に対応する予測モデルも変化するため、運用開始後も機器の制御条件が適切であるか確認するための評価を行う。

評価においてシステムに改善が必要と判断された場合は、原因の検証を検証環境にて行うため、データをコピーし運用環境と同じ状態とする必要がある。この時、全データをコピーするとデータ量の観点から運用環境に負荷を与えるため、コピーするデータを最小限とするが、データ範囲の特定には仕様書等の調査が必要であり、時間を要するという問題点があった。

そこで、本稿では IoT 機器予測制御システムのプログラムとテーブルの関係を RDF クエリ言語で構成可能な属性情報を付属したデータフロー図により管理する。これにより、任意の予測モデルのテーブル名・対象日時から当該予測モデルを生成する際に用いた時系列データの最小限の範囲を決定し、IoT 機器予測制御システムの運用環境から検証環境にコピーするシステムを提案する。

2. IoT 機器予測制御システムの構成と課題

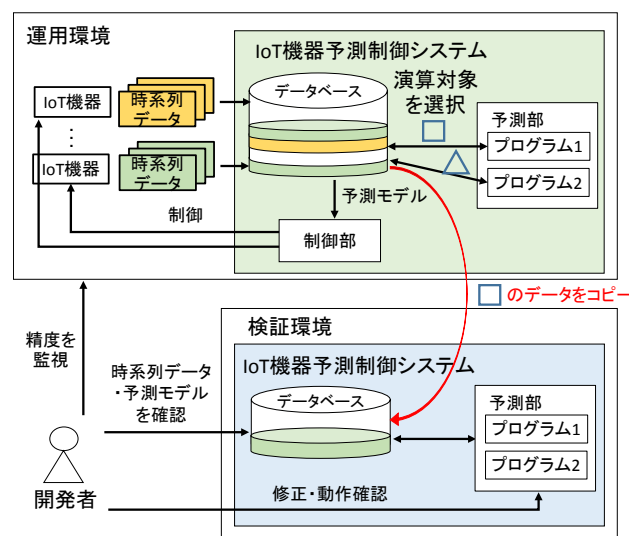


図 1 継続的な改良を考慮した IoT 機器予測制御システムの構成

本稿では IoT 機器予測制御システムの運用環境と検証環境の構成を図 1 とする。IoT 機器予測制御システムの予測部はデータベース内の様々な時系列データから制御対象機器の時系列データを選択し、制御対象機器毎の予測モデルを生成する。データベース内には機器の様々な要素が複数のテーブルに蓄積されており、データは数年間分にもなる。予測部では複数のプログラムが複数回実行されるため、同じ対象日時の予測モデルでも複数の生成日時が存在する。それぞれのプログラムはテーブル内に点在する一部の期間を入力として用いるが、この期間は予測モデルが対象とする日時により異なる。また、プログラムが出力するデータが中間データとしてテーブルに格納され、他のプログラムの入力となる場合もある。そのため、検証に必要な運用環境の状態を検証環境に再現するために必要な最小限のデータ(以下、再現用データ)を格納しているテーブルや期間を特定するためには各種仕様書を調査する必要がある。調査は、システム仕様書から対象のプログラムに対する入出力テーブルを、S/W 設計書やプログラムからアルゴリズムの詳細や、各テーブルの定義を取得し、予測モデルの対象日時から対応した再現用データの日時を判断する。この時、予測部の一部のプログラムの検証を行う場合は、再現用データとして中間データを含む可能性があり、中間データの生成日時への考慮も必要になる。このように再現用データの特定には各種仕様書の調査を行うために時間を要するという課題がある。

3. 時刻属性付きデータフロー図による選択条件記述

2 章の課題を解決するために、任意の予測モデルから再現用データを検索できる属性付きデータフロー図を構成するための表記法を提案する。本稿では属性を時刻属性として定義しているが再現用データを特定するための条件として距離属性といった他のデータの分類においても構成可能である。

3.1 時刻属性付きデータフロー図

予測部におけるプログラムと各プログラムの入出力データの関係を表す図として、構造化システム分析・設計手法であるデータフロー図[1]がある。データフロー図からはプログラムと、入出力するデータが格納されるテーブルの関係を読み取ることができる。しかし、テーブル内のいつのデータかを特定するための情報は記載されておらず、詳細なデータ範囲を知るためには仕様書とつぎ合わせる必要がある。

そこで、データを特定するための情報として、データフロー図に属性情報を追加で定義した時刻属性付きデータフロー図を構成する。時刻属性付きデータフロー図は、通常のデータフロー図に加え、プログラムとテーブルがやり取りするデータの詳細情報として、制御する機器の ID と時刻属性付きの情報であるシステムの実行日時・対象日時を新たに定義する。実行日時・対象日時は予測モデルの日時

[†] 三菱電機株式会社, Mitsubishi Electric Corporation

からの相対値を定義する。これにより任意の予測モデルの機器の ID・テーブル名・対象時刻・実行時刻から、データフロー図を辿ることで入力に用いた時系列データの範囲の検索ができる。

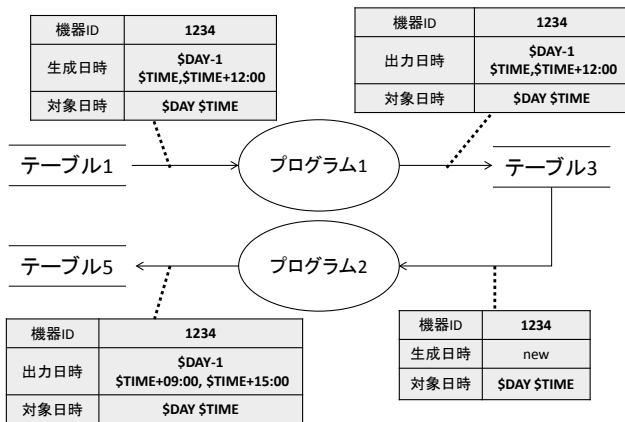


図 2 時刻属性付きデータフロー図

3.2 時刻属性付きデータフロー図の表記法

時刻属性付きデータフロー図は RDF の Turtle 記法で表記し、クエリ言語 SPARQL[2]にて自動処理する。

図 2 におけるテーブル 1 とプログラム 1 の関係の表記例を図 3 に示す。プログラム 1 を実行するためにテーブル 1 から取得する機器 ID は、検索時に指定される任意の機器 ID を示す 1234 である。テーブル 1 にデータが生成された日時である生成日時は、テーブル 5 に格納されている予測モデルの対象日 DAY の 1 日前を意味する DAY-1 である。時刻は特定の時刻、もしくは予測モデルの時刻との差にて示される。データの生成日時においても同様に示される。これらによって定義されたテーブル 1 のデータはプログラム 1 に入力される。テーブルとプログラムの関係は link:により関係を表記する。このように、テーブルがプログラムにデータを入力する関係と、プログラムがテーブルにデータを出力する関係をそれぞれ示すことで、時刻属性付きデータフロー図を表記することが可能である。

```
table: テーブル1 link:取得 [ data: ID "1234";
  data:生成日時 "$DAY-1 $TIME, $DAY-1 $TIME+12:00";
  data:対象日時 "$DAY $TIME";
  link:入力 program: プログラム1 ].
```

図 3 時刻属性付きデータフロー図の表記法

3.3 再現用データ範囲の検索方法

時刻属性付きデータフロー図から予測モデルの情報を用いて再現用データを検索する。予測モデルの情報とは、検証環境にて再現する予測モデルが格納されているテーブル名と機器の ID・出力時刻・対象時刻である。例として予測モデルが図 2 のテーブル 5 に格納されているとしたときの再現用データ検索の流れを示す。予測モデルが格納されているテーブル 5 にデータを出力しているプログラムを「link:」により検出する。検出されたプログラム 2 にデータを入力しているテーブルを同様に検出する。この処理を繰り返すことで、予測部の入力となるデータが格納されているテーブルが検出できる。図 2 ではテーブル 1、2、4 が該当する。それぞれのテーブル内から取得するデータ範囲は、予測モデルの機器 ID・出力時刻・対象時刻の値を

「data:」に対応させることで再現用データの範囲を決定可能である。

図 2 においてテーブル 5 に格納されている予測モデルの再現用データの範囲を検索する例を示す。

予測モデルの範囲として、機器 ID:1234・出力日時:2017/05/13 09:00・対象日時:2017/05/14 00:00~23:59 とする。時刻属性付きデータフロー図の link を辿ることでテーブル 1 を検索する。出力日時が 2017/05/13 09:00 の時点でテーブル 1 に格納されているデータは 2017/05/13 00:00 に生成されたものであることから再現用データを取得する SQL 文は以下のように示される。

```
SELECT * FROM テーブル1
WHERE 生成日時<='2017-05-13 00:00'
AND 対象日時 BETWEEN '2017-05-14 00:00'
AND '2017-05-14 23:59'
```

4. 検証環境へのデータ複製

時刻属性付きデータフロー図を用いることで、任意の予測モデルの範囲から再現用データを検索し、自動で運用環境から開発環境に移行することが可能である。

システム解析者は再現用データ移行システムに、設計時に作成したデータ移行を行う対象の IoT 機器予測制御システムの属性付きデータフロー図を送信する。この処理は対象とする IoT 機器予測制御システム内の処理が変更された場合や、対象システムが異なる場合において実施する。次にシステム解析者は、解析を行う予測モデル情報をデータフロー図管理システムに送信することで、再現用データ範囲が生成される。再現用データ取得システムは再現用データ範囲を基に運用環境から再現用データを取得するクエリを生成し、再現用データを取得する。再現用データ投入システムは取得した再現用データを検証環境に投入するクエリを生成し実行する。これにより、自動で予測モデルの範囲に関する情報から再現用データを運用環境から検証環境に移行することが可能である。

5. 評価

あるシステムで評価したところ、従来では再現用データの調査・取得・投入にかかる時間は 1 件当たり約 4 時間必要としている。再現用データ移行システムを用いた場合、初回時にデータ処理グラフを作成する時間として従来手法の 1 件分の時間を必要とするが、2 件目からは上記の処理を出力データ範囲の入力等を考慮しても約 5 分に短縮することが可能である。

6. おわりに

本稿では、解析対象システムのプログラムとテーブルの関係を RDF クエリ言語により表記することで、システム解析に用いる再現用データを検索可能とした。

今後の課題として、解析対象システムに変更が発生した場合、任意のバージョンに対応した再現用データの検索を可能とする。

参考文献

- [1] 飯塚京土, 佐藤宏之, イコプラムディオノ, 村山隆彦, "RDF データを対象としたグラフ検索におけるクエリ生成方式の検討" 人工知能学会研究会資料, (2004)
- [2] SPARQL Query Language for RDF <https://www.w3.org/TR/rdf-sparql-query/> (参照:2017-06-22)