データマイニング手法によるSE管理に関するマクロエルゴノミクス

○新原 雅也 和多田 淳三 藪内 賢之 (大阪工業大学)

The Macro Ergonomics on Management of SEs by a Data-mining Method.
Masaya Niihara, Junzo Watada, Yoshiyuki Yabuuchi (Osaka Institute of Technology)

1. はじめに

マクロエルゴノミクスの観点の1つに、システムや組織と人とのインターフェイス、システムや組織と人とのスムーズな交渉がある。ここでは、そのような観点からソフトウェア開発に関するSE管理について分析を試みる。

現在、ソフトウェアの量が急増するとともに、大規模複雑化し、そのソフトウェアの開発・保守を組織的に行うことの必要性が強く認識されるようになってきている。なぜなら、ソフトウェアの性能、信頼性もかなり低下するなど、計算機システムの大規模化に伴い、ソフトウェアの生産、保守のコストと所要時間が深刻な問題として注目されるに至ったからである。

これらの問題を解決、改善するために、ソフトウェア開発に携わっているSEの管理データベースを構築し、分析を施すことによって問題点を発見することが重要となる。

そこで本研究では、SEのプログラムバグに関するSE管理情報のデータベースをもとに、データマイニング手法を用いて、マクロエルゴノミクスの観点よりSEの作業・管理の問題点を簡単に導き出すことを目的としている。

2. データマイニング

2.1. データベースにおける知識発見

データベースにおける知識発見(KDD: Knowledge Discovery in Database)の定義は、データから知識を抽出するプロセス全体である。ここで、知識というのはデータの要素間

の関係とそのパターンを意味する。

データベースにおける知識発見の手順は、データの選択(data selection)、データの洗浄(cleaning)、データの補強(enrichment)、データのコード化(coding)、データマイニング(data mining)、報告(reporting)の6つの段階からなっている。データマイニングは、データベースにおける知識発見の中の1つの役割を担っているものである。

2.2. データマイニング

多量のデータ処理を要するデータベース応用分野として、過去の貴重なデータを効率よく活用するためのシステム構築を進める技術を明確にする必要性が高まっている。

そこで、データベースに蓄積されたノイズを含む膨大な生データから、高いレベルで記述された価値ある情報を発掘することを目的とした、データマイニングに関する研究や、関連するシステム構築が盛んになっている。

データマイニングとは、データベースやデータウェアハウス(データの倉庫)に入っているデータを採掘(マイニング)し、情報・仮説・知見・課題などを見つけ出す方法・プロセスをいう。つまり、所有するデータから個々のデータ以上の有効なものを引き出す技術である。また、データマイニングの技術は1つの技術ですべての目的に対応しているわけではなく、さまざまな異なった技術が異なった目的のために使われる。

3. データマイニングによる分析

SE 管理情報のデータベースを、作成規模、プログラム作成参加人数、プログラムの複雑さ等の「プログラム属性に関する項目」と、作業者の質（正確，モラル），プロジェクトの雰囲気，目標の明確さ，作業の明確度といった「人の属性に関する項目」の2項目に分類し，各々の場合において SE 管理に関する知見を得ることを試みる。

本研究では，数量化理論 III 類によるデータマイニング手法を用いる。変数は，修正件数との相関係数が高いものを基にしている。

3.1. 分析結果

修正件数の多くなる要因は，

- 1) プログラム属性に関する項目では，経験年数の少ない人，システムを大幅変更したプロジェクト，仕様変更における連絡票の発行件数の多いもの，開発期間の長期なもの，全体の規模が大きなものであることが分った。
- 2) 人の属性に関する項目では，上司に全く認めてもらえない人，会社で永年働いても実力を発揮できないと思っている人，プロジェクト内で約半数もの働きにくい人がいると考えている人であることが分った。

結果としてデータマイニングによって，プログラムの変更回数の多さ，大規模なプログラム，働きにくい開発環境の各要素において修正件数が多くなるという情報を得ることができた。

3.2. 考察

修正件数（バグ）の発生に関する要因を考察すると以下ようになった。

まず，プログラム属性について考察すると，規模が大きくなり複雑になると，データのやりとりなどが複雑な構造となるためにバグを作りやすくなってしまう。またその中の要因

を個別に考えると，プログラムのロジックの複雑さ，判断分岐の数，インターフェイスの数といったものがあげられる。つまり，プログラムの規模に依存している。

次に，人の属性について考察すると，働きにくい環境に大きく依存している。類似プログラムの経験回数の多さ，プログラムの仕様が詳細まで決まっている等，プログラムを作成する際に構造の理解を容易にさせる項目が，バグの発生を抑える要因となっている。また，作業者の作業に対する満足度を低下させること，同じプロジェクトに働きにくい人がいることなどが，バグの発生要因になっている。よって，職場の組織の要因，環境の要因が，バグ発生につながっているといえよう。

4. おわりに

本研究では，SE のプログラムバグに関する SE 管理情報のデータベースを基に，データマイニング手法により，SE の作業・管理の問題点を分析した。

従来の分析手段と比較しても，本研究よりデータマイニング手法は有用であることを明らかにした。例えば分析したいデータベースが大規模になると，通常の統計手法ではアプリケーションソフトを用いてもそのソフトにおける計算可能な許容量を越えるため，結果が求められなくなる。その場合においても，データマイニング手法では結果を得ることができるからである。これにより，データ量が多いために以前の手法で分析できなかった問題においても，分析可能となるのである。

本研究によりデータマイニングは，POS データや顧客データといったマーケティング方面での分析や，インターネット上での検索システムだけではなく，今回のような管理データを扱う人間工学的な面においても，有効な手法であることを示すことができた。