

RFID データに基づいたアパレル商品の販売数予測

櫻井 茂明[†] 三部 雅法^{††} 渡辺 勝利^{†††}

[†] 株式会社 東芝 研究開発センター 〒212-8582 神奈川県川崎市幸区小向東芝町 1

^{††} 東芝テック株式会社 リテールソリューション事業本部 流通システム研究所 〒410-2392 静岡県伊豆の国市大仁 570

^{†††} 東芝テック株式会社 リテールソリューション事業本部 東京支社 〒141-8664 東京都品川区東五反田 2-17-2
E-mail: [†]shigeaki.sakurai@toshiba.co.jp, ^{††}{Masanori_Sanbe,Katsutoshi_Watanabe}@toshibatec.co.jp

あらまし 本論文では、アパレル店舗に設置された RFID リーダー及びその商品に付与された RFID タグから収集される RFID データに基づいて、商品の次週販売数を予測する方法を提案する。提案法においては、商品の試着回数、商品の棚からの取り上げ回数、当週販売数などを説明変数とし、次週販売数を被説明変数として、予測モデルを生成する。本論文では、2 箇所のアパレル店舗から収集された RFID データに基づいて予測モデルを生成し、当該モデルの予測性能を評価する。

キーワード SVM、RFID、販売数予測、アパレル店舗

Prediction of Sales Volume of Apparel Items based on RFID Data

Shigeaki SAKURAI[†], Masanori SANBE^{††}, and Katsutoshi WATANABE^{†††}

[†] Corporate Research & Development Center, Toshiba Corporation 1, Komukai-Toshiba-cho, Saiwai-ku, Kawasaki, Kanagawa, 212-8582, Japan

^{††} Distribution System Laboratory, Retail Solution Business Group, Toshiba TEC Corporation 570, Ohhito, Izunokuni, Shizuoka, 410-2392, Japan

^{†††} Tokyo Main Branch, Retail Solution Business Group, Toshiba TEC Corporation 2-17-2, Higashi-gotanda, Shinagawa, Tokyo, 141-8664, Japan

E-mail: [†]shigeaki.sakurai@toshiba.co.jp, ^{††}{Masanori_Sanbe,Katsutoshi_Watanabe}@toshibatec.co.jp

Abstract This paper proposes a new method that predicts sales volume of apparel items in the next week based on RFID data. The RFID data is collected from RFID readers set in apparel shops and RFID tags attached to items. The method regards the number fitting the items, the number picking up them from shelves, the sales volume of the items in this week, and so on as independent variables. It regards the sales volume of the items in the next week as a dependent variable. Also, it creates a prediction model which explains a relationship between the independent variables and the dependent variable. This paper applies the method to the RFID data collected from two apparel shops and evaluates the prediction efficiency of the created models.

Key words SVM, RFID, Prediction of Sales Volume, Apparel Shop

1. はじめに

各種センサーや RFID(Radio Frequency IDentification) タグ及び RFID リーダーの小型化、低価格化を背景として、様々な空間及び物体に対して、各種センサーや RFID リーダー及び RFID タグを埋め込むことが可能となっている。このようなユビキタスネットワーク環境の出現に伴って、あらゆる環境から大量のデータを収集、蓄積することが可能となってきた。これらのデータの中には、利用者の行動をタイムリーに予測し

て支援したり、様々な企業活動を実施する上での意思決定を支援したりすることが可能である有用な知見が埋もれている。このため、これらデータに効率的にアクセスし、分析することのニーズが高まっており [8] [11]、これらデータの分析法が活発に研究されている。

例えば、Akaoshi et al. [1] では、ユビキタスネットワークにおいて、関係データベースの枠組みを利用することにより、動的に設置される機器からデータを収集し、当該データにアクセスする方法を提案している。Dass and Mahanti [5] は、幅優

先探索と深さ優先探索を組み合わせることにより、実時間で頻出するパターンを発見する方法を提案している。また、Ihler et al. [7] では、ポアソン分布に基づいた、異常イベントの説明モデルを提案しており、当該モデルのパラメータを時系列データから統計的に推定する方法を提案している。Kuramitsu [10] では、各種センサーから送られてくるデータストリームから、周期性を発見する方法及び、異常値を予測する方法を提案している。Teng and Lin [19] では、動作を表すイベントと時間間隔を同時に含む時間動作パターンを、木構造表現を利用することにより、効率的に発見する方法を提案している。この他、ユビキタスネットワーク環境から得られるデータに限定してはいないものの、時系列データから特徴的な時系列パターンを効率的に発見する方法を、Sakurai et al. [15] [16] [18] は提案している。

一方、RFID リーダー及び RFID タグの実応用としては、アメリカに本部を置く世界最大のスーパーマーケットチェーン Wal-Mart [13]、ヨーロッパを中心とした小売り大手企業であるドイツ・Metro グループ [20]、イギリス最大の小売業チェーンである Marks & Spencer [4] などによる導入事例が知られている。また、国内においても、アパレル業界向けや出版業界向けの RFID リーダー及び RFID タグの実証実験が実施されている [12] [14]。また、三越が自主運営するインポートカジュアルのセレクトショップでは、RFID リーダー及び RFID タグの実運用もなされている [9]。

このような RFID リーダー及び RFID タグの普及により、在庫管理などに対しては、その有効性がある程度検証されている。しかしながら、これら応用だけでは、RFID リーダー及び RFID タグから得られるデータ (RFID データ) を、必ずしも十分に活用しているとはいえない。これらデータを、より深く分析することにより、より付加価値の高いサービスを実現することが求められているものの、どのような分析が必要であるかは、対象とする業界にも依存する問題であり、なかなか有効な分析法が確立されていないのが現状である。

このような背景の下、櫻井ら [17] では、アパレル店舗に設置された RFID リーダー及びその商品に付与された RFID タグから収集されるデータの分析に焦点を絞ることにより、その効果的な分析法を検討した。また、実際のデータに基づいた分析を実施し、データ分析の可能性を検証した。しかしながら、本論文で実施された分析は、商品の移動に伴う回数やその時間を積算し、順位付けされた結果を視覚的に表示するといった、基本的な分析にしか過ぎず、必ずしも高度な分析とはなっていなかった。そこで、本論文では、アパレル業界関係者のニーズを踏まえつつ、より高度な分析として、収集されたデータに基づいて、商品の販売数を予測する方法を検討する。また、2 店舗から収集された RFID データに基づいて予測モデルを生成し、当該モデルにおける予測性能を評価する。

2. 販売数予測

当期における在庫数と次期における販売数が与えられるとすれば、余裕在庫をどのくらい確保するかといった問題があるとしても、当期における発注数がある程度予測することが可能で

ある。一方、当期における在庫数は、RFID リーダー及び RFID タグに基づいた在庫管理によって管理されており、システムに保管されている。このため、次期における販売数を予測することができれば、当期における発注数を予測することができる。このような販売数の予測が可能になれば、多数の在庫を抱える必要がなくなるとともに、欠品による機会損失や売れ残り商品の安値販売を避けることができる。アパレル店舗の経営者にとって、価値ある RFID システムを提供することができる。そこで、以下においては、次期における販売数を RFID データに基づいて予測することを考えることにする。

次期における販売数は、商品の特性、季節の特性、顧客の特性、店舗の特性、イベントなどの外部要因の特性といった様々な影響を受けて決定されている。現段階では、どのような特性が、次期における販売数に大きな影響を与えているかは明らかではない。従って、現段階で利用可能な特性により、どの程度まで次期における販売数を予測可能か検討していく必要がある。現在のところ、アパレル店舗からは、店舗番号、商品の品番、カラー、サイズ、スマートシェルフ (RFID リーダーが設置されている棚) からの商品の取り上げ回数 (ピックアップ数) 及び取り上げ時間 (ピックアップ時間)、スマートシェルフの番号、試着室への商品の持ち込み回数 (試着数) 及び持ち込み時間 (試着時間)、試着室の番号、顧客の来客数、商品の在庫数、商品の販売数といったデータを収集することができる。外部の情報源にアクセスすることにより、販売数に影響を与えそうな気象情報や周辺のイベント情報などの情報も利用可能とは考えられるものの、外部の情報がどの程度、販売数予測に有効であるかは未知数である。一方、情報収集の枠組みを開発するには、多大なる労力が必要となる。そこで、本論文では、収集されているデータの範囲で、まずは、次期における販売数の予測を考えることにする。

アパレル店舗の商品の場合、コンビニエンスストアなどにおける日販品の販売とは異なり、特定の商品が 1 日に何十個も売れるといったことはない。このため、日別の販売を考えた場合には、ほとんどの商品が売れないことになる。従って、売れないと判断するだけで、かなり高い精度の予測が行えることになり、売れる商品を予測することは難しいと考えられる。そこで、本論文では、週をひとつの単位とする一方で、類似の商品をある程度まとめたものを単位として、販売数の予測を試みることにする。具体的には、週を単位とした店舗別のデータ及び週を単位とした商品分類別のデータに基づいて、顧客の来客数、商品の在庫数、当期における商品の販売数などを説明変数、次期における販売数を被説明変数として、説明変数と被説明変数の組からなる学習データを生成する。また、当該データを既存の予測モデルの学習法に適用することにより、説明変数に基づいて被説明変数を説明する予測モデルを学習する。

図 1 は本枠組みに沿った発注予測方式を示している。本方式では、学習データから生成した予測モデルに対して、説明変数の値だけからなる評価データを適用することにより、評価データに対応する次期における販売数を予測することができる。また、予測された次期における販売数と当期における在庫数から、

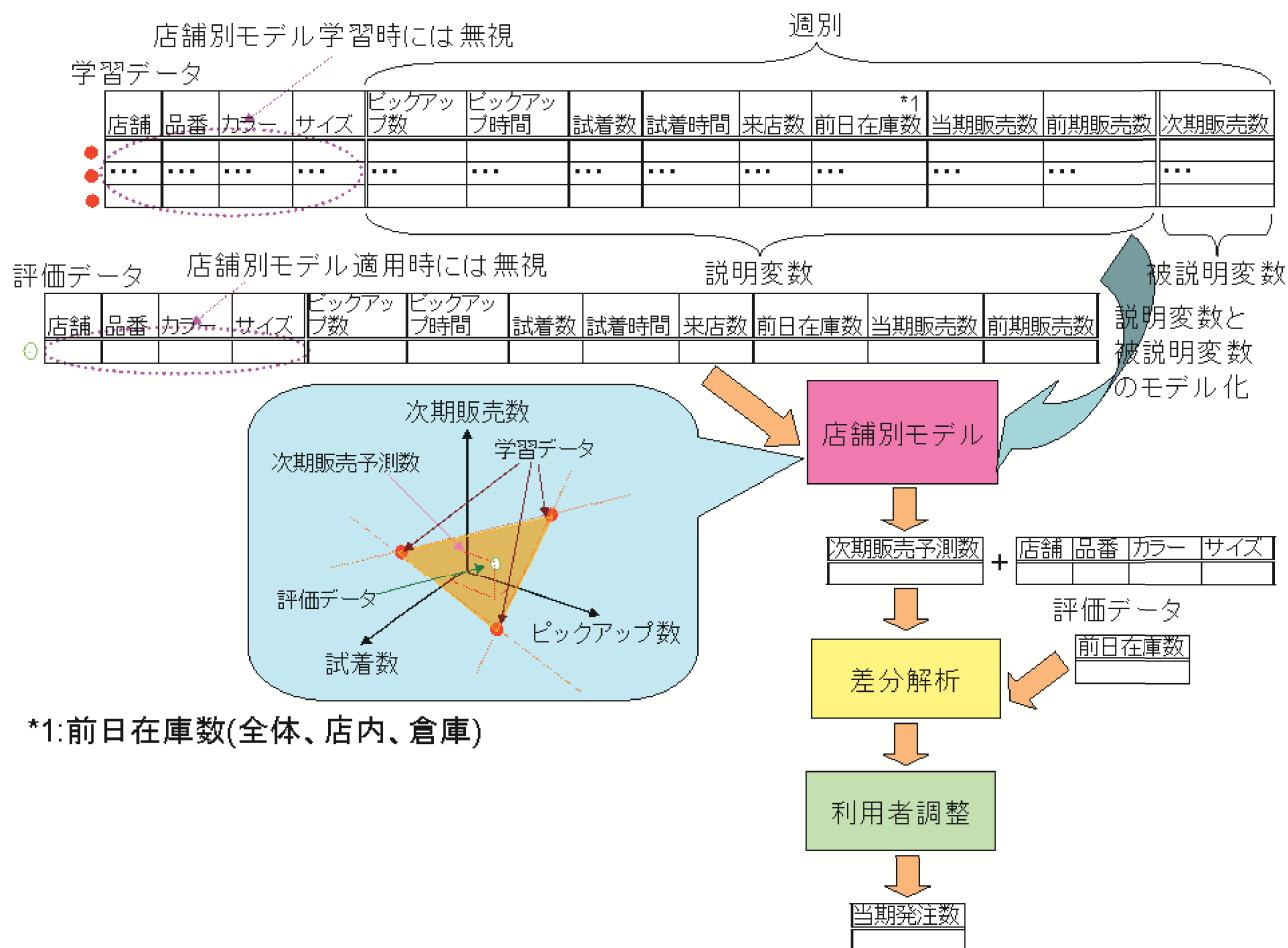


図 1 発注数予測方式

当期における発注数を予測することができる。

3. 実験データの生成

現在、予測モデルの学習に利用可能なデータとしては、表 1 にその収集期間を示す、3 種類のデータが存在している。しかしながら、B 社データに関しては、結果を公開することができないため、以下においては、A 社データに関してのみ扱っていくことにする。

データ名	収集期間
A 社: α 店	2008/11/25(火)~2009/03/25(水)
A 社: β 店	2008/11/25(火)~2009/03/25(水)
B 社: γ 店	2008/09/18(木)~2008/12/18(木)

表 1 データ収集期間

ただし、本データの場合、システムの不備によって一部期間のデータが欠落している。また、本期間の月曜日と木曜日をトリガーとしたデータの収集が行われている。

一方、各データにおいては、店舗に設置された RFID リーダーの違い、その他センサーの違い、データ収集法の違いにより、表 2 に示すように、利用可能なデータに若干の差異が生じている。表 2 においては、「○」がデータの収集を表しており、「×」がデータの未収集を表している。これらデータでは、来客数以外のデータは、品番に対応してデータの収集が行われてい

る。これに対し、来客数のデータは、各店舗単において 15 分単位でデータの収集が行われている。このため、各店舗の各商品に対しては、同一の値が付与されることになる。なお、今回の実験では、どのデータが予測モデルの学習に効果があるかが不明であるため、各データに与えられているすべてのデータを利用して、学習データを生成することにする。

本データに対して、各品番ごとに週を単位としてデータを積算し、学習データの生成を行う。ただし、在庫データに関しては、データ収集のトリガーとなった月曜と木曜以外のデータは収集されていないため、各日における各品番の在庫数を以下に示す基準で決定することにする。また、このようにして決定した各日の値を、該当週単位で積算した値を該当週における在庫データとする。

(1) 該当日がデータ収集日であり、該当日における在庫が記録されている場合には、当該データを在庫データとする。

(2) 該当日における在庫が記録されていない(データ収集日でない場合及びデータ収集日においてデータが欠損している場合を含む)場合には、該当日を起点とした前後において、在庫データが存在する日を探査する。

(a) 前後の在庫データが共に探索される場合には、該当日に対して、探索された値を利用して、日単位に線形補間した値を算出し、該当日の在庫データとする。

(b) その他の場合には、該当日における在庫データを 0 と

	品番	ピックアップ数	ピックアップ時間	試着回数	試着時間	来客数	在庫数 (全体)	在庫数 (店内)	在庫数 (倉庫)	販売数
A 社: α 店										
A 社: β 店		×	×							

表 2 収集対象データ

する。

このようにして週単位に積算したデータに対して、在庫データが存在している期間を品番ごとに算出し、当該期間におけるデータを学習データとして利用する。また、販売数に関しては、対象とする週を起点として、その前の週における販売数と、対象とする週の販売数を説明変数に加えることにし、対象とする週の翌週の販売数を被説明変数とみなすことにする。

4. 数値実験

4.1 SVM の利用

本節における数値実験では、データ間の非線形性の存在を考慮して、SVM(Support Vector Machine) を利用して、予測モデルを学習した実験を主に行うことにする。また、比較のために統計手法のひとつである回帰分析を利用した実験も行うことにする。

元来 SVM は、線形分離可能な分類問題を対象とする学習法であった。本 SVM は、カーネル関数を用いて特徴空間に対象を写像することにより、非線形な分類問題においても高い分類性能を発揮することが示され[2]、その利用が急速に高まった学習法である。また、SVM は、回帰問題を対象とするために、SVR(Support Vector Regerssion) [6] といった拡張がなされている。このような SVM は、多くのフリーソフトが提供されており、今回の実験では、そのうちのひとつである LIBSVM [3] を実験では利用する。LIBSVM では、5 種類の SVM に関する実装がなされているが、そのうち、標準的な分類問題向けの C-SVC と回帰問題向けの epsilon-SVR を利用し、カーネル関数としては、LIBSVM でデフォルトの関数として指定されている動径基底関数 (Radical Basis Function) を利用する。

一方、比較実験で利用する回帰分析としては、MicrosoftTM Office Excel 2003 に搭載されている回帰分析機能を利用する。

4.2 実験方法

アパレル店舗の商品の場合、個々の商品でみれば、その販売数は必ずしも多くない。1 週間及び類似の商品をまとめたものを単位とする学習データを生成してはいるものの、全く販売されなかった商品が比較的多く存在している。このため、被説明変数である次週販売数が 0 となっている学習データが多く、次週販売数が 0 より多くなっている学習データが少なくなっている。学習データにこのような偏りが存在する場合、学習される予測モデルも偏ったものになりがちである。実際、予備実験で利用したニューラルネットワークの場合では、その傾向が顕著に現れたために、すべての評価データに対して、次週販売数を 0 に予測する予測モデルが生成されていた。SVM においては、このような傾向がやや小さかったものの、やはり学習データの偏りの影響はあるものと考えられる。そこで、本論文では、こ

の偏りの影響を回避するために、分類問題向けの C-SVM と回帰問題向けの epsilon-SVR を組み合わせた方法の効果を検証する。具体的には、学習データの被説明変数である次週販売数が 0 のものを 0、0 より大きなものを 1 とするクラスを考え、2 分類問題とした場合の予測モデルを最初に学習する。次に、クラスが 1 である学習データだけを抽出し、epsilon-SVR を利用して、予測モデルの学習を行う。以下においては、本学習法を 2 段階 1 法と呼ぶことにする。

このような 2 段階の学習によって学習された予測モデルを利用した予測では、C-SVM による 2 分類の予測モデルを用いて評価データを最初に評価し、クラスが 0 になると予測されたものに対して、0 を予測結果として出力する。これに対して、クラスが 1 になると予測されたものに対しては、epsilon-SVR による予測モデルを適用し、評価データに対応する推定値を算出し、その値を予測結果として出力する。

上記に説明した 2 段階 1 法に加えて、全学習データに対して epsilon-SVR を適用した方法 (以下、単純法)、2 段階の学習ではあるが、2 段階目の予測モデルの学習の際に全学習データを適用した方法 (以下、2 段階 2 法) を用いて評価する。

この他、C-SVM における特定クラスに対する重みの付与機能を利用して、学習データに占めるクラスの割合を考慮した重みを付与することにする。具体的には、クラス 1 となる学習データとクラス 0 となる学習データの件数を算出して、クラス 0 となる学習データを基準とする重みを求め、クラス 1 に対する重みとして、C-SVM に与えることにする。このような重みを考慮することにより、予測モデルの偏りがある程度補正することが期待できる。

なお、以下の実験では、2009/01/25(日) を起点として生成した学習データを数値実験に利用している。本日時を起点としたのは、データ収集の起点となる日を変えて、回帰分析に基づいて予測モデルを学習する予備実験を実施した際に、当該日を起点とした場合における、説明変数と被説明変数との間の相関が最も高くなったためである。

4.3 実験結果

図 2 及び図 3 に、クラスに変換した学習データにおけるデータが分布している様子に対応する実験結果を示す。図においては、図 2 が、 α 店の結果を示しており、図 3 が、 β 店の結果を示している。

各図におけるグラフ (a) は、クラスが 0 となる学習データ (非販売) 及びクラスが 1 となる学習データ (販売) の件数を、中分類ごとに示したグラフである。ここで、中分類とは、学習データの生成単位となった商品の分類をその特性に従ってある程度まとめた、より上位のコード体系である。一方、各図におけるグラフ (b)～グラフ (g) の結果においては、中分類に基づいた

学習データの分割を実施せずに、すべての学習データを利用した場合における結果が「全体」として示されている。また、中分類ごとに学習データを分割した場合における、各中分類に対応した結果が、中分類ごとに示されている。

ここで、これらグラフをより詳細に見ていくと、グラフ (b) においては、全学習データを利用して、2 分類予測モデルを学習した場合における結果として、学習した予測モデルで学習データを再評価した際の、正解率、販売正解率、非販売正解率、及び販売正解率と非販売正解率の相乗平均 (以下、単に相乗平均) を、学習データの全体及び、商品分類ごとに算出した結果が示されている。ただし、正解率、販売正解率、非販売正解率は、式 (1)～式 (3) によって定義されるとする。これら正解率は参考のために記載しているが、予測モデルをバランス良く評価するといった観点からは、相乗平均がもっとも妥当な評価値と考えられる。なお、以下の式においては、学習データの集合が S 、予測と一致した学習データの集合が $S_{correct}$ 、販売された学習データの集合が S_{sales} で表されている。また、 $|\cdot|$ は集合に含まれる要素の数を算出する演算であるとする。

$$\text{正解率} = \frac{|S_{correct}|}{|S|} \quad (1)$$

$$\text{販売正解率} = \frac{|S_{correct} \cap S_{sales}|}{|S_{sales}|} \quad (2)$$

$$\text{非販売正解率} = \frac{|S_{correct} \cap (S - S_{sales})|}{|S - S_{sales}|} \quad (3)$$

また、グラフ (d) 及びグラフ (f) においては、全学習データを学習時に利用する代わりに、5 交差検定によって、学習データと評価データを分けて評価した場合における、2 分類予測モデルの結果が示されている。ただし、グラフ (d) が学習時における重みの利用が無い場合の結果、グラフ (f) が学習時における重みの利用がある場合の結果に対応している。なお、 n 交差検定においては、与えられた学習データを n 個に分割し、 $n-1$ 個の学習データに基づいて、予測モデルを学習し、残りの学習データで予測モデルを評価している。また、このような評価を、評価データを変えることにより、 n 回繰り返している。本 n 交差検定を利用することにより、未知データに対する予測モデルの性能を、ある程度評価することができる。

加えて、グラフ (c)、グラフ (e)、グラフ (g) においては、学習データにおける次週販売数と予測モデルによって算出された予測次週販売数との差の絶対値の和を、学習データ件数で割った値が示されている。ただし、回帰分析及び SVR において算出される予測次週販売数は実数値として与えられるため、算出された値を四捨五入して整数値化した値を予測次週販売数としている。また、グラフ (c)、グラフ (e)、グラフ (g) においては、それぞれ、全学習データを利用して予測モデルを学習した場合、5 交差検定かつ 2 分類学習時において重み無しで予測モデルを学習した場合、5 交差検定かつ 2 分類学習時において重みありで予測モデルを学習した場合の結果が示されている。

4.4 考 察

本節では、提案法の効果を 5 つの観点から考察してみることとする。

モデルの妥当性: 回帰分析による結果と比較してみると、SVM による結果の方が平均絶対誤差は小さくなっている。以下の考察において、平均絶対誤差があまり良い評価値になっていないことを指摘しているが、回帰分析と SVM との比較においては、同一の学習データを利用した場合の比較になっているため、平均絶対誤差に内在する問題点はあまり表れないと考えられる。このため、SVM による予測モデルの学習の方が、回帰分析による予測モデルの学習よりも妥当であると考えられることができる。

交差検定による結果に着目してみると、平均絶対誤差としては、各データにおいて、1.0 程度の誤差となっており、現在のデータは次週販売数の予測に全く役に立たないと言っていることはできない。しかしながら、2 分類予測モデルにおける性能が高くても 0.7 を超える程度であり、改良の余地はかなり残されているといえる。

ここで、学習データをすべて利用した場合と、5 交差検定によって学習した場合の結果に着目してみると、2 分類の予測モデルで、その差に大きな違いがある。すなわち、5 交差検定による結果がすべての学習データを利用した場合よりもかなり悪くなっている。今回の実験においては、重み以外のパラメータに関しては、デフォルトパラメータを利用しているが、予測モデルの学習に利用可能なカーネル関数やそのパラメータの選択には、調整の余地が残されていると考えられる。

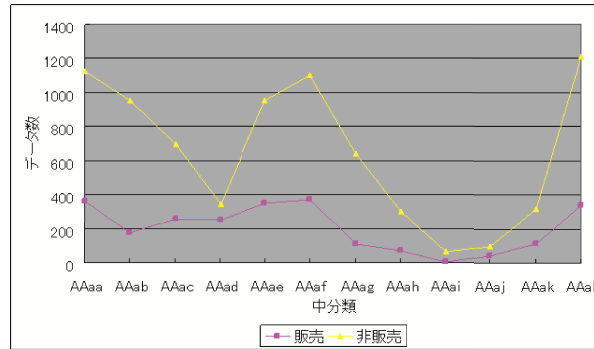
2 段階法の効果: 平均絶対誤差の観点からは、2 段階法の効果はそれ程観測されておらず、2 段階 1 法が最も悪く、2 段階 2 法が単純法よりも若干性能がよいといったレベルである。現状のデータの場合、2 分類予測モデルの分類性能がそれ程高くなっていないため、この結果が平均絶対誤差の悪化に影響したことがそのひとつの原因と考えられる。

また、2 段階法の場合、次週販売数が存在するデータを優遇することになるため、単純な絶対誤差の和の平均では、販売数が 0 となるデータの数の影響から、次週販売数を小さく見積もった方が平均絶対誤差が小さくなる傾向にある。このため、2 段階法における平均絶対誤差が大きくなったとも考えられる。

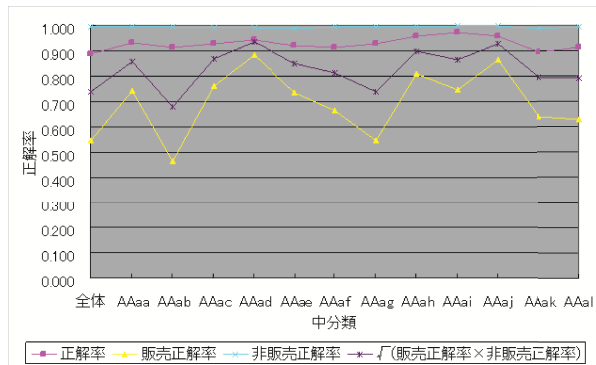
2 段階法の妥当性をより正確に評価する上では、2 分類予測モデルの分類モデルの性能向上を図るとともに、次週販売数における値の分布を考慮した評価基準を検討することも必要であると考えられる。

重みの効果: 2 分類予測モデルにおける相乗平均に着目してみると、重みを付与した学習を行った方が、全般的には、バランスの良い 2 分類予測モデルを学習可能であることを観測することができる。学習データに占めるクラスの偏りを、重みの調整によってある程度緩和できたと考えられる。しかしながら、この差は平均絶対誤差の違いに影響を与える程にはなっておらず、重みの違いによる平均絶対誤差の違いは明確には観測されてはいない。2 分類予測モデルの性能自体があまり高くなかったため、平均絶対誤差の違いに影響を与える程には、改善されなかったことがその原因のひとつと考えられる。

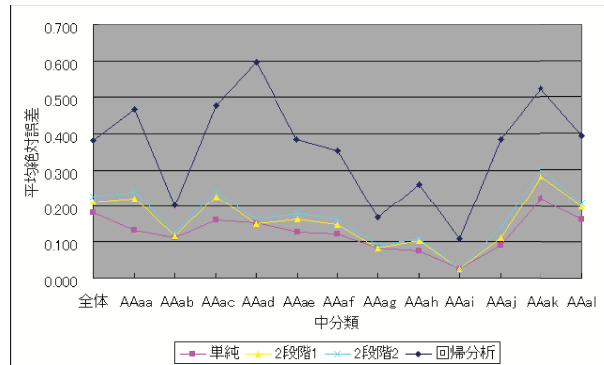
店舗の影響: β 店における相関係数の方が全般的に、 α 店における相関係数よりも高くなっていたが、2 分類の予測モデルに



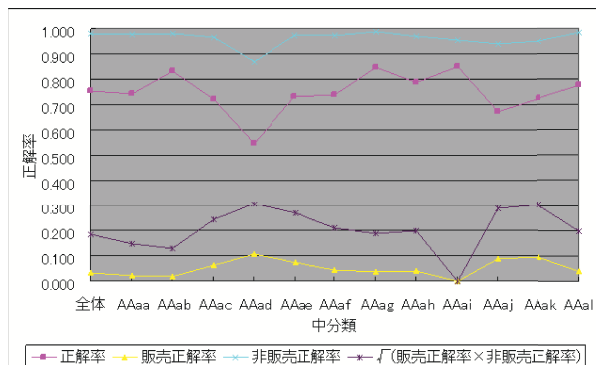
(a) 販売数の分布



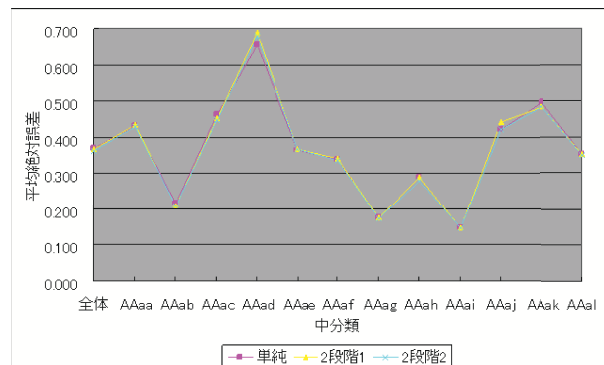
(b) 2分類 (全データ)



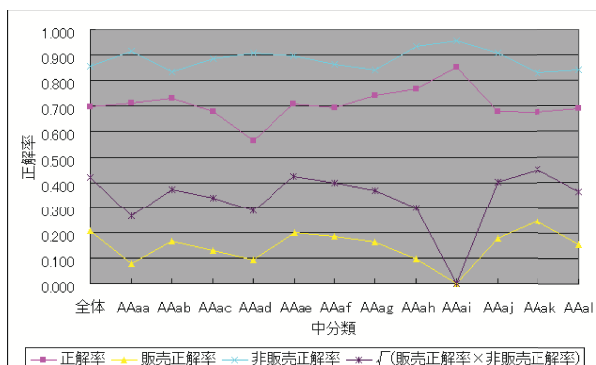
(c) 回帰 (全データ)



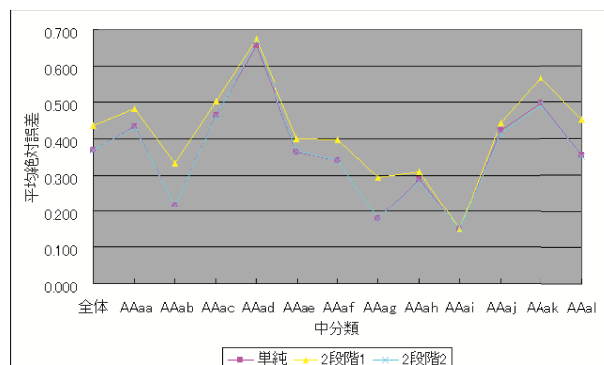
(d) 2分類 (5 交差検定: 重み無し)



(e) 回帰 (5 交差検定: 重み無し)



(f) 2分類 (5 交差検定: 重みあり)

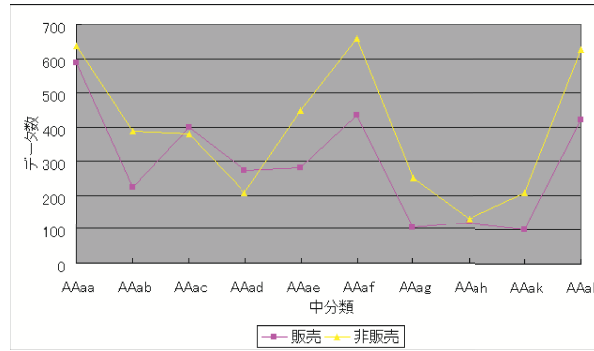


(g) 回帰 (5 交差検定: 重みあり)

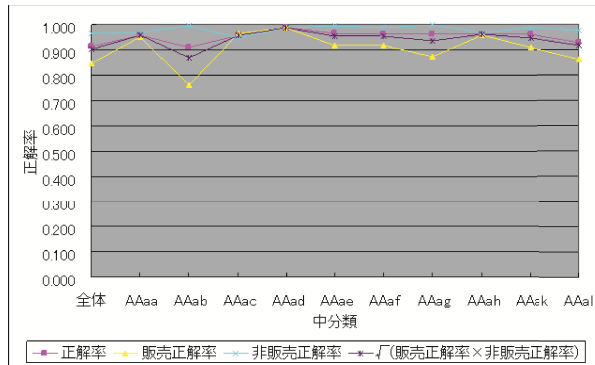
図2 A社: α 店の結果

においても、同様な効果が観測されており、 β 店における2分類予測モデルの性能が高いものになっている。また、店舗によって学習される予測モデルには大きな違いがあるようで、紙

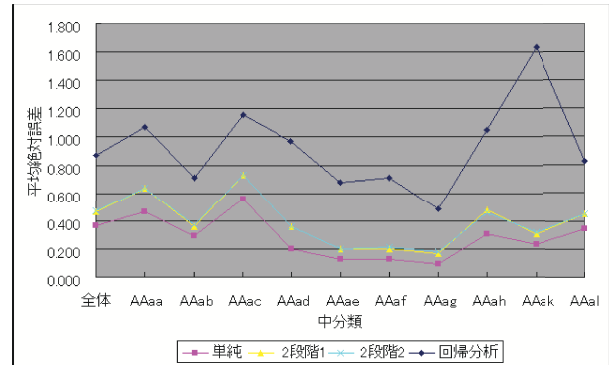
面の関係で割愛されているが、店舗を無視して学習データを統合した場合における、2分類の予測モデルは、 β 店における結果と α 店における結果の中間的な結果となっている。一方、



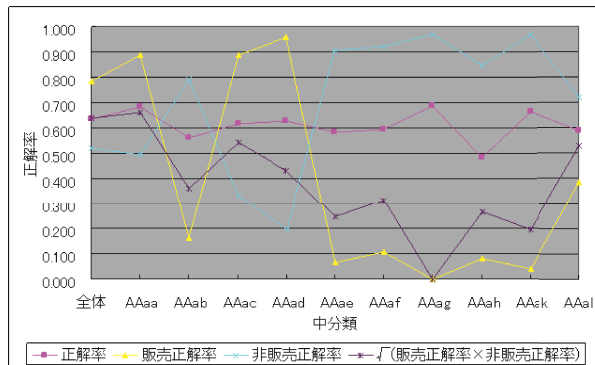
(a) 販売数の分布



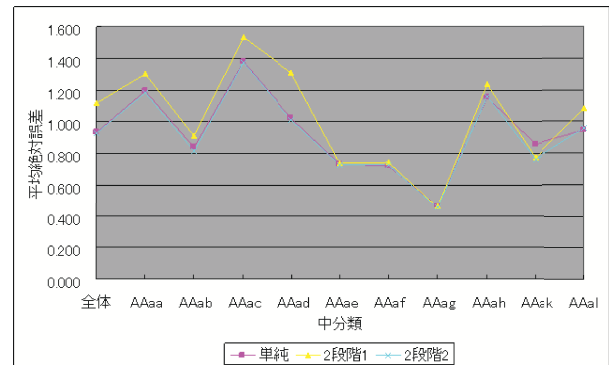
(b) 2 分類 (全データ)



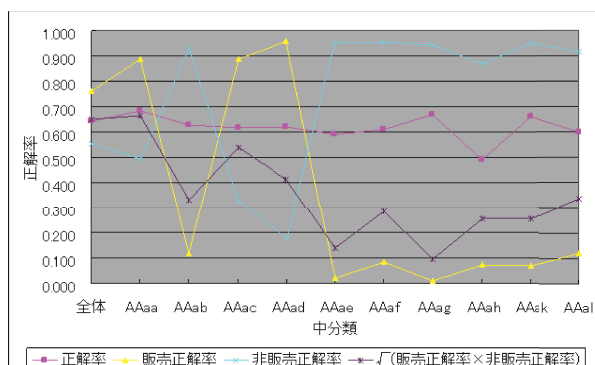
(c) 回帰 (全データ)



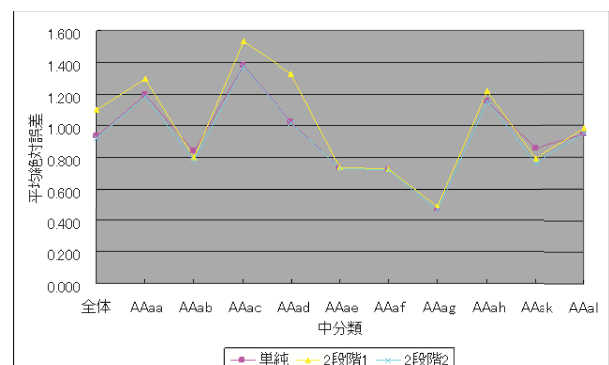
(d) 2 分類 (5 交差検定: 重み無し)



(e) 回帰 (5 交差検定: 重み無し)



(f) 2 分類 (5 交差検定: 重みあり)



(g) 回帰 (5 交差検定: 重みあり)

図 3 A 社: β 店の結果

平均絶対誤差に関しては、 α 店の方が、全般的に小さな値となっている。この原因としては、 β 店における商品の販売数が、 α 店における商品の販売数よりも、かなり大きくなっているこ

とが考えられる。すなわち、学習データに占める販売数が存在するデータの割合が多くなる程、次週販売数と予測した次週販売数の差が大きくなる可能性は高くなると考えられる。この影

響によって、 β 店における平均絶対誤差が上昇したものと考えられる。先の 2 段階法の効果のところでも考察したが、販売数のデータ分布を考慮した評価法を検討することも必要と考えられる。

商品分類の影響: 2 分類予測モデルに着目してみると、学習データを商品分類ごとに分けずに全体として学習した場合の性能が全般的に高くなっており、特定の商品分類において、性能の良い予測モデルが学習されるといった現象は、それ程観測されてはいない。商品分類による予測モデルの差異は、非線形性を考慮したモデルで吸収できる程度の差異でしかなく、学習データの減少が、商品分類ごとの予測モデルにおける性能の劣化に繋がったことがその原因と考えられる。

5. まとめと今後の課題

本論文では、現在利用可能な RFID リーダー及び RFID タグに関するデータに基づいて、商品の販売数を予測する方式の検討を行った。また、提案法における商品販売数の予測の可能性を、ふたつの店舗から収集された RFID データを用いて評価した。その結果、現在の学習データからは十分な予測が実施できる予測モデルが学習できると断言できる程には、その予測性能は高まってはいないものの、アパレル商品の販売数の予測の可能性を示唆することができた。

より一層の精度の向上に向けては、現在のデータ収集項目での精度を高めることや、より多くの学習データを収集することも必要と考えられるが、アパレル関係者に対するヒアリングを通して、次週販売数の予測に必要な属性が不足していないかを確認することも必要と考えられる。このような学習データの改良はとすると、アパレル店舗からの新たなデータ収集が必要となり、必ずしもすぐに実施することは困難な問題もはらんでおり、直近での対応は難しいかもしれない。一方、気象条件や外部のイベント情報などの、販売数に影響を与えられ、外部のデータベースから収集可能なデータの利用も検討していく必要があると考えている。このような学習データの改良により、アパレル店舗から新たな学習データを収集しなくても、より性能の高い予測モデルを構築する可能性は、十分に残されていると考えられる。

加えて、SVM に基づいた学習といった立場では、利用するカーネル関数やそのパラメータの選定に調整の余地が残されている。また、ひとつの学習データから複数の予測モデルを学習するアンサンブル学習の技術を導入することにより、より精度の高い予測モデルを構築する可能性も残されている。これら改良によって、より精度の高い予測モデルの学習を目指していきたい。

文 献

[1] Yuhei Akahoshi, Yutaka Kidawara, and Katsumi Tanaka, “A Database-Oriented Wrapper for Ubiquitous Data Acquisition/Access Environments”, Proc. of the Second Intl. Conf. on Ubiquitous Information Management and Communication, pp. 25-32, 2008.

[2] Bernhard E. Boser, Isabelle Guyon, and Vladimir Vapnik, “A Training Algorithm for Optimal Margin Classifiers”, Proc. of the Fifth Annual ACM Conference on Computational Learning Theory, pp. 144-152, 1992.

[3] Chih-Chung Chang and Chih-Jen Lin, “LIBSVM – A Library for Support Vector Machines”, <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.

[4] Jonathan Collins, “Marks & Spencer Expands RFID Trial”, RFID J., <http://www.rfidjournal.com/article/articleview/791/1/1/>, 2004.

[5] Rajanish Dass and Ambuj Mahanti, “Implementing BDFS(b) with Diff-Sets for Real-Time Frequent Pattern Mining in Dense Datasets – First Findings”, Proc. of the 2005 Intl. Workshop on Ubiquitous Data Management, pp. 113-120, 2005.

[6] Harris Drucker, Christopher J. C. Burges, Linda Kaufman, Alex J. Smola, Vladimir Vapnik, “Support Vector Regression Machines”, NIPS, pp. 155-161, 1997.

[7] Alexander Ihler, Jon Hutchins, and Padhraic Smyth, “Adaptive Event Detection with Time-Varying Poisson Processes”, Proc. of the Twelfth ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining, pp. 207-216, 2006.

[8] 今井 浩, “コンテンツからのデータマイニング”, 日本機械学会誌, vol. 109, no. 1056, pp. 8-10, 2006.

[9] 木實 新一, “三越の事例: 小売のための RFID インタラクションデザイン”, <http://pingmag.jp/J/2006/12/15/mitsukoshi-case-interaction-design-for-rfid-retail/>, 2006.

[10] Kimio Kuramitsu, “Finding Periodic Outliers over a Monogenetic Event Stream”, Proc. of the 2005 Intl. Workshop on Ubiquitous Data Management, pp. 97-104, 2005.

[11] 仲瀬 明彦, “ユビキタスデータからの知識発見に向けて”, 東芝レビュー, vol. 59, no. 8, pp. 52-53, 2004.

[12] 野村 昌弘, 山内 秀樹, 三部 雅法, “「実証実験で明らかになった UHF 帯電子タグを使った個品管理」セミナー”, <http://www.smrj.go.jp/keiei/seni/qr/kn19/021566.html>, 2007.

[13] Mark Roberti, “Wal-Mart Begins RFID Rollout”, RFID J., <http://www.rfidjournal.com/article/articleview/926/1/1/>, 2004.

[14] 流通システム開発センター, “EPC RFID システム導入における検討事項調査報告書 – 電子タグ導入における先進事例研究 –”, 流通システム開発センター, 2008.

[15] Shigeaki Sakurai, Youichi Kitahara, and Ryohei Orihara, “A Sequential Pattern Mining Method based on Sequential Interestingness”, Intl. J. of Computational Intelligence, vol. 4, no. 4, pp. 252-260, 2008.

[16] Shigeaki Sakurai, Youichi Kitahara, Ryohei Orihara, Koichiro Iwata, Nobuyoshi Honda, and Toshio Hayashi, “Discovery of Sequential Patterns Coinciding with Analysts’ Interests”, J. of Computers, vol. 3, no. 7, pp. 1-8, 2008.

[17] 櫻井 茂明, 三部 雅法, 渡辺 勝利, “アパレル店舗から収集された RFID データの分析”, 第 25 回ファジィシステムシンポジウム予稿集, 1A3-03, 2009.

[18] Shigeaki Sakurai, Ken Ueno, and Ryohei Orihara, “Discovery of Time Series Event Patterns based on Time Constraints from Textual Data”, Intl. J. of Computational Intelligence, vol. 4, no. 2, pp. 144-151, 2008.

[19] Vincent S. Tseng and Kawuu W. Lin, “Mining Temporal Moving Patterns in Object Tracking Sensor Networks”, Proc. of the 2005 Intl. Workshop on Ubiquitous Data Management, pp. 1-8, 2005.

[20] Rhea Wessel, “Metro Group’s Galeria Kaufhof Launches UHF Item-Level Pilot”, RFID J., <http://www.rfidjournal.com/article/articleview/3624/1/1/>, 2007.