

ビジュアルプログラミング言語によるビックデータ解析システムの開発

富山県立大学大学院工学研究科情報システム工学専攻
1955016 横井稜

指導教員：奥原浩之

1 はじめに

近年、企業などでは世間に溢れる様々な情報を収集し、ビックデータと呼ばれる非常に巨大で複雑なデータの集合として扱うことが増えてきている。また、ビックデータを扱うには様々な処理や解析によって情報を取捨選択し、自分たちに必要な形で保管する必要がある。しかし、ビックデータの情報は膨大で、人の手で全て解析するのは困難であるため、一般にプログラミング言語を用いて機械に処理させるのが一般的である。このため、ビックデータを扱うためにはプログラミングの知識や技術が必要不可欠であり、プログラミングに触れたことがない人には扱いづらいものになっている。

本研究では、プログラミング初心者でも扱いやすいビジュアルプログラミング言語にビックデータの処理を落とし込み、ブロックを並べるだけで処理できるプログラムを作成することでプログラミングができない人でもビックデータを扱うことができるようにすることを目標にする。

2 ビジュアルプログラミング

2.1 ビジュアルプログラミング言語

【記号】 t : 期 D_t : t 期における需要量 y_t : t 期における目標在庫量 q_t : t 期における発注量; t 期の期末に、 t 期の需要量 (D_t)、当期と翌期の目標在庫量レベルとの差によって決まる。すなわち、次のようになる。

$$q_t = y_{t+1} - y_t + D_t \quad (1)$$

ブルウィップ効果 B は、需要の分散に対する発注量 (発注者の需要量) の分散の比で表す。

$$B = \frac{Var[q_t]}{Var[D_t]} \quad (2)$$

次に、ブルウィップ効果の要因を考える。12 個がブルウィップ効果の要因として考えられる。

- [1] 発注から納入までのリードタイム
- [2] 緊急発注量のバッチ単位の注文方式
- [3] 需要のばらつき
- [4] サプライヤーの生産計画手法
- [5] 意思決定者のオーバーアクション
- [6] 複数の要因を考慮した発注方式
- [7] 週単位計画から日単位計画
- [8] 製造ロットによる上流と下流のマッチングの難しさ
- [9] サプライヤー間の情報共有
- [10] 価格割引
- [11] 内示変動のときのトレンドやばらつきの変動
- [12] 安全在庫目標の決め方
- [13] 緊急注文タイミングのアルゴリズム

それぞれの要因について説明する。リードタイムは、発注から納入までかかる日数を表す。つまり、サプライヤー側は先に生産しておく必要があることが要因となる。バッチ単位の注文とは、部品を 1 個ずつ注文するのではなくて、1000 個ずつなど複数個ずつ注文すること、または、最低注文数を決定してそれ以上注文することを指す。つまり、倍数ロット数による注文方式及び最小ロット数による注文方式である。

需要のばらつきは、需要量の標準偏差が変わることでブルウィップ効果の式に需要量の分散が含まれていることから影響を与える事が分かる。サプライヤーの生産計画手法は、サプライヤーの設備能力や製品品質のばらつき等による供給不足時の対応を指す。オーバーアクションは、通常時に連続で供給不足に陥った時などに、意思決定者による来期の不足の予想及び不安の感情などにより多めに注文することを指す。

複数の要因を考慮した発注方式は、複合的に要因を考えた場合に、要因それぞれが互いに及ぼしあう影響からブルウィップ効果に影響を与えられ思考えられるから要因となる。製造ロットによる上流と下流のマッ

ングの難しさとは、1 次サプライヤー間では 10 個から生産できるが、2 次サプライヤーの部品は 1000 個からしか生産できず、1 次サプライヤーの部品が 2 次サプライヤーの部品は一つだけ必要な場合などに発生する。価格割引とは、月によって部品が安いことがあることを指す。

内示変動のときのトレンドやばらつきの変動とは、内示にトレンド性がある場合に、内示のばらつきを調べる期によって変化することを表し、発注量に影響を与えられ思考えられる。安全在庫目標の決め方は、多くの方法がある。

2.2 直交表による実験計画

主効果とは、それぞれの独立変数がそれぞれ「独自」に従属変数へ与える単純効果のことである。交互作用とは、独立変数を組み合わせた場合の複合効果のこと、要因 A の主効果と要因 B の主効果だけでは説明できない組み合わせ特有の効果がみられることである。

直交表とは、任意の 2 因子 (列) について、その水準のすべての組合せが同数回ずつ現れるという性質をもつ実験のための割り付け表である [4]。図 2 が、4 因子 3 水準の L_9 直交表の例である。

$L_9 (3^4)$				
実験No.	因子列			
	A	B	C	D
1	1	1	1	1
2	1	2	2	2
3	1	3	3	3
4	2	1	2	3
5	2	2	3	1
6	2	3	1	2
7	3	1	3	2
8	3	2	1	3
9	3	3	2	1

図 2 $L_9(3^4)$ 直交表

2.3 MPI による並列分散処理

並列コンピュータのアーキテクチャは、3 つに分類することができる。分散メモリ型、共有メモリ型、分散共有メモリ型である。今回は、クラスタ環境は分散メモリ型並列計算機とする。その並列分散処理を行う方法として Message Passing Interface(MPI) と hadoop と GPGPU がある。MPI と hadoop は、並列コンピューティング利用するための標準化された規格である。

MPI は、プログラミング言語とは独立の通信プロトコルで、並列計算機上で動くプログラムに使用される。MPI を使うメリットは、カスタマイズ性に優れていることで、デメリットは、耐障害性への考慮が少ないことである。hadoop は、java ベースの分散システムで、通信のコードを書く必要がない。hadoop を使うメリットは、耐障害性があることで、デメリットは、ファイルの読み書きが遅いことである。GPGPU は、GPU の演算資源を画像処理以外の目的に応用する技術のことである。GPGPU を使うメリットは、CPU の数十倍のコア数があるため CPU のみの並列処理より早くなることで、デメリットは、コーディングが難しいことである。

3 提案手法

3.1 直交表の作成

本研究の実験と合致する直交表がある場合は、その直交表を実験に用いる。無い場合は、直交表の定義を元に作成する。 $C = \{0, 1, \dots, c-1\}, c \in G$ とおく。C は水準の集合、c は水準数に対応する。直交表のフォーマルな定義は次のように与えられる。

C の元を要素とする $G \times H$ の 2 次元配列が直交表であるとは、ある自然数 q, λ が存在して、任意の $G \times q$ 部分配列 ($0 \leq q \leq H$) が、行

