

LVQ を用いたプロ野球の試合の観客数予測

繁村 崇[†] 黒木 進[‡] 北上 始[‡] 森 康真[‡]

[†] 広島市立大学情報科学部 〒731-3194 広島県広島市安佐南区大塚東 3-4-1

[‡] 広島市立大学大学院情報科学研究科 〒731-3194 広島県広島市安佐南区大塚東 3-4-1

E-mail: [†] shigemura@de.info.hiroshima-cu.ac.jp, [‡] {kuroki, kitakami, mori}@hiroshima-cu.ac.jp

あらまし 現在、観客数等の予測を行う手法として最も一般的なものは、重回帰分析である。しかしながら、重回帰分析では予測値と実測値の間に誤差が出てしまう。その誤差を小さくし、予測値を実測値に近づけるために、誤差の起こりやすい状況に着目し、重回帰分析とデータマイニングを組み合わせた研究が行われた。しかし、予測誤差に多少の改善は見られたが、大きな変化は見られなかった。

本研究では、試合の行われた状況に着目して観客数予測を行うことで、予測値を実測値に近づける方法について検討した。また、その予測結果を従来手法の結果と共に示し、予測精度についての比較を行った。

キーワード クラスタリング, データマイニング, LVQ

Forecasting the Number of Spectators of Professional Baseball Games using LVQ

Takashi SHIGEMURA[†] Susumu KUROKI[‡] Hajime KITAKAMI[‡] and Yasuma MORI[‡]

[†] Faculty of Information Sciences, Hiroshima City University

3-4-1 Ozukahigashi, Asaminami, Hiroshima, Hiroshima, 731-3194 Japan

[‡] Graduate School of Information Sciences, Hiroshima City University

3-4-1 Ozukahigashi, Asaminami, Hiroshima, Hiroshima, 731-3194 Japan

E-mail: [†] shigemura@de.info.hiroshima-cu.ac.jp, [‡] {kuroki, kitakami, mori}@hiroshima-cu.ac.jp

Abstract Multiple regression analysis is often used to forecast statistical data such as the number of spectators of professional baseball games. However, it is inevitable that there are differences between the forecasted values and samples. To minimize the differences, we have tried to find special occasions where the differences are tend to be large using data mining and then forecast values using multiple regression analysis. We have improved the differences, but there is still much room for improvement.

In this research, we have studied to forecast statistical values according to occasions classified by LVQ to improve precision of our forecast. And we also applied our forecast method based on LVQ to forecast the number of spectators of professional baseball games and evaluated our method.

Keyword Clustering, Data Mining, LVQ

1 はじめに

2006年、2009年に野球世界一決定戦 WBC (World Baseball Classic) が開催され、日本代表が2連覇という偉業を達成したことは、まだ記憶に新しい。また、セ・パ交流戦や、両リーグのプレーオフ実施等近年のプロ野球界は大きく変動し、衰退しつつあると言われている。プロ野球人気も、回復の兆しが見え始めたと思われる。

プロ野球界の変動の一つとして、2005年から各球団とも球場の観客数を実数に近い数字で発表し始めたこ

とが挙げられる。2004年以前にも観客数の発表は行われていたが、観客数の水増し等のため、信頼できない数値であった(日本プロ野球オフィシャルサイトにも2004年以前のデータは掲載されていない)。そのため、2005年からの実質的な観客数発表と言える。

したがって、2005年からの過去のデータを踏まえた上で、観客数を予測できないかと考えた。

現在、このような予測を行う手法として一般的であるのは重回帰分析である。また、重回帰分析と相関規則を組み合わせて予測を行うという研究が行われた。

本研究では、LVQを用いて、会場・天候等のプロ野球の試合の行われる状況、順位・スター選手等の試合を行う球団のチーム状況、観客数をクラスタリングすることにより、観客数を予測する方法を検討する。また、予測精度についても検討する。

2 扱うデータ

扱う標本として、観客数を実数発表になった、2005年の開幕戦から、2009年5月17日までのセ・リーグの試合を対象とした。なお、セ・パ交流戦、主催球団の本拠地以外の球場で行われた試合、月曜日に行われた試合については、データ数が少なく、有意な結果が得られないと思われるため、標本から除いている。また、広島で行われた試合についてであるが、広島は2009年から新球場となった。新球場効果もあり、旧本拠地球場である広島市民球場とは、観客数が例年とはまったく異なっている。そのため、2008年までの広島市民球場でのデータのみを扱っている。

1つの試合の行われた状況を表すものとして、その試合の観客数に影響を及ぼすと考えられる、試合の行われた曜日、試合会場、対戦相手、天候を球場ごとに収集した。また、チームの状況に影響を及ぼすものとして、ホームグラウンドチームの貯金の有無、平日休日、ホームグラウンドチームの順位、両チームのスター選手の人数合計、ホームグラウンドチームの連勝連敗（前日勝敗）についてもデータとして用いた。

曜日は、火曜日から日曜日までの6、球場は、東京ドーム、甲子園球場、神宮球場、ナゴヤドーム、広島市民球場、横浜スタジアムの6、対戦相手は自球団以外の5、天候は、晴れ、曇り、雨の3、貯金は有、無の2、休平日は休日、平日の2、順位は1・2位、3・4位、5・6位の3、スターは5人以下なら少、6～9人なら中、10人以上なら多の3、連勝連敗は前日負け、前日勝ちの2カテゴリーである。

標本の例（広島市民球場）を表1に示す。

表1：標本の例

	曜日	相手	天候	観客数
1	木	中日	晴れ	8114
2	金	阪神	晴れ	26263
3	土	阪神	雨	24807
4	日	阪神	雨	19550
5	火	ヤクルト	曇り	8013

貯金	休平日	順位	スター	前日
有	平日	3・4	中	●
有	休日	3・4	多	○
有	休日	1・2	多	●
有	休日	1・2	多	○
無	平日	3・4	中	●

なお、標本が質的データであるため、実際にはダミー変数を用いている。ダミー変数の例を表2に示す。

表2. ダミー変数の例（曜日）

	1	2	3	4	5
火	1	0	0	0	0
水	0	1	0	0	0
木	0	0	1	0	0
金	0	0	0	1	0
土	0	0	0	0	1
日	0	0	0	0	0

試合の行われた曜日・試合会場・対戦相手・観客数、前日勝敗のデータは、日本プロ野球オフィシャルサイト^[1]から、天候についてのデータは気象庁^[2]から得た。順位・貯金有無については、プロ野球貯金グラフ虎ノ軌跡2009^[3]を参考にした。また、スター選手についてはオールスターゲーム出場者の人数で判断した。なお、シーズン前後半の境界はオールスター戦開催日とし、シーズン前半は前年度のオールスター出場者数、後半は翌年度の出場者数を用いた。

3 関連研究

比較を行うため、関連研究^[4]について紹介しておく。

3.1 重回帰分析

観客数等の予測を行う手法として最も一般的なものである重回帰分析について説明する。

3.1.1 重回帰分析の定義

重回帰分析とは、従属変数（目的変数）と独立変数（説明変数）の関係を調べ、関係式を作成し、これを用いて次の事柄を明らかにする方法である。

- (1) 予測、潜在能力、評価
- (2) 従属変数と独立変数の相関関係
- (3) 説明変数の重要度

重回帰分析に適用できるデータは従属変数、独立変数ともに量的データであるが、ダミー変数を用いれば質的データも扱える。

回帰分析によって算出される式のことを回帰式という。1つの独立変数からなる回帰式を単回帰式、2つ以上の独立変数からなる式を重回帰式といい、以下のような式で表される。

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki} \quad (i = 1, 2, \dots, n)$$

\hat{Y}_i : 予測値

$\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$: 回帰係数

X_1, X_2, \dots, X_k : 独立変数

k : 説明変数の個数

また、実測値 Y と予測値 \hat{Y} の関係は以下のようなになる。

$$Y_i = \hat{Y}_i + \hat{U}_i$$

\hat{U}_i : 残差

重回帰式はこの残差（予測値と実測値の誤差）の2乗の和、つまり残差平方和が最小になるような式である。

3.1.2 決定係数

重回帰式から算出される予測値と実績値が近くなるほど、分析の精度が良いといえる。その精度は決定係数 R^2 として表され、以下のように表される。

$$R^2 = 1 - \frac{\sum_{i=1}^n \hat{U}_i^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

\bar{Y} : Y の平均

R^2 は $0 \leq R^2 \leq 1$ の数値で表され、1に近いほど回帰式によって従属変数が説明出来ているということになる。ただし、決定係数だけで考えたとき、独立変数を増やすほど1に近づくという欠点があり、これを調整したものが自由度修正済みである。自由度修正済み決定係数 R^{*2} は以下の式で表される。

$$R^{*2} = 1 - (1 - R^2) \frac{n-1}{n-k}$$

決定係数がいくつ以上あればよいのかという統計的基準は無いが、一般的には決定係数は0.8以上であれば精度が非常に良いといわれている。また、0.5以上は必要とされている。

3.1.3 重回帰方程式の検定

重回帰方程式の有意性の検定を行うには、回帰係数 X_1, X_2, \dots, X_k が Y を説明しないという帰無仮説を

$$\text{帰無仮説: } \beta_2 = \beta_3 = \dots = \beta_k = 0$$

として、 X_1, X_2, \dots, X_k のいずれか1つの変数が Y を説明できるという対立仮説を

対立仮説: $\beta_2, \beta_3, \dots, \beta_k$ の少なくとも1つが0でない。

として、有意水準（通常は0.05） α を定めて両側のF検定を行う。検定統計量Fは以下の式で表される。

$$F = \frac{\frac{1}{k-1} \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\frac{1}{n-k} \sum_{i=1}^n \hat{U}_i^2}$$

帰無仮説が正しい場合、Fは自由度(k-1, n-k)のF分布 $F(k-1, n-k)$ に従うので、有意水準 α とFのF分布 $F(k-1, n-k)$ における有意確率 p について比較を行う。 p が α 以下なら、重回帰方程式は有意であり、それ以外

は有意でない。

3.2 データマイニング

データマイニングとは、大容量のデータの中から規則性や法則性を発掘することである^[4]。データマイニングには、重回帰分析などの統計解析や、決定木（ディジション・ツリー）、ニューラルネットワーク、遺伝的アルゴリズムなど多くの手法があり、その中でアイテム（項目）間や、数値属性間の相関を探る、相関規則の抽出について焦点を当ててみた。

3.2.1 相関規則

例えば、

$$(\text{Balance} \in [v_1, v_2]) \Rightarrow (\text{CardLoan} = \text{yes})$$

というような式がある。これは、銀行における顧客のデータベースから、預金残高が v_1 から v_2 の間である顧客を条件として絞り込み、その前提条件を満たす客ならば、カードローンを使用するという意味の式である。ここで、条件Aを満たす顧客の数のことを、A支持数といい、条件Aを満たした顧客のうち条件Bも満たす顧客の割合のことを確信度という。現実の問題では、少なすぎる支持数では確信度が100%ということは何も有り得ないので、一定の確信度を満たす式を有効であると見なし、有効な相関規則とみなす。

3.2.2 前提条件の組み合わせと誤差

重回帰分析結果と実測値の間には誤差があり、特に誤差が多いと思われる場所や要素を特定できれば、その誤差を改善することが出来ると考えた。

独立変数AとBがあるとき、 $A \wedge B$ なら誤差Cは v_1 と v_2 の間にあるという式を以下のように表す。

$$(A \wedge B) \Rightarrow (C \in [v_1, v_2])$$

このとき、前提条件に当てはまる標本数が指定した最小支持数を満たし、その内 $(C \in [v_1, v_2])$ を満たす確信度が指定した最小確信度を満たすなら、その規則は有効であるとみなせる。

例えば今回使用した標本のうち、広島東洋カープの試合の中で、場所が広島、対戦相手が横浜という組み合わせの試合は19試合あったが、そのうち12試合で誤差が-1000以上であった。最小支持数を10、最小確信度を0.6とすると、これは支持数19の確信度12/19で、広島と横浜の組み合わせで誤差が-1000以上になるという有効な規則になる。

これを曜日、場所、対戦相手、天気などの組み合わせに適用して、それぞれの組み合わせにおける誤差を、+2000以上、+1500以上、+1000以上、-1000以上

+1000 未満, -1000 未満, -1500 未満, -2000 未満に小分けして, 最大の確信度を持つ組み合わせを有効であるとし, 予測に反映させることで, 実測値に近い数値を出す。

4 提案手法

関連研究では, 誤差の発生しやすい状況に着目して研究がすすめられた。本研究では, 観客数との関係性の大きい要素を見つけることで, 改善できると考えた。

4.1 LVQ

LVQ とは, 高次元の属性データを扱う場合に適した競合型ニューラルネットワークモデルである。入力データと空間上の各ユニットの距離に応じてユニットの位置の修正を行い, 学習を繰り返すことで, 理論的に最適な識別境界を形成する。ノイズの多い属性データを入力しても, 精度の高い識別が可能である^[5]。

4.1.1 LVQ のアルゴリズム^[6]

LVQ は入力層と競合層から構成され, 入力層には入力データベクトル x_i が入力される ($i = 1, 2, \dots, N$)。競合層のユニットは, いずれかの対象物が属している分類カテゴリーを示しており, 各分類カテゴリーには複数のユニットが存在可能である。

以下の図 1 に LVQ の基本構造を示す。

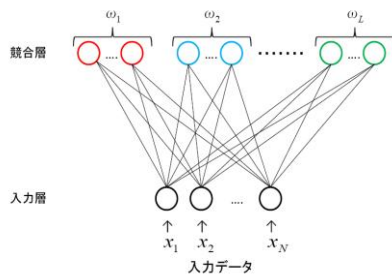


図 1. LVQ の基本構造

LVQ には LVQ1, LVQ2, LVQ3, OLVQ 等が存在する。一般に LVQ の精度と学習に必要な時間は, 以下の 2 点に依存する^[7]。

- (1) 各カテゴリーに割り当てられたコードブックベクトルのほぼ最適な数とその初期値
- (2) 詳細なアルゴリズム, 学習回数中に加えられた適切な学習率や学習の停止のための適切な基準

LVQ1 は, より頑丈な処理過程を行うようにアルゴリズムが定義されており, コードブックベクトルは非常に多くの学習回数の後でも一定値に留まると仮定されている。また, 学習率は早く収束するようにだい

たい最適化することが可能であるため, 本研究では LVQ1 を使用した。

4.1.2 LVQ1 のアルゴリズム

LVQ1 の学習は, 入力データベクトルと最もユークリッド距離の小さい勝者ユニット c が教師データと同じカテゴリーに属する場合には, その結合重みベクトルを入力ベクトルに近づける方向に修正, 異なるカテゴリーに属する場合には入力ベクトルから遠ざける方向に修正することで行う。

結合重みベクトルの初期値を $w_i(0)$, $i = 1, 2, \dots, M$ とし, 教師データの入力回数が t 回の時の結合重みベクトルを $w_i(t)$ とする。 t 回目の教師データ $x(t)$ が入力された後, 入力データベクトルと各ユニット間のユークリッド距離が最も小さいユニットを勝者ユニット c とする。勝者ユニット c の算出式は以下である。

$$\text{勝者ユニット } c = \arg \min_i \|x(t) - w_i(t)\|$$

そして, 以下の更新式に従って結合重みベクトルを更新する。

- (1) $x(t)$ と c の属するカテゴリーが等しい場合

$$w_c(t+1) = w_c(t) + \alpha(t)[x(t) - w_c(t)]$$

$$w_i(t+1) = w_i(t), i \neq c$$
- (2) $x(t)$ と c の属するカテゴリーが異なる場合

$$w_c(t+1) = w_c(t) - \alpha(t)[x(t) - w_c(t)]$$

$$w_i(t+1) = w_i(t), i \neq c$$

ここで, t は学習時の学習回数である。 $\alpha(t)$ は学習率を示し,

$$\alpha(t) = \alpha_0 \left(1 - \frac{t}{T}\right)$$

である。ここで α_0 は初期学習率, T は全学習回数であり, $0 < \alpha(t) < 1$ である。学習回数が増えるほど α は 0 に近づく。これは, 学習初期には結合重みベクトルの修正量が多く, 学習を重ねるごとに修正量が少ないことを示している。

4.2 処理の流れ

LVQ1 の C 言語のプログラムはヘルシンキ大学工学部 Natural Network Research Center で公開されている LVQ_PAK^[8] を利用した。扱う標本は二次元配列に格納する。二次元配列の横一列にある試合の状況を格納し, それを試合の数だけ縦に並べたものになる。データの冗長性をなくすためにダミー変数を用いているので, 20 個の変数と 1 個のラベルから二次元配列は構成される。

まず LVQ1 を用いて観客数を基にクラスタリングを行う。LVQ1 によって, 入力された 20 個の変数が縦

横の2変数に変換され、ラベルとして入力しておいた、試合の行われた日付が平面上に出力される。

以下の図2にLVQ1の出力例として、広島市民球場の出力を載せる。

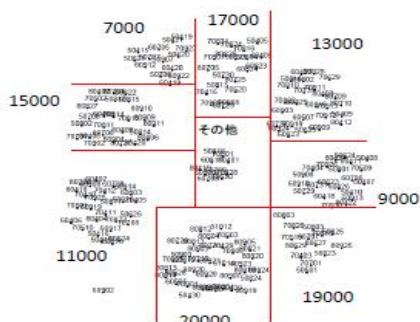


図2. LVQの出力例

図内の赤線は、LVQ1を実行した際の2変数を基にした境界線である。この2変数は、それぞれ2次元平面上におけるx座標、y座標に該当する。

それぞれのカテゴリで観客数.datというデータファイルを作成し、2変数の値によって、該当のファイルにラベルである日付を入力していく。その後、観客数.txtファイルに該当の日付の試合の行われた状況と実測観客数を二次元配列として入力する。以下の表3に試合状況と観客数を入力した二次元配列の例を載せる。

表3. 二次元配列の例

	曜日	相手	天候	貯金
0	10000	1000	10	1
1	00100	1000	01	1
2	00000	0100	10	0

休日	順位	人気	連勝敗	実測値
0	1	0	1	15645
0	1	1	1	15320
1	0	0	1	15879

予測したい試合の状況を入力し、input.datに格納する。input.datの形式は表3と同様である。入力された試合会場のファイルを開き、観客数.txtとinput.datを比較し、天候、貯金、休日、順位、人気、連勝敗の中で異なる値が2個以下であれば実測値を出力し、出力された同じカテゴリ内の平均値を出力し、予測値とした。曜日、相手で誤差を許すと、試合状況が大きく変わってしまい、類似として扱えないので、誤差は許していない。また、天候から連勝敗までであるが、誤差1以下を許容すると、ほとんど類似データが見つからず、誤差2以下にするとほとんどの状況で出力された。誤差2でも出力されない、つまり類似データが見つからない場合は、誤差3まで許し、再度比較を行う。

以下の図3に、出力の例を載せる。

```

ave3= 0.000000
*****
11950.000000 (誤差数: 2) in class 4
ave4= 11950.000000
*****
13775.000000 (誤差数: 2) in class 5
13111.000000 (誤差数: 2) in class 5
ave5= 13443.000000
*****
16061.000000 (誤差数: 2) in class 6
ave6= 16061.000000
*****

```

図3. 出力の例

図3のように、複数のカテゴリにまたがって出力された場合には、出力された試合数が一番多いカテゴリであるclass5を採用し、その平均値である13443を予測値とした。

5 性能評価・比較

性能評価・比較のため、重回帰分析のみ、関連研究（重回帰分析と相関規則を組み合わせたもの）、LVQの3手法について実験を行った。

5.1 性能評価・比較に用いたデータ

性能評価に使用したデータは、2009年7月1日～31日に行われた試合を対象とした。また、広島戦について、既知データは広島市民球場のものを使用している。しかしながら、2009年よりMAZDA ZOOM-ZOOMスタジアムで試合が行われている。したがって、試合会場が異なるため、広島で行われた試合は除いた。

5.2 球団ごとの比較

3つの手法で球団ごとにそれぞれ予測を行った。最も平均誤差が少なくなった球団と最も平均誤差が多くなった球団の結果を以下に示す。

最も平均誤差が少なくなったのは巨人戦だった。その出力結果の誤差をまとめたものを図4に示す。

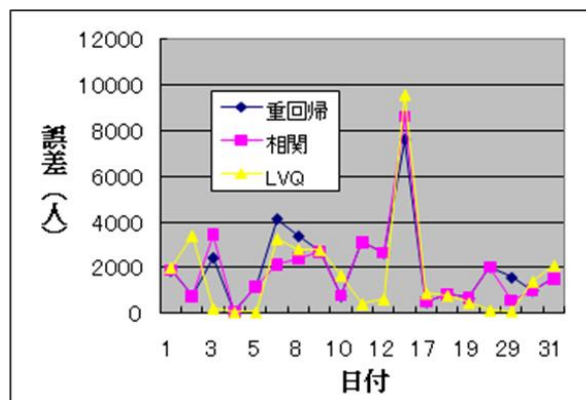


図4. 巨人戦における誤差比較

最も平均誤差が多くなったのはヤクルト戦だった。その出力結果の誤差をまとめたものを図5に示す。

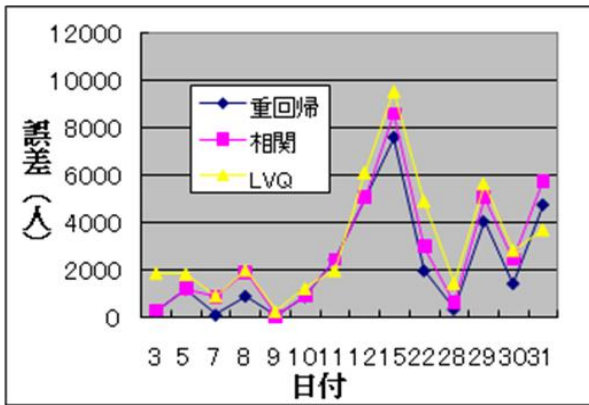


図5. ヤクルト戦における誤差比較

各球団の平均誤差と平均誤差率についても調査した。平均誤差率は以下の式より算出した。

$$\text{平均誤差率(\%)} = \frac{\text{誤差の合計}}{\text{実測値の合計}}$$

比較結果を表4、図6に示す。

表4. 球団ごとの平均誤差 (平均誤差率)

	巨人	阪神	中日	ヤクルト	広島	横浜
重回帰	2048 (4.83)	2486 (5.76)	2632 (8.65)	2239 (9.43)	1959 (7.47)	2343 (8.45)
相関	1943 (4.58)	2819 (6.53)	2904 (9.54)	2739 (11.53)	2319 (8.85)	2205 (7.95)
LVQ	1740 (4.10)	2319 (5.38)	2030 (6.67)	3189 (13.43)	2254 (8.59)	2094 (7.55)

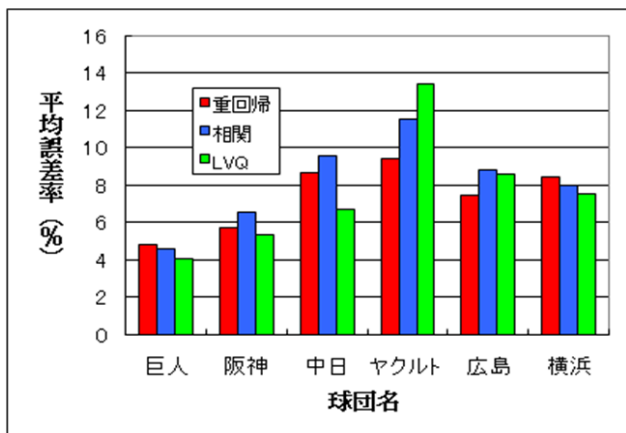


図6. 球団ごとの平均誤差率

図6より、6球団中4球団で平均誤差率が最小となった。その中でも中日戦での改善が一際大きくなっている。

5.3 球場ごとの比較

試合の行われた球場によって予測精度にばらつき

があるので、球場ごとにも同様の比較を行った。なお、広島での試合については予測から除いているため、比較対象としていない。

最も平均誤差が少なくなったのは東京ドームだった。その出力結果の誤差をまとめたものを図7に示す。

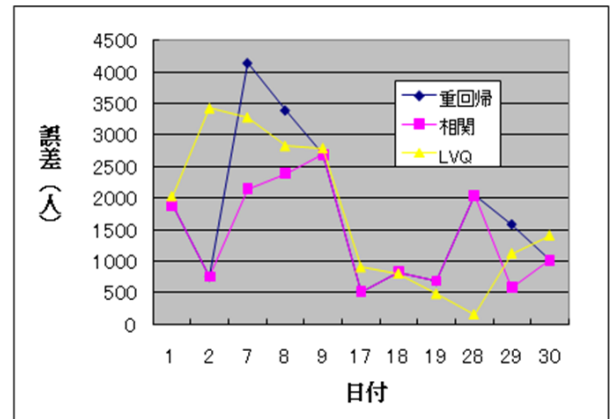


図7. 東京ドームにおける誤差比較

最も平均誤差が多くなったのは神宮球場だった。その出力結果の誤差をまとめたものを図8に示す。

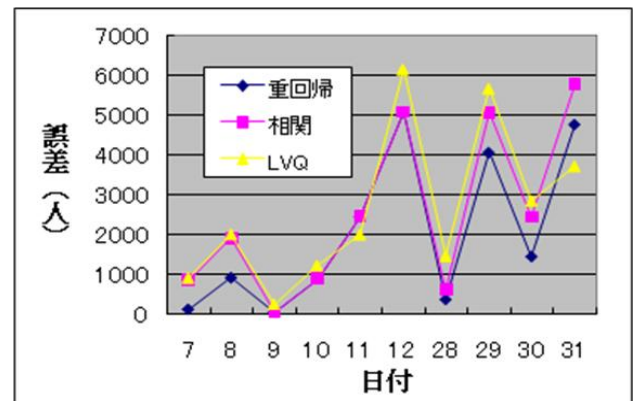


図8. 神宮球場における誤差比較

各球場の平均誤差と平均誤差率についても調査した。比較結果を表5、図9に示す。

表5. 球場ごとの平均誤差 (平均誤差率)

	東京	甲子園	中日	神宮	横浜
重回帰	1776 (4.09)	1911 (4.24)	3336 (10.04)	2023 (11.95)	2087 (13.28)
相関	1412 (3.25)	2372 (5.27)	3636 (10.95)	2522 (14.90)	1909 (12.14)
LVQ	1746 (4.02)	1825 (4.05)	2601 (7.83)	2638 (15.58)	1152 (7.33)

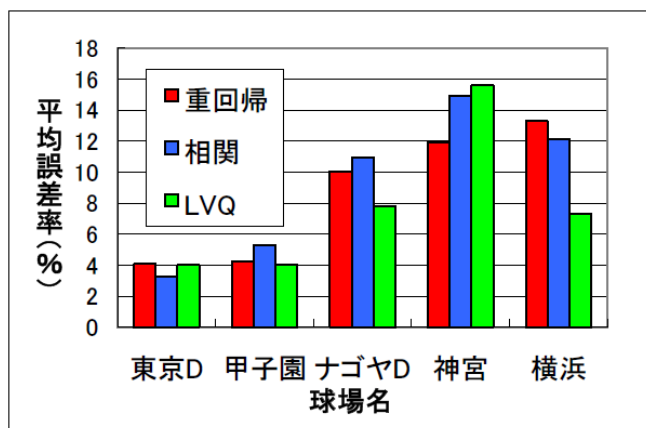


図 9. 球場ごとの平均誤差率

図 9 より、3 手法の中で、LVQ を用いた場合、5 球場中、甲子園球場、ナゴヤドーム、横浜スタジアムの 3 球場で予測誤差が最小になった。また、横浜スタジアムについては最も悪い結果である重回帰分析の半分程度まで誤差が減った。神宮球場の誤差がやはり比較的大きい。神宮球場では 7 月下旬から 8 月終わりまで、試合中に花火が打ち上げられる。出力結果を見ると大きな誤差が出力された時期と同時期であるため、観客数に影響を与えたと考えられる。東京ドーム、甲子園球場は人気の根強い球団の本拠地であるため、激しい観客数の増減があまり見られなかった。また、この球場ごとの結果が、球団ごとの結果に大きく影響していた。

6 おわりに

本研究では、試合の行われた状況やチーム状態から、プロ野球の試合の観客数を予測することを目的とし、LVQ を用いた手法を提案した。また、既存の手法と本研究での手法の予測誤差について比較を行い、結果を示した。

結果として、今回提案した手法は、全球団・全球場で最善の結果とはならなかったが、半数以上の球団・球場で最善の予測となり、ある程度の成果をおさめることができた。

また、今シーズンから広島の本拠地球場は広島市民球場から MAZDA Zoom-Zoom スタジアム広島にかわった。新球場への注目度の高さから、新球場の入場者数は旧広島市民球場の 2 倍近くになり、球団創設以来最速で入場者数 100 万人を突破した。また、総入場者数も過去最多の 178 万人を記録した。例年とは異なり、今後どう変化していくかわからないこの状況にどう対処していくか、ということが今後の課題に挙げられる。

パ・リーグの試合の観客数予測や、今後試合数が十分集まってからセ・パ交流戦、他の競技における予測

についても検討を行うと、より価値のあるものになると思われる。

先にも述べたように、今回はデータをプロ野球の試合に限定して実験を行い、ある程度の成果をあげることができた。しかし、他の数値データ（異種競技・売上げ等）についても有用かどうかは、実験を行っていないため分からない。したがって、他の数値データを収集し、実験を行う必要がある。

謝辞

インターネット経由で、気象データや試合状況、観客数等のデータ、LVQ を提供してくださった、気象庁、日本プロ野球機構、ヘルシンキ大学工学部 Natural Network Research Center のスタッフの方々に深く感謝いたします。

また、本研究の一部は、日本学術振興会、科学研究費補助金（基盤研究（C）課題番号 20500137）の支援により行われた。

参考文献

- [1] 日本プロ野球機構オフィシャルサイト
<http://www.npb.or.jp/>.
- [2] 気象庁
<http://www.jma.go.jp/jma/index.html>
- [3] プロ野球貯金グラフ虎ノ軌跡 2009
<http://www25.big.or.jp/~hidea/tigers/>
- [4] 福原徹：重回帰分析とデータマイニングによるプロ野球の試合の観客数予測，広島市立大学情報科学部卒業論文，2007.
- [5] 株式会社エイム 学習推論型エンジン”NEXTe”
http://www.aim-inc.co.jp/service/pdf/NEXTe_engine.pdf
- [6] 小西智久：リモートセンシングを用いた水稲作付領域の抽出と生育状況の監視に関する研究，大阪府立大学博士論文，2008.
- [7] T. コホネン：自己組織化マップ 改訂版，シェプリンガー・フェアラーク東京，2005.
- [8] LVQ ソフトウェアパッケージ
http://www.cis.hut.fi/research/lvq_pak/