

# Témata diplomových prací

Ing. Jiří Vomlel, Ph.D.  
<http://www.utia.cz/vomlel>

12. července 2007

## 1 Adaptivní testování znalostí pro e-learning

Cílem práce je vytvořit prototyp adaptivního testu pro e-learning. Test by měl být navržen tak, aby si studenti studující určitý e-learningový program mohli ověřit získané znalosti a zjistit oblasti, ve kterých se musí ještě zlepšit. Test by byl navržen jako adaptivní, což znamená, že otázky by byly vybírány vždy na základě správnosti či nesprávnosti předchozích odpovědí.

Prvním úkolem je navrhnout bayesovskou síť modelující závislosti mezi různými dovednostmi potřebnými k úspěšnému řešení určitého typu úloh či otázek. Tento model by pak byl použit při návrhu adaptivního testu, jehož cílem je, pomocí co nejmenšího počtu otázek, dosáhnout potřebné informace o dovednostech testovaného studenta.

Řešitel bude mít k dispozici systém Hugin, který umožňuje aktualizaci bayesovské sítě při pozorování nové informace. Tento systém má aplikační rozhraní pro programy psané v programovacích jazycích C, C++, Java a ActiveX-server.

## 2 Detekce vloupání do počítačového systému

Pro bezpečnost mnohých systémů je důležité včasné rozpoznání nelegálního průniku do systému. Chování vetřelce obvykle vykazuje určité charakteristické znaky, které je možné použít pro rozpoznání.

Cílem práce by bylo pro vybraný systém zkonstruovat pravděpodobnostní klasifikátor, který by byl schopen na základě vybraných atributů vyhodnotit, zda-li pozorované chování odpovídá nelegálnímu průniku do systému, nebo zda-li se jedná o standardní chování. Cílem je podchytit všechna vloupání i za cenu určitého počtu falešných poplachů.

Modelovaným systémem může být například počítačová síť, jednotlivý počítač, systém kreditních karet, síť mobilních telefonů, apod. Výhodou je, pokud řešitel má určité znalosti o typických problémech v bezpečnosti modelovaného systému a je schopen získat data, která by bylo možné použít pro naučení parametrů (případně i struktury) pravděpodobnostního klasifikátoru.

### **3 Dolování znalostí z dat pomocí bayesovských sítí**

Cílem práce je použít metody učení bayesovských sítí pro vyhledání zajímavých souvislostí v datech. Předpokladem je, že řešitel má k dispozici zajímavá data z některé oblasti, v které se dobře orientuje.

Protože řešitel bude mít k dispozici systém Hugin, práce nevyžaduje naprogramování algoritmu učení bayesovských sítí. Důraz bude kladen na předzpracování dat a na jejich názornou interpretaci pomocí bayesovské sítě.

### **4 Učení struktury bayesovských sítí**

Nejnovější algoritmy pro učení struktury bayesovských sítí z dat hledají obecné pravděpodobnostní závislosti mezi veličinami pomocí hladového prohledávání stavového prostoru modelů. Pro efektivitu algoritmu je klíčové navrhnout efektivní prohledávání takzvaného inclusion okolí modelu. Cílem práce je vytvořit efektivní implementaci algoritmu učení struktury bayesovských sítí. Důraz bude kladen na rychlost implementace - tomuto požadavku by měl odpovídat i výběr programovacího jazyka.

### **5 Učení bayesovských sítí obsahujících deterministické závislosti**

Znamé algoritmy pro učení bayesovských sítí hledají obecné pravděpodobnostní závislosti mezi veličinami. V praxi se ovšem často vyskytují speciální typy pravděpodobnostních závislostí. Častým případem jsou takzvané zašuměné funkční relace (noisy-or, noisy-and, noisy-max, noisy-min, noisy-add).

Cílem práce je navrhnout a naprogramovat algoritmus, který by byl schop naučit z dat bayesovskou síť obsahující speciální typy pravděpodobnostních závislostí mezi veličinami.

### **6 Zobecnění noisy-or klasifikátoru**

Noisy-or klasifikátor patří mezi jednoduché klasifikátory z rodiny bayesovských sítí. Ukazuje se však, že dosahuje výsledků srovnatelných s jinými metodami. Cílem práce je zobecnit klasifikátor z binárních na celočíselné veličiny a otestovat chování klasifikátoru na reálných datech - např. souboru textů agentury Reuters.