

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 16 日現在

機関番号：82632

研究種目：挑戦的萌芽研究

研究期間：2012～2014

課題番号：24650404

研究課題名(和文)スパコン等の並列計算環境を用いた野球選手の評価手法に関する研究

研究課題名(英文) Numerical Evaluation of Baseball Players on Parallel Computational Environments

研究代表者

大澤 清 (OSAWA, Kiyoshi)

独立行政法人日本スポーツ振興センター国立スポーツ科学センター・スポーツ科学研究部・契約研究員

研究者番号：60601135

交付決定額(研究期間全体)：(直接経費) 3,200,000円

研究成果の概要(和文)：本研究ではマルコフ連鎖モデルを野球の試合に適用し、選手ごとにその成績から定義される状態遷移行列を用いて得点の確率分布を求め、その結果から試合に勝つ確率を求めて選手を評価する手法を提案した。従来手法では打撃成績のみを用いて計算を行っていたが、本研究では守備成績の失策数をモデルに組み込み、より実際の野球に即した計算を行った。

その結果、失策の多い野手が少ない野手に比べて1シーズンで2勝分程度の損失を守備成績によって与えていることが示された。またモデルに守備成績を加えたことで増加した計算量に対応するため並列計算環境を利用し、約2億1千万の打順から1時間22分程度で最適な打順を選ぶことが可能となった。

研究成果の概要(英文)：A probability distribution of runs scored in a baseball game and a winning percentage of the game are able to be computed by multiplying matrices, which are defined by some results of players.

We proposed the method to evaluate players by the relation between the winning percentage and their results. The existing model to compute the probability distributions was based on only batting results, so that fielding results were not applied to the computation. We incorporated fielding results into the extended model.

The numerical experiment indicated that impact of difference in winning percentages, which were calculated from the model with fielding results of players who have good and poor fielding skill, corresponded to about 2 wins for a season. While this problem had large computational complexity due to adding fielding results to the model, the parallel computational environment enabled us to find an optimal batting order among about 210 million orders in 1 hour 22 minutes.

研究分野：情報工学

キーワード：ハイパフォーマンスコンピューティング オペレーションズリサーチ

1. 研究開始当初の背景

プロ野球(NPB)は国民的スポーツともいわれるが、一部の人気球団を除いた多くの球団経営にはオーナー企業からの損失補填が欠かせず、その経営状況は決して良いものとはいえない。このような状況に至る原因の一つとして、選手の勝利に対する貢献度とその選手に支払われる年俵とのミスマッチが挙げられる。例えば、球団の成績低迷に伴いスタジアムの入場料収入が減少しているにもかかわらず、チームの編成部は伝統的な指標を選手との契約の判断材料とし、かつ必ずしもその指標は勝利への貢献度とは比例していないため、本来高い給料を払うべきではない選手に億単位の年俵が費やされるという状況が発生している。

このような状況から脱却すべく、セイバーメトリクスとよばれる新しい、選手の勝利への貢献度をよく表した指標が米大リーグの球団や日本の一部の球団において、選手との契約の判断材料として使われつつある。2000年代前のオークランド・アスレチックスや近年の北海道日本ハムファイターズなど、セイバーメトリクスを用いてチーム編成を行ったとされる球団が少ない予算規模で上位に進出するなど、一定の成果を収めた。

2. 研究の目的

低予算でも好成績を収めることが可能になるような球団編成において、選手の獲得基準として利用されたいわゆるセイバーメトリクスと呼ばれる指標であるが、この指標の計算方法にはいくつかの問題点が存在する。

- ・出塁率などの勝利との相関が「高そう」な成績を人間が選択し、それを用いて指標を計算するため、客観的ではない。

- ・仮に勝利への貢献度との相関が高い指標が得られたとしても、それは過去のシーズンの結果に基づくものであり、2011年シーズンのようにボールの材質が変わり、それに伴い野球の「質」が変化した場合に対応できない。

以上より、主観の部分を極力排除し、かつ野球の質の変化にもすばやく対応できる新たな評価指標が必要となる。本研究では、比較的試合中の各状態が離散化されているためモデル化しやすい野球の特性を活かし、試合中の様々な状態の間をある確率で遷移するマルコフ連鎖モデルを採用して解析を行う。より具体的には選手の成績(打撃成績、守備成績、投手成績)がある値からある値へ変化することで、試合に勝つ確率にどの程度影響を与えているのかを定量的に評価する。このような解析は行列計算を伴うため本質的に多大な計算量を必要とするが、スーパーコンピュータ(スパコン)等の並列計算環境を利用することで現実的な時間内で計算を行うことが可能となる。

このような評価指標を導入することで、例えば守備成績など、勝つ確率に与える影響が

定量的に評価されていない成績について厳密な評価が可能となり、さらに打率や本塁打以外の打撃成績、例えば犠打や進塁打とよばれる走者を塁上に置いた状況での内野ゴロなど、今まで勝利への貢献度が明らかになっていなかった成績についても、勝つ確率への影響が評価可能となる。この評価手法によりセイバーメトリクスを超える精度の選手評価が可能となり、費用対効果の点でより効率的な選手を発掘することで、少ない予算でも高い勝率を挙げるチームの編成が可能となる。

3. 研究の方法

本研究では、過去の申請者の研究で使用していない守備成績の一部項目をモデル化の際に考慮することにより、その成績がどの程度勝利へ貢献しているかを評価する。多くの項目をモデル化の際に考慮することにより生じる計算量の増加には、新しい並列計算環境を導入して対応する。具体的にはGraphics Processing Unit(GPU)を利用して従来の並列計算環境よりも圧倒的に費用対性能の優れた環境を構築して計算を行う。また大学や研究機関に設置されているスパコンを利用して現実的な時間で解くことができない規模の問題にも取り組む。

より具体的には、実際の野球に即したモデル化を行うために、従来のモデル化では使用していなかった守備成績における失策の発生確率を考慮してモデル化を行った。

並列計算環境に関し、従来大学や研究機関に設置されているスパコン、PCクラスタなどの並列計算機、あるいは地理的に分散しているこれらの計算機を接続して利用するグリッドコンピューティング環境は、計算機科学分野の研究者によるベンチマーク(計算機・計算環境自体の性能評価)や計算物理学分野の研究者による分子動力学(MD)シミュレーションなど、計算と関連の深い分野の研究者により主に利用されていた。並列計算機の中でもその時代の最先端技術を投入して構築されるスパコンの性能向上の速度は目覚ましく、並列計算機の性能評価ランキングとして知られるTop500プロジェクトでは、2010年5月に米国Cray社のXT5システムが1.759PFlops(1秒間に1.759兆回の演算を実行)を記録して1位を獲得したが、1年後の2011年5月には富士通の「京」システムが8.162PFlopsを記録し、4倍以上の性能向上を示している。

このような近年の並列計算機環境の性能向上を勘案すると、もはやMathematicaやMATLABなどの基本的に単体のPCをプラットフォームとして仮定している数値計算ツールを使って小規模な演算を行う必要は無く、組合せ最適化問題のような本質的に計算量が膨大となる問題は並列

計算機に委ねることで、従来計算時間の点で解くことが諦められていた問題から新たな知見が得られることが期待できる。

本研究ではマルコフ連鎖モデルを野球の試合に適用し、選手ごとに定義される状態遷移行列を用いて行列積の演算を行うことで試合に勝つ確率を計算する。より具体的には、行列積の演算から9イニング終了時点でチームが挙げている得点の確率分布が求められ、これを対戦している2チームそれぞれについて求め、これらから一方がもう一方に対して勝利する確率が計算できる。行列積の演算量は行列サイズの3乗に比例するためそのサイズが増加するにつれて急激に増加することになり、かつ野球においては打順が挙げる得点に影響を与えるため、すべての打順について得点の確率分布を計算しようとすれば、最低でも362, 880(9の階乗)通りの打順について計算を行う必要がある。打順を構成する選手の候補が9人を超える場合、この人数を n 人とすると、 $nP9$ 通りの打順(P は順列の数を表す)について計算を行う必要があり、 $n=10$ の場合は3, 628, 800通りと9人の場合の10倍になり、 $n=11$ の場合は約2, 000万通り、 $n=12$ の場合は約8, 000万通り、 $n=13$ の場合は約2億6, 000万通りとなり、打順(順列)の数は急激な増加を示す。

このような爆発的な計算量の増加に対応するために、上記のスパコンのような並列計算環境を利用して現実的な時間内で計算を行う。ただし、現状大学や研究機関に設定されている並列計算機を利用するには一般的にその機関が定める利用目的に本研究テーマが合致するか審査を受ける必要があり、審査の結果利用できない場合は自前で並列計算環境を用意する必要がある。また利用可能であった場合も、研究機関の運用上の都合等で常に計算機が利用可能とは限らないためテスト環境は必要であり、いずれにしても自前の並列計算環境が必要となる。

この並列計算環境の構築のために、2009年に長崎大で開発されたスパコンDEGIMA(0.158PFlops)でも採用されているGPUの利用が考えられる。GPUは従来3Dゲームの描画用に開発されたハードウェアであるが、近年その処理能力の向上により一般的な数値計算用途にも利用が広まりつつある。GPUはコンピュータの周辺機器の一つであるため一般的な店舗で入手可能であり、計算能力に対する価格や消費電力は一般的なPCで使用されているCPUよりも圧倒的に優れている。GPUを数台接続することで一般的に大学の研究室などで利用されているPCクラスタの性能に並ぶ並列計算環境が安価に、かつ低い消費電力で構成可能である。本研究ではこのようなGPUを用いた並列計算環境を導入することで、研究機関などの定める利用基準に合致しないような萌芽的研究についても大規模

な計算が実行可能となり、そこから新たな知見が得られることを示す。

4. 研究成果

野球選手の勝つ確率に対する貢献度を測るために野球における様々なプレーをモデル化し、選手の各種成績をそのモデルに当てはめてどの程度その成績が勝つ確率に影響しているかを定量的に評価した。

具体的には、野球の試合中にとり得る状態(無死一塁や一死二塁、二死満塁など)を定義した上で打撃成績、走塁成績、相手の守備成績を考慮に入れて状態間の遷移行列を選手ごとに定義し、マルコフ連鎖を用いて1イニングに挙げる得点の確率分布とそれから得られる9イニング終了時点の得点の確率分布(延長戦に対応するため10~12イニングについても計算する)、与えられた2チームの得点の確率分布から得られる両チームの勝つ確率を計算する。ここでいう「勝つ確率」とは、一般に勝率として用いられている勝利数を勝利数と敗戦数で割った値ではなく、上記計算を行って得られる得点の確率分布から計算される「試合終了時に相手の得点を上回る確率」を指している。この勝つ確率がある選手の成績を変化させた場合にどの程度向上・減少するかを計算することで、変化させた成績が勝つ確率に与える影響を定量的に測ることが可能となる。

過去の申請者の研究で使用している打撃成績(二塁ゴロやセンターフライ、レフトへの二塁打のような打球を処理した野手まで区別した成績)や走塁成績(走者としてある塁に出塁した際、打者による打撃でいくつ先の塁まで進塁したか、または封殺・タッチアウトとなったか、あるいは占有していた塁にとどまったか)、守備成績(野手として打球を処理した際の相手選手の上記のような走塁成績)に加え、本研究では別の守備成績、具体的には通常捕球あるいは失策の発生確率を計算に組み込み、この成績が勝つ確率に与える影響を定量的に測った。

その結果、2012年シーズンに最多失策を記録した広島・堂林三塁手は、守備成績が良好なヤクルト・宮本三塁手に比べてシーズンの勝利数換算で約2試合分勝つ確率を低下させる影響を守備成績のみで与えていることが示され、打撃成績の向上でこの低下を埋め合わせようとする、1シーズンで約9本の本塁打が必要になることが示された。このようにして、守備成績が勝つ確率に与える影響を測ることが本研究により可能となった。

本研究では二つの並列計算環境上(スパコン、GPUクラスタ)に提案したモデルに基づく計算を行うプログラムを実装することを目標としている。

スパコン上での実装を行うため、東京大学情報基盤センターに設置されたFX10スーパーコンピュータシステム(以下、FX1

0システム)のアカウントを取得して数値実験を行った。FX10システムでは並列計算環境として一般的に用いられているMPI(計算ノード間の並列化に利用されるインタフェース)とOpenMP(計算ノード内の並列化に利用されるインタフェース)が用意されているが、本研究課題で利用するプログラムはPOSIX Threadsとよばれるインタフェースを利用してより計算ノード内の並列化効率を高めることを目指し、その実装を行った。並列処理特有の排他制御(あるプログラムが利用しているデータを他のプログラムが利用しないようにする等の、計算の正当性を確保する処理)を本実装においても適切に行う必要がありその部分の作成に時間を要したが、同インタフェースを適切に配置して動作させることにより希望する並列処理が可能となった。その結果、上記の守備成績が勝つ確率に与える影響の評価や、平成25年3月に開催された野球の国際大会(2013 World Baseball Classic)における日本代表の最適な打順をFX10システムの計算ノード96台を利用して約2億1千万通りの打順から1時間22分程度の計算時間で選ぶことが可能となった。

GPUクラスタ上の実装については、購入したGPUクラスタが最終年度にハードウェア故障を起こし利用不可能となったため本格的な数値実験の環境として利用することはできなかったが、予備実験として最小の問題サイズ(打者9人による打順の数:362,880通り)・従来用いられていたモデル(走塁成績・守備成績を組み込んでいない、打撃成績のみを用いたモデル)で相手に対して勝つ確率が最高となる最適な打順を求めたところ、GPU特有のプログラムにおける高速化手法(CPUの2次キャッシュに相当するGPU内の共有メモリを有効活用するようなプログラムを作成するなど)を適用しないプログラムを使った場合においても17分程度(995秒)で計算が完了し、十分実用的な時間内で最適解を求められることが示された。

さらに、ある選手のあるプレーが勝つ確率に与える貢献度を測るためのモデル化を行った。具体的には、例えば一死一塁において打者が二塁ゴロを放った場合について、過去数年分の記録(日本のプロ野球における全記録)から40%が二死一塁、30%が二死二塁、29%が併殺により三死(攻撃終了)、1%が二塁手のエラーによりアウトカウントが増えない状況(一死一二塁など)にそれぞれ遷移したとすると、各遷移後の状況から計算により得られる勝つ確率を40%等の各遷移確率で重み付け平均した値が「日本のプロ野球における平均的な一死一塁で二塁ゴロを打者が放った後の勝つ確率」となる。この値と打者が二塁ゴロを放つ前の一死一塁における勝つ確率との差分を取れば二塁

ゴロによる守備側の勝つ確率の平均的な利得(攻撃側は損失)となり、これを投手の勝つ確率に対する貢献度とする(打球を二塁手がどのように処理したかは投手が関与していないため、実際にその二塁ゴロの後にはどの状況に遷移したかは問わない)。ゴロを処理した二塁手は一塁走者を二塁で封殺、または打者走者を一塁で封殺するというプレーのいずれかを選択することになり、前述の勝つ確率を重み付け平均する方法を用いて同様に「走者が一塁にいる状況で二塁手がゴロを処理した後の平均的な勝つ確率」と「同状況で二塁手実際に選択したプレー後の状況から計算で得られる勝つ確率」との差分が二塁手の勝つ確率に対する貢献度(二塁手がエラーした場合はマイナス)とする。以下同様にして他の野手の貢献度も算出できる。以上のような一つのプレーが勝つ確率に与える貢献度を測る方法の提案を行った。この定量的評価手法に関しては近日中に公開する予定である。

5. 主な発表論文等

[学会発表](計2件)

- ① Kiyoshi Osawa and Nobuyoshi Hirotsu: the Technique of Finding an Optimal Batting Order in Baseball with Base Running Results, International Symposium on Scheduling 2015, Jul. 5, 2015, Kobe Industrial Promotion Center Convention Hall, Kobe, Hyogo, Japan
- ② Kiyoshi Osawa: Effects of Fielding Results on a Winning Percentage of a Baseball Game, the 9th International Symposium on Computer Science in Sport (IACSS 2013), Jun. 21, 2013, Marmara University, Istanbul, Turkey

6. 研究組織

(1) 研究代表者

大澤 清 (OSAWA, Kiyoshi)

独立行政法人日本スポーツ振興センター
国立スポーツ科学センター・スポーツ科学研究部・契約研究員

研究者番号: 60601135

(3) 連携研究者

合田 憲人 (AIDA, Kento)

国立情報学研究所・アーキテクチャ科学研究系・教授

研究者番号: 80247212