INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Název projektu: Mezinárodní centrum pro informaci a neurčitost

Registrační číslo: CZ.1.07/2.3.00/20.0060

Zpráva z účasti na stáži

Datum konání stáže:	5.2.2012 – 18.2.2012
Navštívené pracoviště:	NRU HSE, Moskva, Rusko
Zahraniční garant:	prof. Sergei O. Kuznetsov
Účastník stáže:	Mgr. Jan Outrata, Ph.D.

Stručný popis navštíveného pracoviště

Zaměřením katedry analýzy dat a umělé inteligence (Department of Data Analysis and Artificial Intelligence) při Škole aplikované matematiky a informatiky (School of Applied Mathematics and Information Science) na National Research University, Higher School of Economics (NRU HSE) v Moskvě je výzkum v oblastech modelování socioekonomických, politických a organizačních procesů a vývoj matematických modelů pro rozhodování. Zejména v souvislosti druhým uvedeným tématem je však toto pracoviště známé jako jedno z nejvýznamějších světových výzkumných center v oblasti formální konceptuální analýzy (FCA) a příbuzných oblastí. Z tohoto pracoviště jsou známé algoritmy FCA (CbO prof. Kuznetsova a AddIntent S. A. Obiedkova aj.), které patří k nejvýkonnějším svého druhu, a výsledky využití FCA v machine learning a data mining. V současné době je na katedře 15 pracovníků (z toho 4 profesoři a 2 docenti) a 5 doktorandů (1 pod vedením i z jiné univerzity). Výzkum členů katedry je financován převážně na lokální univerzitní úrovni státem a částečně také průmyslovým sektorem.

Během stáže jsem diskutoval vědecké problémy s prof. Sergeiem O. Kuznetsovem a dalšími spolupracovníky a doktorandy z jeho týmu. Prof. Sergei O. Kuznetsov se s nimi dlouhodbě věnuje FCA (algoritmy FCA a složitost problémů ve FCA) a jejího využití v data mining (biclustering, pattern structures) a machine learning/artificial intelligence (rozhodování, rough sets). V těchto oblastech dosáhli a dosahují zásadních výsledků. Granty a projekty získané skupinou prof. Kuznetsova jsou financované převážně z národních zdrojů, z menší části i z průmyslu. Skupina kolem prof. Kuznetsova a ostatní lidé z katedry spolupracují s pracovišti na univerzitách UP v Olomouci (skupina prof. Bělohlávka), TU Dresden (B. Ganter), UBP Clermont-Ferrand (L. Nourine), University of London, University of Portugal a výzkumným týmem LORIA Nancy ve Francii.

Průběh stáže

S prof. Kuznetsovem a jeho spolupracovníky a doktorandy, zejména S. A. Obiedkovem, M. Babinem a D. Ignatovem, jsme konzultovali především algoritmy pro výpočet množiny všech formálních konceptů z objekt-atributových relačních dat vyvinuté na katedře informatiky PřF UP (algoritmus FCbO, vylepšující algoritmus CbO prof. Kuznetsova, a nový algoritmus používající třídění atributů) a na jejich pracovišti (algoritmus AddIntent S. A. Obiedkova aj.). Po vysvětlení algoritmů jsme probírali zejména otázky složitosti algoritmů, jejich možná vylepšení a implementační záležitosti pro zvýšení efektivity algoritmů. Na začátku stáže bylo také dohodnuto experimentální porovnání efektivity algoritmů již během stáže (v testovacím FCA frameworku vyvinutém na jejich pracovišti), k tomu ale nakonec nedošlo, neboť v rámci férovosti porovnání by pro řešení stejné úlohy vyžadoval náš algoritmus FCbO větší úpravy. Pro vlastní porovnání na svých pracovištích mi proto byla poskytnuta implementace jejich algoritmu AddIntent S.A Obiedkovem, implementace našeho algoritmu FCbO je dostupná ke stažení na internetu.

Dále jsme diskutovali možnosti využití Boolovské faktorové analýzy (BFA) dat pro FCA v jejich výzkumu v oblasti doporučovacích systémů (recommendation systems) s využitím FCA. Se spolupracovníkem A. Neznanovem jsem ke konci stáže konzultoval ještě výzkum metod kreslení svazů a byla dohodnuta výměna dosavadních výsledků a další společný výzkum v této oblasti.

Zúčastnil jsem se také přednášky doc. Vychodila z katedry informatiky PřF UP, který byl na stejném pracovišti v prvním týdnu stáže také na stáži (z jiného projektu), na jejich vědeckém semináři, následné diskuze a také části prezentace o implementaci jejich testovacího FCA frameworku.

Rovněž jsme s prof. Kuznetsovem nezávazně probírali plánované vzájemné pobyty kolegů z katedry informatiky PřF UP (Bartl, Macko) na NRU HSE a jeho doktorandů (J. Poelmans, D. Ignatov aj.) na katedře informatiky PřF UP. Návštěva J. Poelmanse je plánovaná již na začátek března 2012.

Přednášky na navštíveném pracovišti

Na NRU HSE jsem přednesl celkem tři přednášky: První, o indukci rozhodovacích stromů z konceptuálního svazu, pro postgraduální studenty ve výuce prof. Kuznetsova, další dvě, o použití formálních konceptů (boolovských faktorů) jako nových atributů pro machine learning a o rychlé faktorizaci podle podobnosti ve FCA s fuzzy atributy, na vědeckém semináři 'Mathematical Models of Information Technologies' (http://ami.hse.ru/en/seminars). V komentářích po přednáškách zazněly souvislosti s existujícími výsledky a náměty na další výzkum, zvláště po třetí přednášce následovala bouřlivá diskuze nad návrhy posluchačů k dalšímu výzkumu v prezentované problematice. K problematice prvních dvou přednášek mi byly poskytnuty relevantní reference na související výsledky.

1) *Inducing decision trees via concept lattices*

2) *Formal concepts as attributes for machine learning*

Abstrakt: We present an utilization of formal concept analysis (FCA) in input data preprocessing for machine learning. The preprocessing consists in using formal concepts as

new attributes which replace the original attributes describing objects/records in input data. The formal concepts used are so-called factor concepts obtained by boolean factor analysis, decribed by means of FCA. The preprocessing means reduction of dimensionality of data since the number of factor concepts is smaller than the number of original attributes. We will demonstrate the idea on decision tree induction as the most commonly used method of machine learning.

3) *Fast factorization by similarity in FCA of data with fuzzy attributes*

Abstrakt: The results on fast factorization by similarity in formal concept analysis (FCA) of data with fuzzy attributes, achieved at Dept. Computer Science, Palacky University in Olomouc, Czech Republic, will be presented. The basic idea of factorization by similarity is to have, instead of a possibly large original fuzzy concept lattice, its factor lattice. The elements of the factor lattice are maximal blocks of clusters which are pairwise similar to degree exceeding a user-specified threshold. The factor lattice thus represents an approximate version of the original concept lattice. Two ways to compute the factor lattice directly from input data (without the need to compute the possibly large original concept lattice) will be presented.

Publikace rozpracované během stáže

Nebyly rozpracované žádné publikace.

Navázání kontaktů

Stáž významně přispěla k posílení a prohloubení vědecké spolupráce se skupinou prof. Kuznetsova. Byl diskutován stávající výzkum nad společnými tématy s vytipováním konkrétních bodů spolupráce (algoritmy, BFA) a navržena nová témata pro další vzájemnou spolupráci (kreslení svazů).

Shrnutí stáže

Stáž významně splnila svůj účel. Byla posílena vědecká spolupráce s jedním z nejvýznamnějších center v oblasti formální konceptuální analýzy (FCA) a příbuzných oblastí (machine learning). Na stáži byly předneseny tři přednášky, pro postgraduální studenty a na vědeckém semináři, po kterých následovaly živé diskuze s návrhy dalšího výzkumu.

Fotografická dokumentace





Fotografie z vědeckého semináře, na snímku vlevo prof. Kuznetsov (vlevo) a dr. Outrata (vpravo), na snímku vpravo po přednáškách dr. Outrata.