



情報工学研究院
知能情報工学研究系
准教授
平田 耕一先生

データマイニング

ー 大量のデータの中から有用な規則を見つけよう ー

情報化が進んだ現在、さまざまな学術分野において大量のデータが蓄積されています。ただ、あまりにデータが大量になったため、人間が直接見て調べようとしてもすべてのデータに目を通すことすら困難となってきました。

そこで、処理速度が向上したコンピュータを利用して、大量のデータから有用な規則を見つける「データマイニング」の研究が発展してきました。ただ、いくらコンピュータの処理速度が向上したとはいえ、大量データを扱う際にはやはり工夫が必要です。そこで、データマイニングの研究は、「対象とするデータや見つける規則に応じて高速に動くプログラムの設計」ということを軸に、さまざまな方向に発展してきました。このようなデータマイニングの研究の中でも、私は特に「時系列データマイニング」と「木構造データマイニング」に取り組んでいます。

時系列データマイニングとは、「Aが出現した後にBが出現する」というような時間経過にともなう変化を表す規則を見つけるデータマイニングです。私の研究では、さまざまな薬剤が菌に効いたかどうかという情報を記述した、「薬剤感受性検査データ」を対象としています。そしてそのデータから、院内感染の要因と考えられている、薬剤に対する菌の耐性の変化である「薬剤耐性変化」と薬剤投与の対象である菌に代わって別の菌が生じる「菌交代」を表現する規則を見つけることに取り組んでいます。

一方、木構造データマイニングとは、WebのHTMLデータやXMLデータ、生物学のRNAデータや糖鎖データのように、「木」とよばれる構造で表現できるデータを対象としたデータマイニングです。私の研究では、木構造データがどれだけ似ているのかを表す「類似度」を高速に計算するプログラムを設計し、それによって木構造データを類別することで、特に「糖鎖データ」を特徴づけるような規則を見つけることに取り組んでいます。

さらに、データマイニングの結果を専門家と共に解析することで、専門家も思いつかなかったような興味深い知見を得ることができました。このように、データマイニングの研究は、情報技術をさまざまな分野へ応用する学際的な研究として今も発展しています。