

Adatbányászási technikák műszaki alkalmazása

Tolvaj Béla, Műszaki informatika szak V. évfolyam

**Konzulens: Dr. Pataricza András,
Méréstechnika és Információs Rendszerek Tanszék**

(Összefoglaló)

Napjaink nagybonyolultságú információs rendszereiben komplex folyamatok lejátszódásakor, igen sok naplózott adat keletkezik. E nagymennyiségű információ egyrészt a rendszerbe érkező kéréseket, másrészt a rendszer válaszait, belső állapot változásokat és a fellépő hibákat tartalmazza. A mérési eredményeknek csak csekély részéből nyerhetők olyan adatok, amelyekből a rendszer kritikus helyzetekben való viselkedésére következtethetnénk. Rendezetlen adathalmaz feldolgozására és a rejtett törvényszerűségek felismerésére dolgozták ki a Knowledge Discovery in Databases ill. Data mining módszereket, amelyeket számos szakterületen eredményesen alkalmaznak. A KDD technikák alkalmazásával nemcsak a nagy adathalmazok elemei közötti összefüggések feltárása lehetséges, de következtetések vonhatók le a rendszer jövőbeli viselkedésére és állapotára. Továbbá egy adott állapotsorozat bekövetkezése után megjósolható a következő állapot bekövetkezésének valószínűsége. A gyakorlati alkalmazás során hatékonyabb riasztási algoritmusokat készíthetünk, illetve optimalizálható a szükséges erőforrások elosztása.

Példaképpen egy a Helsinki Egyetemen kifejlesztett alkalmazással egy távközlési hálózatban bekövetkező riasztások közötti összefüggéseket tárták fel ily módon.

A szakirodalomban számos módszere ismert az adatbányászatnak, ezért a dolgozat megírását alapos irodalomkutatózás előzte meg. A dolgozat foglalkozik az adatbányászat lehetőségeivel, többek között a trend analízissel, a clusteringgel és a multidimenzionális OLAP technikával. Ismerteti a műszaki problémák megoldásának tipikus követelményeit. Majd megvizsgálja milyen alkalmazásai lehetnek az adatbányászatnak e téren.

Ezen technikával felismert szabályszerűségek alapján a rendszerről modell alkotható. A dolgozatban a modellezés sztochasztikus petri-hálóval történik, amely paramétereinek meghatározására készítettem alkalmas eljárást.

A modell elkészítése után viselkedését szimulációs szoftverrel vizsgáltam. Majd az így nyert eredményeket és az eredeti rendszerről OLAP-pal készült elemzéseket hasonlítottam össze. A vizsgálatához szükséges adatokat egy hálózati eszközön végzett mérések adták.