

Abstrakt

Jak využít metody dobývání znalostí z databází (data miningu) v oblasti vzdělávání? Data mining či dobývání znalostí z dat je procesem získávání zajímavých a dosud neznámých informací z dat. Tento proces je založen na použití analytických metod, jako jsou analýza hlavních komponent, shlukování či využití neuronových sítí. V rámci projektu pracujeme s daty získanými z didaktických testů, v nichž byla sledována úroveň studentů v předem definovaných dovednostech. Dovednosti byly odvozovány z odpovědí studentů na základě klíče definovaného experty z řad pedagogů. S výstupy získanými z dat metodami data miningu a metodou IRT – teorie odpovědi na položku bude tento experty definovaný klíč srovnáván.

V rámci evaluace procesu vzdělávání a hodnocení studentů se často provádí testování žáků a studentů didaktickými testy, u kterých experti z řad pedagogů a psychologů obvykle definují klíč k určení úrovně jednotlivých typů dovedností. Tento dovednostní model spočívá v přiřazení jednotlivých testovacích otázek k patřičným dovednostem, které otázky se podle jejich názoru vyžadují k úspěšnému vyřešení. Tyto expertní analýzy jsou ale velmi pracné a nákladné – nedalo by se vytváření dovednostního modelu automatizovat aplikací statistických postupů a data miningu přímo na výsledky didaktických testů?

Charakteristika datové základny

- výsledky testování studentů z projektu diagnostika stavu znalostí v příhraniční oblasti,
- testování žáci 7. ročníků ZŠ,
- testované dovednosti z matematiky.

Metody řešení

- Data Mining (Educational Data Mining),
- IRT (Item Response Theory),
- Jednparametrický logický model.

IRT (Item Response Theory)

IRT neboli teorie odpovědi na položku (Item Response Theory), je metodou pro analýzu a skórování testů a dotazníků pro měření žákových schopností či dovedností. Metoda IRT se uplatňuje při vytváření a zdokonalování testů nebo při vyrovnávání obtížnosti testů. Prostřednictvím modelu IRT rovněž určujeme, do jaké míry model koresponduje s daty.

IRT je souborem matematických modelů, které určují pravděpodobnost správné odpovědi na položku v závislosti na úrovni studentových schopností a charakteristikou konkrétní položky.

Cíl projektu

Cílem projektu je aplikovat dostupné metody extrakce vlastností na pokusná data z didaktických testů a výsledky porovnat s klíčem k dovednostem, jak jej definovali experti z řad pedagogů.

Výsledkem budou:

- návrhy na zlepšení dosavadního klíče,
- srovnání výsledků jednotlivých data miningových metod s klíčem definovaným experty,
- a vypracování metodiky k ověřování podobných modelů dovedností na datech.

Data Mining

Data mining je proces, kterým získáváme – „dolujeme“ zajímavé, ale dosud neznámé informace z dat. Mezi nejpoužívanější techniky data miningu patří například rozhodovací stromy, asociační pravidla, shluková analýza, neuronové sítě či regresní analýza.

Poměrně novou výzkumnou oblastí, která nabízí zajímavé možnosti pro aplikaci různých metod data miningu je „Educational Data Mining“ neboli dolování znalostí z dat ve vzdělávání či získaných v kontextu se vzděláváním.

Jednparametrický logický model

Model 1 PL (jedinparametrický model) obsahuje pouze parametr obtížnosti položek. Určuje vztah mezi úrovní latentního rysu probanda θ a pravděpodobností určité odpovědi na položku b_i , který je upravován pouze obtížností položky.

Závěr

Na základě odhadů pravděpodobnosti výskytů jednotlivých odpovědí na položku můžeme určit, které otázky v testu jsou pro žáky obtížnější a které naopak snadné. Potvrdí nám to také hodnota samotného parametru obtížnosti položky.

Pro ověření využitelnosti metod data miningu a IRT budou získané výstupy porovnány s klíčem definovaným experty.