

# 測距センサーデータを対象とした外れ値検出手法

永島 寛子<sup>1</sup> 加藤 由花<sup>1</sup>

**概要：**近年、情報技術の進化と共に、ビジネスで使用するデータの量や種類が飛躍的に増加した。産業界における活用用途も、店舗内での顧客動向の分析、ロボットの自律的行動、機器の故障検知、新規ビジネスの創出のように多岐に渡る。しかしながら、収集したデータをそのまま分析してしまうと、望んだ結果が得られない。そのため分析前に、外れ値の除去や欠損値の補完、使用データの形式の統一、分散しているデータの統合・正規化など、データの曖昧性や不統一性を除くための前処理が不可欠である。本稿では、センサーデータを含む大量データを対象とし、前処理で行う処理を分類することにより、自動化領域を組み込んだフレームワークを提案する。3次元距離画像センサーで収集したデータによる人流解析を想定した評価項目について既存フレームワークと提案フレームワークを比較することで、センサーデータを含む大量データの分析の場合、提案フレームワークが既存フレームワークより適していることを評価した。

## 1. はじめに

近年、各種センサーやウェアラブル端末など、データを収集するための機器が増えるにつれて、収集できるデータの量や種類が増え、それらのデータを活用した多角的な分析を行えるようになってきた。それに伴い、産業界では、マーケティングのための顧客動向予測、ロボットの自律的行動、クレジットカードの不正利用検知、製造機器の故障検知、新規ビジネスの創出のように様々な分野でデータが活用されている。しかしその一方、収集したデータは、外れ値や欠損値、機器に依存した計測データ単位の差異や、表記の揺れなどを含んでおり、そのまま Business Intelligence ツールや機械学習エンジンで分析することができない。例えば、身長（概ね3桁の整数）と年齢（概ね2桁の整数）を考慮のだけでも、データが示すサイズが異なることがわかる。そのため、分析の目的を事前に検討し、分析対象データの状態を見極め、分析内容・分析ツールに適したデータを基に分析モデルを生成する必要がある。

これまで、ある分析プロジェクトが発生してから、分析モデルを生成するまで、さらには生成したモデルによる運用を見据えた維持やナレッジの展開など、データマイニングの一連の流れを示したフレームワークが提案されてきた [1]。フレームワーク内の処理は、例として、外れ値の除去やデータ抽出などの前処理や、データ分析モデルの生成・評価があり、その全てを分析者が行っている。そのため、分析対象データが大量の場合、データの量と種類、分

析手法の多様化により、前処理で行うべき処理が増え、既存のフレームワークでは分析モデルを検討・生成できる知識を有した分析者の負荷が高くなってしまふ。そこで本稿では、センサーデータを含む大量データを対象とし、前処理フェーズを自動化するプロセスを組み込んだデータマイニングフレームワークを提案する。

とくに本稿では、前処理で行う処理のうち、統計学的な分析への影響が大きい「外れ値」について着目し、外れ値を自動検知する処理を含めた。外れ値は、不正検知などの異常検知や新たな顧客の動向発見につながる重要な値である [2] が、分析対象データに外れ値を含めたまま処理してしまうと、正しい結果を算出できない。外れ値とは、データ収集機器の不具合や計器誤差の他、不正行為や人的ミス、または単に収集データの偏りによって生じる、他と異なる特徴を持った値 [3][4] であり、センサーのようなデータ取得時にノイズが入りやすいデバイスの場合、特に留意する必要がある。例えば、1時間ごとにある地点の気温を測定し、摂氏 20 度、21 度、21 度、42 度、18 度という5つの測定データを得た場合、単に、42 度という値を除外して分析を行うのか、42 度を含めた分析を行うのか、42 度という他と異なる値に焦点をあてた分析を行うのかは、分析用途に依存するが、分析時に別処理を行わなくてはならないことは自明である。

これらの問題を解決するために、提案フレームワークでは、前処理フェーズで外れ値の自動検出を定め、外れ値の対応を判断するための情報を得ることができるビジネス理解フェーズを、前処理の前に定めた。本稿では、外れ値の対処を含む評価項目を既存のフレームワークと比較するこ

<sup>1</sup> 東京女子大学 大学院理学研究科

とにより、提案フレームワークを評価する。

## 2. 関連研究

代表的なデータマイニングプロセスには、Knowledge Discovery in Database (KDD) [5][6], CRISP-DM[7], Sample, Explore, Modify, Model and Assess (SEMMA) [8], Analytics Solutions Unified Method for Data Mining/Predictive Analytics (ASUM-DM) [9][10] などがある。

以下、それぞれのフレームワークについて詳述する。

### 2.1 KDD

KDD は、1996 年に Fayyad et al により提案された、「選択」「前処理」「加工」によりデータのパターンを生成するフレームワークである。実行フェーズは 9 つに分かれており、分析プロセスの目的を決めてからデータマイニングにより抽出されたモデルとパターン決定までが 1 回のサイクルとして定義され、任意の 2 つのフェーズは繰り返し実施されるフローとなっている。

KDD フレームワークの概要を図 1 に、各フェーズの処理内容を以下に示す。

- (1) アプリケーション領域と関連している事前知識を理解し、顧客観点での最終目的を特定する。
- (2) データソースからモデル生成に必要なサブセットの抽出、またはサンプリングにより、データセットを生成する。
- (3) ノイズの除去と欠損値を補完するためのデータの収集と方針の決定、時系列情報および既知の変化の対処のためのデータの収集などの前処理を行う。
- (4) 最終目的のために重要な意味を持つ変数のみを抽出し、次元削減や変数の変換など、データの削除と射影を行うことで不変的なデータ表現を見つける。
- (5) データ統合、分類、クラスタリングなど、顧客の最終目的のための手法を選択する。
- (6) データマイニングパターンの検証、調査を行い、分析モデルの検討と、仮説を立てる。
- (7) ツリー構造、回帰、クラスタリングなどの分析を行い、データセットに含まれるデータの特徴を調査する。
- (8) 生成されたパターンの解釈と、抽出されたパターンとモデルの可視化を行う。また、必要に応じて 1 から 7 のフェーズを繰り返す。
- (9) ナレッジの発見を行う。発見したナレッジは、直接使うだけでなく、文書化して蓄積する。既知のナレッジとのコンフリクトのチェックも行う。

KDD のメリットは、どのフェーズからでも任意フェーズに戻れる点にある。前述のように、1 回のサイクルはデータマイニングのパターンモデルを生成するまでであるが、任意のフェーズは繰り返し実施されると定められており、繰り返しが生じた場合は、どのフェーズからでも修正を要

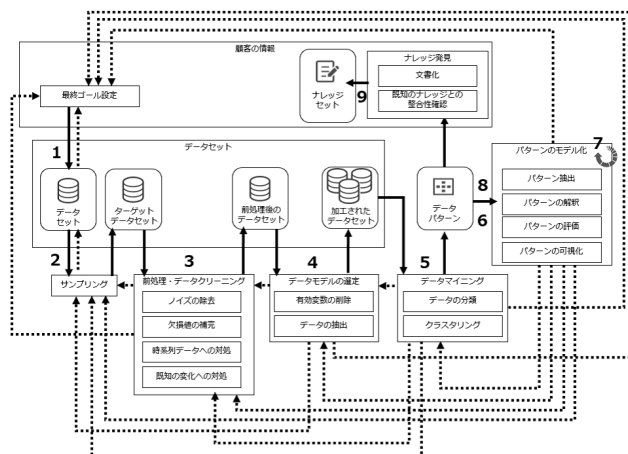


図 1 KDD フレームワーク概要 ([5] の Figure 1 を参考に作成)

するフェーズに戻ることができる。しかし、その反面、どのフェーズで後戻りを検討すればよいか、どこに戻ればよいか煩雑でわかりにくい。そのため本稿で対象とするような大量データを処理する場合、1 つの処理に時間がかかってしまうため適用が難しい。

### 2.2 CRISP-DM

CRISP-DM は、NCR, SPSS, DaimlerChrysler など、データマイニングを実際に活用している企業やコンサルタント企業による、同名のコンソーシアム (CRISP-DM) で開発され、1996 年に発案、1999 年に公開された。データマイニングのフェーズを、「Business Understanding」「Data Understanding」「Data Preparation」「Modeling」「Evaluation」「Deployment」6 つのフェーズで表現されている。6 つのフェーズのうち、Deployment を除く 5 つのステージを 1 回のサイクルとして繰り返し行われる他、顧客のビジネス目的を理解するフェーズと収集したデータの調査フェーズの繰り返し処理、モデル生成のためのデータ準備フェーズとモデル自身の調整・評価フェーズの繰り返し処理も定められている。データの前処理以降のフェーズは、ビジネスの目的から使用データと達成条件を定めることにより処理が進められるため、最初に明確化しておく必要がある。

CRISP-DM フレームワークの概要を図 2 に、各フェーズの処理内容を以下に示す。

- (1) Business Understanding : ビジネス観点で、顧客の最終目的、達成したいことの優先度や成功基準を明らかにする。
- (2) Data Understanding : 収集したデータをリスト化するなど、プロジェクトで使われているデータを洗い出し、取得したデータが要件を満たしているか評価する。
- (3) Data Preparation : 最初に収集した生データから、分析に使う最後のデータセットを作るまでのすべての処理をカバーする。具体的には、欠損値の補完、データ

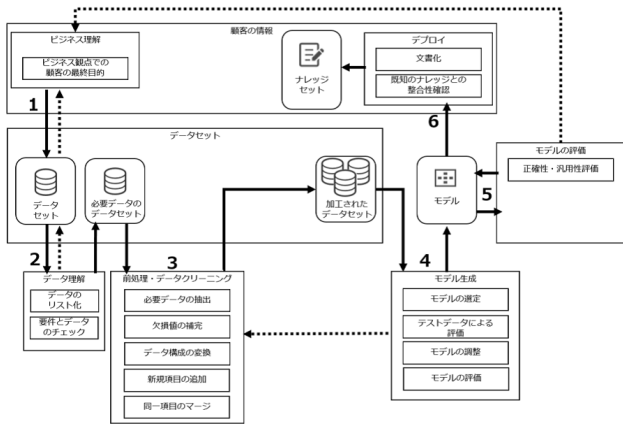


図 2 CRISP-DM 概要 ([7] の図を参考に作成)

構成の変換、新規項目の追加、同一項目のマージなどがある。

- (4) Modeling : モデリングの手法 (ディシジョンツリー, ニューラルネットワークなど) を選択し, テストデータにより評価することで, モデルを調整・確定する。
- (5) Evaluation : モデルの正確性・汎用性の評価を行う。実際のアプリケーションを使い, ビジネス目的を達成できる適切なモデルになるまで 1 から 5 のフェーズを繰り返す。
- (6) Deployment : 一般的な手順は再利用できるようにすることで, データマイニングプロセスに関するナレッジを蓄積する。

CRISP-DM のメリットは, 最初にビジネス観点での理解を確認するフェーズをフレームワークに含んでいる点である。KDD のように単に分析の目標を決めるだけでなく, 顧客が達成したい最優先の結果は何か, 成功の基準は何かなど, ビジネスとしての最終目的がより明確になるため, その後の評価基準も定めやすくなる。一方, 「Data Preparation」フェーズに外れ値に関する記述がなく, センサーなど, ノイズの影響を受けやすい方法で収集したデータを扱った場合, その外れ値の影響が最後まで組み込まれたままになってしまう可能性がある。そのため, 本稿で対象としているようなセンサーデータも含む場合, そのまま適用することは難しい。

フレームワークには組み込まれていないが, CRISP-DM コンソーシアムでは, 外れ値・欠損値について CRISP-DM Methodology として検討している [11]。データマイニング処理を, 「Application Domain」「Data Mining Problem Type」「Technical Aspect」「Tool and Technique」の 4 つの観点で分類しており, 外れ値は「Technical Aspect」に定められている。しかし, CRISP-DM のデータマイニングフェーズにはマッピングされていない。

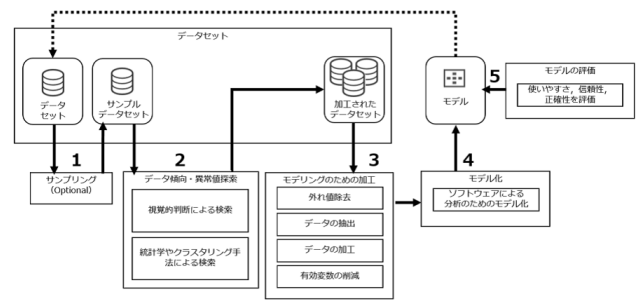


図 3 SEMMA 概要

### 2.3 SEMMA

SEMMA は, SAS Institute により提唱されたプロセスであり, データマイニングに必要なプロセスを「Sample」「Explore」「Modify」「Model」「Access」の 5 つのフェーズで表現している。本フレームワークは, SAS Institute のデータマイニングツールである「SAS Enterprise Miner」に搭載された機能にフォーカスした分類となっているが, 産業界で使われることが多い。データマイニングの対象となるデータを最初にランダムサンプリングで抽出している点, データの傾向調査, モデル化手法に統計解析手法を用いることが明記されている点が特徴である。

SEMMA フレームワークの概要を図 3 に, 各フェーズの処理内容を以下に示す。

- (1) Sample : 任意で実行されるフェーズであり, 分析対象のデータサイズが大きい場合に, 分析に重要なデータが入ったデータセットをランダムサンプリングにより生成する。サンプルを対象にデータマイニングを実施することで, データ全てを対象にするより短時間で結果を得ることができ, 何かしらのパターンが見えたときもトレースすることができる。
- (2) Explore : データの傾向を調査する。視覚的な判断が困難な場合は統計学, コレスポネンス分析 (CA), クラスタリング手法などを用いて検知する。
- (3) Modify : モデル選定のため, データを実際に新規項目を生成したり, 抽出したり, 加工する。必要に応じて, 分析のために最も重要なデータに絞るため, 外れ値や有効変数の数を減らす。
- (4) Model : ソフトウェアが, モデリングの手法 (ディシジョンツリーやニューラルネットワーク, 時系列分析) が行えるよう, モデル化する。
- (5) Assess : 生成したモデルの使いやすさ, 信頼性, 正しく動作しているかを評価する。

SEMMA のメリットは, データ量が多い場合はサンプリングをするフェーズが含まれており, 大量データを処理する場合でも, 最初から全データを対象とせず, サンプリングにより絞られたデータセットでモデルの検証を行える点である。そのため, データセットからモデルを作るまでの, 1 サイクルの試行を実施する場合に有効である。ただ

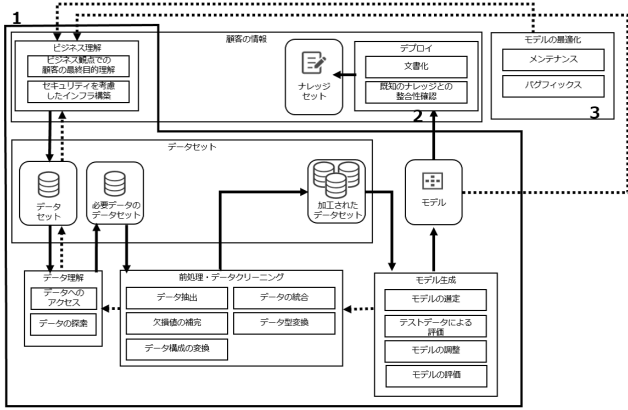


図 4 ASUM-DM 概要 ([10] Figure 1 を参考に作成)

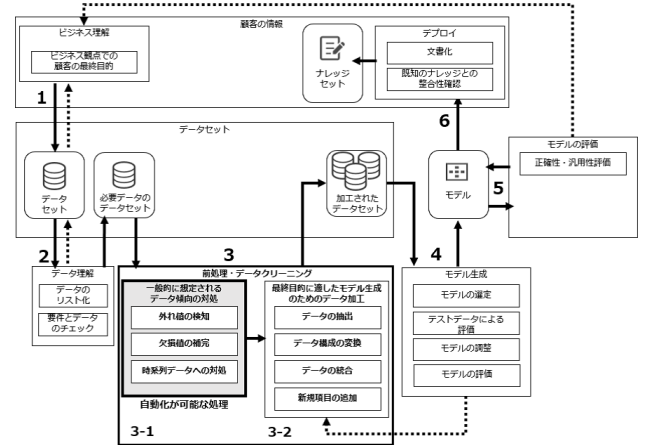


図 5 提案フレームワーク概要

し、製品の機能にフォーカスしたフェーズになっているため、ビジネスの目的理解、モデル生成後のナレッジ蓄積などの機能ではないフェーズがなく、顧客の情報やモデルの生成に関するナレッジは、別で管理する必要がある。そのため、本稿で対象としているような顧客の目的から外れ値の扱いを決定する場合、そのまま適用することは難しい。

## 2.4 ASUM-DM

IBM が提唱しているフレームワークであり、IBM の製品「SPSS」のためのモデリングである。CRISP-DM で繰り返し処理に含まれていない「Deployment」を繰り返し処理に取り込むことで、ナレッジの蓄積・共有や生成したモデルのメンテナンスフェーズを強化したモデルである。「Analyze」「Design」「Configure & Build」「Deploy」「Operate & Optimize」の 5 つのフェーズに分かれているが、Analyze と Design と Configure & Build フェーズはデータマイニングプロセスにおいて繰り返されることは自明と考え、1 つの統合されたフェーズとして表現されている。セキュリティについて説かれている点、モデル決定後のメンテナンス、バグフィックスについて説かれている点が特徴である。

ASUM-DM フレームワークの処理コンポーネントを図 4 に、各フェーズの処理内容を以下に示す。

- (1) Analyze-Design-Configure & Build: 顧客の最終目的、要望を理解し、セキュリティを考慮した基盤の構築する。予期しないエラーを避けるためにデータを理解した上で前準備を行い、モデルの生成と評価を繰り返し実施する。
- (2) Deploy: 単一ユーザーが継続運用できるように、さらに、グローバルレベルに展開できるように整理する。
- (3) Operate & Optimize: モデルのメンテナンスとバグフィックスなど、プロセスの正しい状態を維持する。

ASUM-DM のメリットは、「Deploy」フェーズを繰り返すに含めることで、モデル生成のナレッジを蓄積・更新していくことができる点である。実際にビジネスの場合は、

一度作ったシステムを数年間運用し続けるのが一般的である。そのため、ナレッジの蓄積は重要なファクターである。一方、CRISP-DM をベースにしているため、CRISP-DM 同様、外れ値に関する記述がなく、外れ値の影響を最後まで組み込んでしまう可能性がある。そのため、本稿で対象としているようなセンサーデータも含む場合、そのまま適用することは難しい。

## 3. 提案フレームワーク

本稿では、センサーデータを含む大量データの分析を対象とし、前処理フェーズで行う処理のうち、外れ値検知や欠損値の補完など、一般的に想定されるデータに対する処理の自動化を組み込んだデータマイニングフレームワークを提案する。

### 3.1 フレームワークの概要

フレームワークの処理の一部を自動化するため、モデル生成のためのデータを生成する「前処理」フェーズに 2 つのサブフェーズを定めた。前処理フェーズは、外れ値検出や欠損値の補完など一般的に想定されるデータ対処のための処理と、生成したいモデルに特化したデータ項目の抽出やデータ構成の変換などの処理に分類できる。一般的に想定されるデータの対処処理は統計手法やクラスタリングにより自動化が可能であるため、本サブフェーズを自動化対象処理として定め、フレームワーク内に自動化プロセスを組み込んだ。提案フレームワークは、大量データを対象として開発され、特定の製品に特化しないフレームワークである「CRISP-DM」をベースとし、前処理箇所を拡張する。提案フレームワークの概要を図 5 に示す。

### 3.2 フレームワークの詳細

前処理フェーズを分類した 2 つのフェーズは並列には行われず、始めに (3-1) 一般的に想定されるデータ傾向の対処を実施し、次に (3-2) 最終目的に適したモデル生成のた

めのデータ加工を実施する流れとする。

外れ値の対処は(3-1)で行う。ベースとしている CRISP-DM では、2.2 で述べたように CRISP-DM は前処理で外れ値による対処を行っていないが、本稿ではセンサーデータ含むデータを対象としているため、統計手法に大きな影響を及ぼすことが分かっている外れ値は、早めに特定しておく必要がある。しかしながら、外れ値は不正検知などの異常や顧客の突発的な行動の検知において、重要な役割を担う値である。したがって、当該フェーズでは除去は行わず、検出に留めることにする。

提案フレームワークは、大量データを対象としているため、はじめに対象データの選定ができるように、CRISP-DM 同様、最初のフェーズでビジネスの理解と、次のフェーズでデータの理解を行う。

各フェーズの処理内容を以下に示す。

- (1) Business Understanding : ビジネス観点で、顧客の最終目的、達成したいことの優先度や成功基準を明らかにする。
- (2) Data Understanding : 収集したデータをリスト化するなど、プロジェクトで使われているデータを洗い出し、取得したデータが要件を満たしているか評価する。
- (3) Data Preparation
  - (3-1) 外れ値の検出など、一般的に想定されるデータ傾向の対処を自動で実施する。
  - (3-2) 複数データの統合や、モデル生成のためのデータ抽出や変換など、最終目的に適したモデル生成のためのデータ加工を行う。
- (4) Modeling : モデリングの手法(ディビジョンツリー、ニューラルネットワークなど)を選択し、テストデータにより評価することで、モデルを調整・確定する。
- (5) Evaluation : モデルの正確性・汎用性の評価を行う。実際のアプリケーションを使い、ビジネス目的を達成できる適切なモデルになるまで1から5のフェーズを繰り返す。
- (6) Deployment : 一般的な手順は再利用できるようにすることで、データマイニングプロセスに関するナレッジを蓄積する。

## 4. 評価

### 4.1 評価の方法

センサーデータを含む大量データを対象とした分析フレームワークであること、外れ値の自動検出ができること、という理由から、「ビジネスの理解を行えること(ビジネスの理解)」、「外れ値を含むデータに対応していること(外れ値)」、「フレームワークの途中でデータの構成変更による項目追加ができること(項目追加)」、「前処理とモデル生成フェーズの繰り返し時の追加作業量が少ないこと(反復性)」、「大量データに対応していること(大量データ)」の

5つの項目において、既存のフレームワークと提案フレームワークを比較する。

以下、評価項目の選定理由を述べる。[12]のデータを使い、ショッピングモール内の顧客の動向と各店舗の購買実績から、モール内で商品を購入してくれる可能性が高い顧客の傾向を分析し、ショッピングモール内のディスプレイにおすすめ商品を投影するプロジェクトを例とする。本プロジェクトは顧客の一般的な行動の分析を行いたいため、外れ値は除去して構わない。このような判断は、ビジネスの理解に起因するため、評価項目とした。顧客の動向は、3次元距離画像センサーで顧客の動きを収集できるが、センサーデータのため、計測値に外れ値や欠損値、ネットワークの状態により時系列が前後した値を含んでいる可能性がある。そのため、外れ値データへの対応を評価項目とした。また、設置場所・設置環境により、複数種類のセンサーが必要であるため、前処理として、データの形式を統一する必要がある。さらに、モデル生成のために、分析対象期間の顧客の動向と購買実績を結合し、店舗ごとの購買金額項の追加や集計値の算出など、フレームワークの途中でデータ構成変更、項目追加も重要な条件である。よって、項目追加を評価項目とした。分析モデル生成後、モデルの評価結果によりモデルで扱うデータに変更が生じた場合、データ生成フェーズに戻る必要があるが、再度モデルの評価を行うまでのフェーズが少ない方が、分析者の作業量が減るため、反復性を評価項目とした。最後に、本稿は大量データでの分析を想定しているため、大量データへの対応を評価項目とした。

### 4.2 評価結果

本稿で定めた評価項目における、フレームワークの比較を表1に示す。以下、各評価項目の理由を詳述する。

まず、ビジネスの理解について比較する。CRISP-DM, ASUM-DM, 提案フレームワークは顧客理解に関するフェーズが存在するため「○」である。KDDも達成する目標を定めるフェーズが存在するが、顧客の優先事項や達成条件までは定義されていないため「△」とする。SEMMAは、ビジネス理解に関する記述がないため「×」とする。

次に、外れ値について比較する。SEMMAはModifyフェーズで、統計学やクラスタリングによる外れ値検出ができるため「○」である。提案フレームワークは前処理フェーズの自動化対象フェーズで検出することができるため「○」である。KDDは、前処理フェーズでノイズが除去されてしまうことから、外れ値に関する分析に対応できないため「△」とする。CRISP-DMとASUM-DMは、外れ値に関する記述がないため「×」とする。

次に、項目追加について比較する。CRISP-DM, ASUM-DM, 提案フレームワークは、前処理フェーズで新規項目の追加ができるため「○」である。SEMMAはModifyフェー

表 1 フレームワーク比較

| 項目      | KDD | CRISP-DM | SEMMA | ASUM-DM | 提案フレームワーク |
|---------|-----|----------|-------|---------|-----------|
| ビジネスの理解 | △   | ○        | ×     | ○       | ○         |
| 外れ値     | △   | ×        | ○     | ×       | ○         |
| データ追加   | ×   | ○        | ○     | ○       | ○         |
| 反復性     | ○   | △        | ×     | △       | ○         |
| 大量データ   | △   | ○        | △     | ○       | ○         |

ズで新規データの追加が可能となっているため「○」である。KDDは、追加に関する記述がないため「×」とする。

次に、反復性について比較する。KDDは、任意の2つのフェーズを繰り返すことができ、前処理とモデル選定のためのデータ加工処理も分かれていることから、最小単位での繰り返し処理を組み込むことができるため「○」である。提案フレームワークは、前処理フェーズのうち、最終的に適したモデル生成のためのデータ加工フェーズに戻ることができ、繰り返し対象作業を限定できるため「○」である。CRISP-DM、ASUM-DMは、前処理として定めている範囲が広く、繰り返し時の作業量は他より多くなるため「△」とする。SEMMAは、データマイニングプロセス自体を繰り返し処理であると定めているものの、各フェーズの小さい繰り返し処理については記載がないため「×」とする。

最後に、大量データについて比較する。CRISP-DM、ASUM-DM、提案フレームワークは、データの理解フェーズで分析に必要なデータだけに絞ることができること、前処理フェーズで同一項目のマージ（データの統合）を行うことにより、複数のデータを扱えることから「○」である。KDD、SEMMAは、データセットのサンプリングフェーズは存在するが、複数のデータに関する記述がないため「△」とした。

#### 4.3 考察

5つの評価項目について評価した結果、提案フレームワーク以外、評価項目を全て満たしているフレームワークは存在しなかった。よって、センサーデータを含む大量データの分析において、提案フレームワークの有効性が評価できた。提案フレームワークにより、ビジネス領域でのデータマイニングプロセスにおいて、分析担当者が行っていた作業の一部を自動化でき、作業負荷を下げる事が可能である。

## 5. おわりに

本稿では、センサーデータを含む大量データの分析を対象とし、前処理で必要な処理を、外れ値など一般的に想定されるデータ傾向のための加工と対処とモデル生成のためのデータ加工に分類することで、自動化を組み込んだデータマイニングフレームワークを提案した。ショッピング

モールの分析プロジェクトを例とし、選定した評価項目について既存フレームワークと比較を行い、本稿で対象としている分析において、提案フレームワークは既存のフレームワークより、適していることを評価した。今後は、提案フレームワークの有効性を実データを用いて検証する予定である。

#### 参考文献

- [1] Azevedo, A. and Santos, M. F.: KDD, SEMMA and CRISP-DM: a parallel overview, *IADIS European Conference on Data Mining 2008*, pp. 182–185 (2008).
- [2] 山西健司: 異常検知: 外れ値検知と変化検知 (異常検知と変化点検出), *日本信頼性学会誌 信頼性*, Vol. 37, No. 3, pp. 134–143 (2015).
- [3] Hodge, V. J.: A Survey of Outlier Detection Methodologies, p. 44 (2004).
- [4] Qi, Z., Wang, H., Li, J. and Gao, H.: Impacts of Dirty Data: and Experimental Evaluation, *arXiv:1803.06071 [cs, stat]*.
- [5] Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P.: From Data Mining to Knowledge Discovery in Databases, *AI Magazine*, Vol. 17, No. 3, p. 37 (1996).
- [6] Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P.: The KDD Process for Extracting Useful Knowledge from Volumes of Data, *Commun. ACM*, Vol. 39, No. 11, pp. 27–34.
- [7] Cross Industry Standard Process for Data Mining Consortium: CRISP-DM by Smart Vision Europe, <http://crisp-dm.eu/reference-model/>.
- [8] SAS Institute: SAS Enterprise Miner, <https://web.archive.org/web/20120308165638/http://www.sas.com/offices/europe/uk/technologies/analytics/datamining/miner/semma.html>.
- [9] International Business Machines Corporation: Have you seen ASUM-DM? - SPSS Predictive Analytics, <https://developer.ibm.com/predictiveanalytics/2015/10/16/have-you-seen-asum-dm/>.
- [10] International Business Machines Corporation: Analytics Solutions Unified Method Implementations with Agile principles (2016).
- [11] Cross Industry Standard Process for Data Mining Consortium: CRISP-DM Methodology, <http://crisp-dm.eu/home/crisp-dm-methodology/>.
- [12] Brscic, D., Kanda, T., Ikeda, T. and Miyashita, T.: Person Tracking in Large Public Spaces Using 3-D Range Sensors, *IEEE Transactions on Human-Machine Systems*, Vol. 43, pp. 522–534.