

Adatbányászati módszerek felhasználása hibainjektálási adatok feldolgozására

Önálló laboratórium feladat összefoglalója

Kollár Nándor (BUK8Y7)

Konzulens: Pintér Gergely

**BME Méréstechnika és Információs Rendszerek Tanszék
Rendszertervezés ágazat, 2007/2008. II. félév**

Önálló laboratórium munkám témája hibainjektálási adatok elemzése volt adatbányászati módszerekkel. A félév során egy alkalmazás hibainjektálással történő vizsgálatának eredményeként keletkezett adatokon végeztem vizsgálatokat.

Az informatika területén gyakori feladat, hogy különböző eszközöket, programokat szeretnénk összehasonlítani teljesítményük alapján. Ez úgy végezzük el, hogy a céleszközökön méréseket végzünk munkaterhelés mellett, és ezután a mérések eredményét hasonlítjuk össze. Ezt nevezük benchmarkingnak. Biztonságkritikus rendszerek esetén lényeges szempont az összehasonlítás során annak megállapítása is, hogy a céleszköz hogyan viselkedik, milyen teljesítményt nyújt hibák esetén. Ezért a vizsgálat során a nem csak munkaterheléssel, hanem hibaterheléssel is ellátjuk a céleszközt. Ez a dependability benchmarking lényege. Egy ilyen vizsgálat eredményeként egy terjedelmes adathalmaz áll elő. Ezen adathalmaz önmagában nem képvisel nagy értéket, ahhoz, hogy számunkra fontos, gyakorlatban is alkalmazható következtetéseket alkothassunk meg, az adatok részletesebb vizsgálata szükséges. Ezt azonban az adatok nagy száma miatt kézzel már túlságosan időigényes elvégezni, tehát valamilyen automatizált módszerre van szükségünk, ehhez nyújt eszközt az adatbányászat.

A félév első felében irodalomkutatást végeztem, megismerkedtem az adatbányászattal, technikáival, a rendelkezésre álló adatbányász programokkal, illetve a hibainjektálás fogalmával, folyamatával. A félév második felében egy kalkulátor program hibainjektálással történő vizsgálata során keletkezett adatokon végeztem elemzéseket. A vizsgálathoz az adatbányászat egy módszerét az osztályozást használtam, az osztályozó algoritmus által felépített döntési fát vizsgáltam. Az elemzés elkészítéséhez az MS SQL Server adatbányász kiegészítését használtam fel. Az injektált hibák a szoftverfejlesztés és program futtatás során előforduló leggyakoribb hibatípusokat foglalták magukban. Ez négyféle hibatípust jelentett: implementációs, modellfinomítási és fizikai hiba. Az összes adat illetve az egyes hibatípusokhoz tartozó adatok alapján is felépítettem egy-egy döntési fát. Ezután a döntési fákat vizsgáltam meg, amely lehetőséget nyújtott az egyes hibatípusok jellegzetességeinek azonosítására.