



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Mezinárodní konference CLA 2013, La Rochelle, Francie

Jan Outrata

15.10.2013 – 18.10.2013



The Tenth Int. Conference on Concept Lattices and Their Applications

- jedna z hlavních konferencí v oblasti Formální konceptuální analýzy, konceptuálních svazů a jejich aplikací
- přední osobnosti z této oblasti výzkumu analýzy a zpracování relačních dat zde prezentují a diskutují svůj nejnovější výzkum:
 - Radim Belohlavek** (Palacký University, Olomouc, Czech Republic)
 - Vincent Duquenne** (Université Pierre et Marie Curie, Paris, France)
 - Sergei O. Kuznetsov** (State University HSE, Moscow, Russia)
 - Engelbert Mephu Ngifo** (LIMOS, University Blaise Pascal, Clermont-Ferrand, France)
 - Amedeo Napoli** (INRIA NGE/LORIA, Nancy, France)
- konference byla v minulosti dvakrát pořádána také katedrou informatiky PŘF UP

CLA 2013 – zajímavá čísla

Acceptance rate:	59.46 %
Počet účastníků:	kolem 40
Počet prezentovaných příspěvků:	27 (22 článků, 5 posterů)
Počet zemí autorů příspěvků:	13
Počet zvaných přednášek:	4
Počet paralelních sekcí:	1

Zvané přednášky (1/4)

Ralph Freese (University of Hawai): Projective Lattices.

- přednášející známý především jako autor používané metody a software na kreslení svazů
- v první části přednášky představen software UACalc pro práci se svazy
- ve druhé části již téma přednášky – projektivní svazy a jejich vlastnosti; jedna z charakterizujících podmínek projektivních svazů je tzv. D-relace, která je využívána na více dalších místech teorie svazů jako např. volné nebo konečně reprezentovatelné svazy, s aplikacemi ve svazech kongruencí nebo v teorii databází
- abstrakty přednášek jsou dostupné na <http://cla.inf.upol.cz/papers.html> a <http://ceur-ws.org/Vol-1062/>

Zvané přednášky (2/4)

Vincent Duquenne (Université Pierre et Marie Curie, Paris): Some applications of Lattice Analysis (1983-2013).

- přednášející se výzkumu v oblasti FCA věnuje od počátku formalizace metody na začátku 80. let a patří zde mezi nejznámější přední odborníky
- řada poznámek a doporučení z raného vývoje metody zejména k prezentaci a aplikaci konceptuálních svazů, ve kterých by se mělo více využívat svazově teoretických strukturálních vlastností jako dekompozice pomocí svazových kongruencí a relací tolerance, a k algoritmům pro výpočet konceptuálního svazu a atributových implikací, kde by měla být snaha o algoritmy provádějící co nejméně výpočtů nad minimálně reprezentovanými daty
- těchto doporučení se držíme při vývoji metody na kreslení (konceptuálních) svazů a algoritmů na jejich výpočet a poznámky jsou tak pro nás velmi cenné
- abstrakty přednášek jsou dostupné na <http://cla.inf.upol.cz/papers.html> a <http://ceur-ws.org/Vol-1062/>

Zvané přednášky (3/4)

Michel Grabisch (Université Paris I): Cooperative Games on Lattices.

- teorie kooperativních her nad distribuovanými svazy jako koalicemi hráčů, mezi kterými existuje hierarchické uspořádání nebo precedenční omezení
- jako řešení hry prezentován koncept jádra (core), zajišťující stabilitu hry, a zejména jeho konečného ohraničení; prezentovány byly i výsledky nad uzávěrovými svazy a představeny byly také hry nad konceptuálními svazy a prvotní výsledky k jádru
- přednášející byl zvaným přednášejícím také na letošní letní škole a workshopu pořádaných v rámci projektu MCIN

- abstrakty přednášek jsou dostupné na <http://cla.inf.upol.cz/papers.html> a <http://ceur-ws.org/Vol-1062/>

Zvané přednášky (4/4)

Bart Goethals (University of Antwerp): Cartification: from Similarities to Itemset Frequencies.

- přednášející známý svými algoritmy v oblasti data mining
- nová metoda transformace vícedimenzionálních dat na ménědimenzionální data pomocí metody analýzy nákupního košíku (tzv. kartifikace)
- metoda spočívá ve vytvoření popisu objektů pomocí jemu nejpodobnějších objektů – nad takto transformovanými daty je pak pomocí dostupných metod vyhledávání frekventní množiny prvků (frequent itemset mining) snadnější provádět shlukování

- abstrakty přednášek jsou dostupné na <http://cla.inf.upol.cz/papers.html> a <http://ceur-ws.org/Vol-1062/>

Vybrané řádné přednášky (1/4)

Jaume Baixeries, Mehdi Kaytoue, Amedeo Napoli: Computing Similarity Dependencies with Pattern Structures.

Představeny atributové závislosti založené na relacích podobnosti mezi hodnotami atributů ($X \rightarrow Y$ pro množiny atributů X, Y je similarity dependency, když pro všechny objekty x_1, x_2 platí když $r_X(x_1, x_2)$, pak $r_Y(x_1, x_2)$, kde $r_X(x_1, x_2)$ je když pro všechny atributy y z X platí $r_y(I(x_1, y), I(x_2, y))$, kde r_y je relace podobnosti mezi hodnotami atributu y) a jejich charakterizace a výpočet pomocí struktur vzorů (pattern structures).

Alexandre Bazin, Jean-Gabriel Ganascia: Enumerating Pseudo-Intents in a Partial Order.

Představen algoritmus výpočtu pseudointentů (a Guigues-Duquenne a Luxenburger bází atributových implikací) využívající svazovou strukturu množiny intentů a pseudointentů místo tradičního lineárního uspořádání atributů (lectic order) algoritmem NextClosure. Novým výpočtem se počítá méně uzávěrů.

Vybrané řádné přednášky (2/4)

Marketa Krmelova, Martin Trnecka: Boolean Factor Analysis of Multi-Relational Data.

Představena Booleovská faktorová analýza (BFA) multirelačních dat (více relačních dat se vztahy mezi sebou) založená na BFA řešené pomocí FCA. Faktory multirelačních dat jsou tvořeny propojením faktorů jednotlivých relačních dat (na základě vztahů mezi relačními daty), což je hlavní rozdíl oproti RCA (Relational formal concept analysis, FCA na multirelačních datech).

Cristina Alcalde, Ana Burusco, Juan Carlos Díaz, Ramon Fuentes-González, Jesús Medina: On information retrieval in morphological image and signal processing.

Představena metoda získání původních obrazů a signálů po přenosu zašuměným kanálem pomocí fuzzy property-oriented konceptuálních svazů (neboli konceptuálních svazů založených na rough sets) a fuzzy relačních rovnic.

Vybrané řádné přednášky (3/4)

Ondrej Krídlo, Patrik Mihalčín, Stanislav Krajčí, Lubomír Antoni: Formal Concept Analysis of higher order.

Představeny formální kontext druhého řádu, s objekty a atributy jako disjunktivními sjednoceními objektů a atributů jiných formálních kontextů, koncept-formující operátory nad tímto kontextem a výpočet formálních konceptů druhého řádu (s objekty a atributy jako formálními koncepty jiných kontextů) pomocí bondů mezi jinými kontexty nebo pomocí metod heterogenních formálních kontextů. V závěru bylo rovněž ukázáno, jak jsou tímto zobecněny homogenické fuzzy formální kontexty a jejich koncept-formující operátory. Článek byl oceněn cenou Best paper award.

Jan Konečný, Manuel Ojeda-Aciego: Isotone L-bonds.

Představeny vlastnosti L-bonds (morfismus mezi formálními fuzzy kontexty) pro FCA s fuzzy atributy s isotonními koncept-formujícími operátory, výpočet L-bonds pomocí direktních součinů a vztah k L-bonds pro FCA s antitonními koncept-formujícími operátory.

Vybrané řádné přednášky (4/4)

Cynthia Vera Glodeanu: Attribute exploration with fuzzy attributes and background knowledge.

Představeno rozšíření metody attribute exploration s předchozí znalostí (background knowledge) některých atributových implikací z dvouhodnotové logiky do fuzzy logiky.

Ondrej Krídlo, Manuel Ojeda-Aciego: CRL-Chu correspondences.

Představeno zobecnění L-Chu korespondence a L-bonds (morfismy mezi formálními fuzzy kontexty) na základě isotonních Galoisových konexí zachovávajících residuum. Zobecnění byla ilustrována na jednoduchém příkladu ze školského prostředí.

- shrnutí dalších přednášek jsou ve Zprávě z účasti na konferenci (všech 22)
- články jsou dostupné na <http://cla.inf.upol.cz/papers.html> a <http://ceur-ws.org/Vol-1062/>

Postery

Bruce R. Barkstrom: Applying User-Guided, Dynamic FCA to Navigational Searches for Earth Science Data and Documentation.

Klasifikace a navigace v archivu Earth Science Data pomocí FCA.

Pablo Cordero, Manuel Enciso, Angel Mora, Manuel Ojeda-Aciego: Computing Left-Minimal Direct Basis of implications.

Metoda výpočtu z množiny všech implikací mezi uzávěrem a jeho min. generátory.

Lenka Pisková, Tomáš Horváth: Comparing Performance of Formal Concept Analysis and Closed Frequent Itemset Mining Algorithms on Real Data.

Experimentální porovnání výkonu algoritmů na datech z UCI ML a FIMI repozitářů.

- články jsou dostupné na <http://cla.inf.upol.cz/papers.html> a <http://ceur-ws.org/Vol-1062/>, shrnutí dalších jsou ve Zprávě z účasti na konferenci (všech 5)

Vlastní prezentace

*Jan Outrata: A lattice-free concept lattice update algorithm based on *CbO.*

Představeny nové rychlé algoritmy pro

- 1 aktualizaci množiny všech formálních konceptů objekt-atributových binárních relačních dat při změně těchto dat (přidání nových a odebrání nebo modifikace existujících objektů nebo atributů) s výpočtem pouze nových/odebraných a modifikovaných konceptů a
- 2 zjištění relace uspořádání konceptů, tj. výpočet konceptuálního svazu.

Dohromady algoritmy tvoří algoritmus pro aktualizaci konceptuálního svazu, pouze ze vstupních dat, bez potřeby svazu před změnou dat. Algoritmy jsou vytvořeny jako modifikace, resp. rozšíření, algoritmu FCbO pro výpočet množiny všech formálních konceptů.

Shrnutí

- prezentovány příspěvky z těchto oblastí: teorie, algoritmy (včetně experimentálního porovnání), aplikace a software (fuzzy) FCA, atributové závislosti a implikace, pattern structures, data mining a konceptuální KDD aj.
- možné směry dalšího výzkumu ve FCA a souvisejících oblastech lze usuzovat morfizmy mezi fuzzy kontexty (L-bonds) nebo RCA (FCA na multirelačních datech)
- dlouhodobě perspektivní jsou pattern structures a konceptuální KDD

Navázání kontaktů

- *Vincent Duquenne (Université Pierre et Marie Curie)*
 - konzultace algoritmů výpočtu konceptuálního svazu a reprezentaci dat pro výpočet
- *Marianne Huchard*
 - společná konzultace s M. Krmelovou a M. Trnečkou Booleovské faktorové analýzy a jejího využití v jejím výzkumu
- *Bruce R. Barkstrom*
 - diskuze o možných aplikacích FCA a Boolovské faktorové analýzy v jeho výzkumu
- většinu lidí z této konference znám z dřívějších ročníků konference nebo spolupráce





INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Mezinárodní konference

IEEE ICDM 2013, Dallas, Texas, USA

Jan Outrata

7.12.2013 – 10.12.2013



IEEE 13th International Conference on Data Mining

- přední konference v mnoha oblastech data mining a machine learning jako např. feature analysis, klasifikace, strojové učení, shlukování, sumarizace, doporučování big data, bioinformatika a medicínská informatika, bussiness intelligence, analýza časových řad, reprezentace dat a dolování dat ve videu, obrazu a textu
- hlavní a nejcitovanější osobnosti z této oblasti informatiky:
Hans-Peter Kriegel (Universität Munich, Germany)
Geoff Webb (Monash University, Melbourne, Australia)
- sponzorována sdruženími IEEE a IEEE Computer Society a grantovými agenturami a velkými formami (letos USA National Science Foundation a IBM a Toshiba)
- mezi účastníky jsou jak vědci z akademické sféry i z výzkumných oddělení vůdčích firem

IEEE ICDM 2013 – zajímavá čísla

Acceptance rate:	19.65 %
Počet účastníků:	přes 500
Počet prezentovaných příspěvků:	187 (94 regular papers, 65 short papers, 18 workshopů, 7 demonstrací, 3 tutoriály, 1 panel, 1 soutěž a 1 PhD fórum)
Počet zemí autorů příspěvků:	55
Počet plenárních přednášek:	3
Počet paralelních sekcí:	4 (u workshopů 11)

Plenární přednášky (1/3)

Alexander Tuzhilin (NYU): Opportunities and Challenges Facing Recommender Systems: Where Can We Go From Here?.

- některé málo probádané směry vývoje doporučovacích (recommender) systémů, které dle jeho názoru představují slibné příležitosti pro další výzkum této oblasti
- výzvy, které je třeba zvládnout k překonání omezení stávajících systémů a vytvoření nové generace doporučovacích systémů

- abstrakty přednášek jsou dostupné na <http://icdm2013.rutgers.edu/keynotes>

Plenární přednášky (2/3)

Joydeep Ghosh (University of Texas, Austin): Predictive Healthcare Analysis under Privacy Constraints.

- nejdříve náhled otázky ochrany soukromí vs. využitelnost údajů v kontextu zdravotní péče
- pak prezentace dvou jejich přístupů k prediktivnímu modelování zachovávajícímu ochranu soukromí, u nichž je malá degradace modelů navzdory omezením sdílení a analýzy údajů
- první přístup predikuje z agregovaných údajů, zatímco druhý, nový, představuje generátor realistických, ale ne reálných, dat z původních chráněných údajů
- abstrakty přednášek jsou dostupné na <http://icdm2013.rutgers.edu/keynotes>

Plenární přednášky (3/3)

Jianchang (JC) Mao (Microsoft): Large-scale Learning in Computational Advertising.

- několik velkých problémů strojového učení z velkých dat (big data) v oblastech reklamy, porozumění dotazu, uživateli a dokumentu, modelování relevance výsledků dotazu, predikce uživatelské reakce, doporučování klíčových slov apod.
- nedávná řešení několika problémů, zejména shlukování dotazu (EM metoda založená na KL-divergenci dotazů) a doporučování klíčových slov (multi-label random forest learning)
- abstrakty přednášek jsou dostupné na <http://icdm2013.rutgers.edu/keynotes>

Vybrané řádné přednášky (1/4)

A. Dukkupati, G. Pandey, D. Ghoshdastidar, P. Koley, D.M.V. Satya Sriram: Generative Maximum Entropy Learning for Multiclass Classification.

Představena metoda klasifikace využívající maximální entropii s výběrem vlastností (feature selection). Pro redukci dimenzionality je využito “maximální diskriminace” (Jeffrey nebo Jensen-Shannon divergence) mezi hustotami podmíněné pravděpodobnosti tříd objektů.

E. Aksehirli, B. Goethals, E. Müller, J. Vreeken: Cartification: A Neighborhood Preserving Transformation for Mining High Dimensional Data.

Představena transformace objektů na vektory atributů, které reprezentují okolí objektu. Transformace zachovává podobnosti mezi objekty při více pohledech na data. Podle experimentů dosahují frequent itemset mining metody lepších výsledků nad takto transformovanými daty.

Vybrané řádné přednášky (2/4)

N. Tatti: Itemsets for Real-Valued Datasets.

Představena metoda dolování vzorů (pattern mining) pro reálně-hodnotová data pomocí transformace na binární data, kde prahy jsou chápány jako náhodné veličiny a support vzorů je průměrný.

B. Micenková, R. T. Ng, X.-Hong Dang, I. Assent: Explaining Outliers by Subspace Separability.

Představena metoda určení možných vysvětlení nalezených tzv. outliers, tj. mimořádných objektů v datech, např. chyb. Vysvětlení, užitečná pro interpretaci či validaci outliers, jsou vyjádřena formou podprostorů, ve kterých jsou outliers separované od inliers.

Vybrané řádné přednášky (3/4)

R. Belohlavek, M. Krmelova: Beyond Boolean Matrix Decompositions: Toward Factor Analysis and Dimensionality Reduction of Ordinal Data.

Představena dekompozice matic s ordinálními daty založená na formální konceptuální analýze a algoritmus dekompozice. Technicky metoda spočívá v nahrazení dvouhodnotové Booleovy algebry pro binární data obecnějšími strukturami, které ale přináší nové problémy.

N. Djuric, S. Vucetic: Efficient Visualization of Large-Scale Data Tables through Reordering and Entropy Minimization.

Představena metoda, EM-ordering, pro přeskládání řádků a sloupců tabulkových dat podle jejich podobností tak, aby tzv. heatmap (vizuální inspekce dat) poskytovala hlubší pohled na data. Uspořádání minimalizuje entropii prediktivního kódování uspořádané tabulky dat. Pro řešení se využívá heuristické řešení problému obchodního cestujícího, kde řádky tabulky jsou města.

Vybrané řádné přednášky (4/4)

R. Shams, R. E. Mercer: Classifying Spam Emails with Text and Readability Features.

Představena klasifikace spamu používající atributy založené na přirozeném jazyku emailu (angličtině) a čitelnosti, v kombinaci s tradičními atributy založenými na obsahu emailu.

D. Erdős, P. Miettinen: Walk 'n' Merge: A Scalable Algorithm for Boolean Tensor Factorization. Představen algoritmus pro Booleovskou tensorovou faktorizaci, tj. faktorizaci binárních dat na binární faktory při zachování binárních rekonstruovaného tensoru dat. Algoritmus provádí CP a Tucker dekompozici.

- shrnutí dalších navštívených přednášek jsou ve Zprávě z účasti na konferenci (celkem 16)

Vybrané prezentace na workshopech a tutoriálu

E. M.L. Peters, G. Dedene, J. Poelmans: Empirical Discovery of Potential Value Leaks in Processes by means of Formal Concept Analysis (workshop EEML: Experimental Economics and Machine Learning).

Využití metod machine learning v teorii lidských strategických interakcí.

P. Jiang, M. T Health: Pattern Discovery in High Dimensional Binary Data (workshop HDM: High Dimensional Data Mining).

Nové směry v řešení problémů zpracování dat s až miliony atributů, kde tradiční metody selhávají.

H. Tong, F. Wang, Ch. Ding: Applied Matrix Analytics: Recent Advance and Case Studies 1. (tutoriál).

Přehled nových maticových data mining algoritmů s aplikacemi v sociální oblasti a zdravotní péči.

Další prezentace na workshopech, PhD fórum a panel

- shrnutí dalších prezentací z workshopů, které mě zaujaly, jsou ve Zprávě z účasti na konferenci (celkem 7)
- demonstrace, které jsem se chtěl zúčastnit, neproběhla
- podíval jsem se i na některé příspěvky na PhD fóru a také na panel o dolování velkých dat (big data)

Shrnutí

- prezentovány průlomové příspěvky zejména z těchto oblastí: klasifikace a strojové učení (hlavně rozhodovací stromy), feature selection a transformace atributů, doporučovací (recommender) systémy a prediktivní modelování. Významnými tématy byly také dolování vzorů (pattern mining) a dekompozice (faktorizace) matic
- lze usuzovat, že perspektivními směry dalšího výzkumu jsou: klasifikace a strojové učení s učitelem (supervised) i bez něj (unsupervised, semi-supervised), feature selection a extraction a prediktivní modelování s aplikacemi v doporučovacích (recommender) systémech a částečně také v oblasti dolování vzorů (pattern mining) dekompozice matic
- tedy témata spíše z oblasti machine learning než data mining

