



evropský
sociální
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání
pro konkurenceschopnost

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 konferenci

název konference: MLDM 2012 (8th International Conference on Machine Learning and Data Mining)
datum konání: 16. 7. - 20. 7. 2012
místo: Berlín, Německo
účastník konference: Mgr. Tomáš Kühn

Stručný popis konference:

MLDM je významnou mezinárodní konferencí, která se zabývá tématy související s oblastmi machine learningu a data miningu, a to jak z teoretického, tak aplikovaného pohledu. Konference se koná každoročně a bývá součástí kongresu DSA (The Frontiers in Intelligent Data and Signal Analysis).

Zajímavá čísla

Acceptance rate:	33,5 %
Počet účastníků:	NA
Počet prezentovaných příspěvků:	71 (z toho 8 posterů)
Počet plenárních přednášek:	4
Počet paralelních sekcí	1

Zajímavé přednášky

A. Plenární přednášky

Graham Williams: Aspects of Data Mining.

Přednáška se zabývala informačními systémy pro správu daní a daňových přiznání v Austrálii. Problém v principu spočívá v tom, že je ekonomicky náročné, aby jednotlivá přiznání kontrolovali experti úřadu. Za využití technik data miningu se již podařilo s vysokou úspěšností oddělit podezřelá přiznání od těch bezproblémových a snížit tak počet přiznání, která musí zpracovávat experti.

Xiaoqing Ding: Face Recognition.

Přednáška se zabývala problematikou detekce postavy, jejího obličeje a rozpoznání tohoto obličeje vzhledem k datům (obličejům) v databázi. Byl představen sofistikovaný software, který umožňuje v reálném čase rozpoznat v záběru kamery

člověka, zpracovat jako obličej (odstranit brýle a jiné předměty) a rozpoznat jej ve své databázi. Software je možné využít při vstupech do metra nebo na letištích.

Hualan Zhang: Quantitative Measurement of Cellular Events.

Přednáška se věnovala odborné studii, ve které vědecký tým vedený autorem, zkoumal možnosti rozpoznání a následného spočítání velmi podobných typů buněk za pomoci kombinace několika algoritmů z oblasti data miningu. Cílem bylo alespoň z části zautomatizovat vyšetření prováděná v nemocnicích a uvolnit tak čas expertů pro náročnější případy.

Patrick S. P. Wang: Intelligent Pattern Recognition and Applications - An Interactive E-Learning System, Modeling and Simulation.

Přednáška představila sofistikovaný systém pro rozpoznávání vzorů v datech a jeho možná využití v e-learningu, modelování a provádění nejrůznějších simulací. Představený systém vykázal v některých studiích srovnatelné a v dalších dokonce lepší výsledky než stávající algoritmy pro danou problematiku.

B. Řádné přednášky

Matias Fernando Garcia-Constantino, Frans Coenen, P-J Noble, Alan Radford, and Christian Setzkor: A Semi-Automated Approach to Building Text Summarisation Classifiers.

Príspevek pojednával o nové poloautomatické metodě extrakce informací z textů a jejich následnou automatickou sumarizaci. Autoři představili také provedené experimenty a výsledky svého přístupu srovnali s alternativními technikami.

Guang Jiang, Xi Liu, and Zhi Zhong Shi: Multi-label Image Annotation Based on Neighbor Pair Correlation Chain.

Přednáška se zabývala novou metodou pro anotování obrázkových dat. Představená metoda je založena na využití bipartitních grafů. Byly představeny také experimenty a srovnání s běžně používanými algoritmy.

Muhammad Sulaiman Khan, Frans Coenen, Clare Dixon, and Subhieh El-Salhi: Finding Correlations Between 3-D Surfaces: A Study In Asymmetric Incremental Sheet Forming.

Autoři představili studii týkající se problematiky klasifikace lokální geometrie tělesa. Cílem studie bylo zefektivnit výrobu jistého typu kovových součástek.

Sahar Changuel and Nicolas Labroche: Content independant metadata production as a machine learning problem.

Přednáška pojednávala o problematice automatické extrakce metadat s využitím asociačních pravidel. Autoři provedli studii na datech obsahujících popisy publikací v knihovně.

Jianfeng Si, Qing Li, Tieyun Qian, and Xiaotie Deng: Discovering K Web User Groups with Specific Aspect Interests.

Přednáška se týkala problematiky analýzy online recenzí s využitím metadat týkajících se uživatelů, kteří tyto recenze psali.

Pavel Turkov, Olga Krasotkina, and Vadim Mottl: Bayesian approach to the concept drift in the pattern recognition problems.

Příspěvek se zabýval představením nového matematického a algoritmického přístupu k problematice rozpoznání vzorů v datech.

Michelangelo Ceci, Annalisa Appice, Herna L. Viktor, Donato Malerba, Hongyu Guo, and Eric Paquet: Transductive Relational Classification in the Co-Training Paradigm.

Autoři prezentovali novou metodu pro automatické generování pohledů z multirelační databáze obsahující data s popisky i bez nich a její využití pro tzv. transduktivní učení.

László Kovács: Reduction of distance computations in selection of pivot elements for balanced GHT structure.

Přednáška pojednávala o problematice výběru pivota v algoritmech týkajících se metrických prostorů. Autoři představili novou metodu a srovnali ji se stávajícími.

Lucrezia Macchia, Michelangelo Ceci, and Donato Malerba: Mining Ranking Models from Dynamic Network Data.

Příspěvek se zabýval novými modely při práci se senzorovými sítěmi, cílem bylo zlepšit efektivitu využívání těchto typů sítí.

Iwan Syarif, Ed Zaluska, Adam Prugel-Bennett, and Gary Wills: Application of Bagging, Boosting and Stacking to Intrusion Detection.

Prezentace popisovala možná vylepšení klasických algoritmů pro klasifikaci dat. Konkrétně autoři představili metody Bagging, Boosting a Stackingu.

Aixiang Li, Makoto Haraguchi, and Yoshiaki Okubo: Top-N Minimization Approach for Indicative Correlation Change Mining.

Přednáška se zabývala problematikou nalezení nekorelovaných vzorů v transakční databázi.

Rui Leite, Pavel Brazdil and Joaquin Vanschoren: Selecting Classification Algorithms with Active Testing.

Příspěvek se týkal nové metody inteligentního výběru klasifikačního algoritmu s využitím testu křížovou validací.

Amit Thombre: Comparing logistic regression, neural networks, C5.0 and M5' classification techniques.

Příspěvek se zabýval srovnávací studií několika vybraných klasifikačních technik, konkrétně se autoři věnovali zejména logistické regresi a několika typům neuronových sítí.

Upendra Sapkota, Barrett Bryant, and Alan Sprague: Unsupervised Grammar Inference using the Minimum Description Length Principle.

Přednáška představila novou metodu tzv. Minimum Description Length Principle pro zjednodušení popisu dat prostřednictvím kontextově nezávislé gramatiky.

Thais Mayumi Oshiro, Pedro Santoro Perez, and José Augusto Baranauskas: How Many Trees in a Random Forest?

Přednáška pojednávala o možnostech nastavení algoritmu Random Forest pro zpracování rozsáhlé datasety.

Hai Thanh Nguyen and Katrin Franke: A General L_p -norm Support Vector Machine via Mixed 0-1 Programming.

Prezentace se zabývala možnostmi nastavení algoritmu Support Vector Machine. Autoři se zabývali tzv. L_2 normovanými SVM, ale předpokládají, že se jejich výsledky dají zobecnit i pro L_p .

Dieter William Joenssen and Udo Bankhofer: Hot Deck Methods for Imputing Missing Data -The Effects of Limiting Donor Usage.

Prezentace se zabývala problematikou práce s neúplnými daty. Autoři představili několik metod a jejich vzájemné srovnání.

Saket Bharambe, Harshit Dubey and Vikram Pudi: BINDER: Binary search based efficient regression.

Autoři představili novou metodu pro účinné binární prohledávání, která využívá regresní algoritmy a vykazuje nízkou složitost ve srovnání s alternativními metodami.

Kartick Chandra Mondal, Nicolas Pasquier, Anirban Mukhopadhyay, Ujjwal Maulik, and Sanghamitra Bandhopadhyay: A New Approach for Association Rule Mining and Bi-clustering using Formal Concept Analysis.

Prezentace se zabývala několika problémy, které se často vyskytují v oblasti bioinformatiky. Autoři představili nový algoritmus FIST.

Tao Xu, David Chiu, and Iker Gondra: Constructing Target Concept in Multiple Instance Learning Using Maximum Partial Entropy.

Příspěvek se zabýval problematikou grupování instancí, byla představena nová metoda využívající maximalizaci parciální entropie.

Nikolaos Pitelis and Anastasios Tefas: Discriminant Subspace Learning Based on Support Vectors Machines.

Přednáška představila novou metodu učení klasifikátorů, která je založená na principu metody Support Vector Machine.

Hoa Thi Nong and The Duy Bui: A new learning strategy of general BAMs.

Příspěvek pojednával o možnostech využití obousměrné asociativní paměti, neuronové síti složené ze dvou Hopfieldových sítí. Autoři ukázali také experimenty a srovnání s jinými metodami.

Michael Glodek, Friedhelm Schwenker, and Günther Palm: Detecting Actions by Integrating Sequential Symbolic and Sub-Symbolic Information in Human Activity Recognition.

Autoři prezentovali metodu pro rozpoznávání činnosti člověka na základě informací o poloze těla. Software, který autoři vyvinuli, využívá systém Kinect.

Davide Maiorca, Giorgio Giacinto, and Iginio Corona: A Pattern Recognition System for Malicious PDF Files Detection.

Príspevek se zabýval návrhom a implementáci systému pro detekci potenciálně nebezpečných PDF souborů na základě specifických vzorů v jejich metadatech.

Konstantinos Kalpakis, Shiming Yang, Peter F. Hu, Colin Mackenzie, Lynn Stansbury, Deborah M. Stein, and Thomas M. Scalea: Outcome prediction for patients with severe traumatic brain injury using permutation entropy analysis of electronic vital signs data.

Přednáška popisovala možnosti predikce úspěšnosti léčby pacientů s traumatickým zraněním mozku a to jak z pohledu přežití, tak z pohledu kvality dalšího života.

Tomasz Staroszczyk, Stanislaw Osowski, and Tomasz Markiewicz: Comparative analysis of feature selection methods for blood cell recognition in leukemia.

Príspevek se zabýval analýzou vlivu výběrové metody na úspěšnost rozpoznání krevních buněk člověka trpícího leukemií.

Henry Joutsijoki and Martti Juhola: DAGSVM vs DAGKNN: An Experimental Case Study with Benthic Macroinvertebrate Dataset.

Autoři prezentovali srovnávací studii dvou významných dataminingových metod: DAGSVM (Directed Acyclic Graph Support Vector Machine) a DAGKNN (Directed Acyclic Graph K-th Nearest Neighbor).

Andreia Silva and Cláudia Antunes: Semi-Supervised Clustering - A Case Study.

Přednáška se týkala studie efektivity vybraných metod tzv. Semi-Supervised Clusteringu. Jako testovací příklad byl použit dataset s filmy a predikování které z nich mohou získat v budoucnu nějakou cenu.

Charlie Isaksson, Margaret H. Dunham, and Michael Hahsler: SOSStream: Self Organizing Density-Based Clustering Over Data Stream.

Príspevek predstavil nový algoritmus SOSStream pro práci s datsety obsahujícími datové proudy, který využívá automatickou adaptaci prahů.

Minh Thuy Ta, Hoai An Le Thi, and Lydia Boudjeloud Assala: Clustering Data Stream by a sub-window approach using DCA.

Autoři představili nový algoritmus nazvaný Difference of Convex function Algorithm, jedná se o velmi účinný klastrovací nástroj využívající rozdělení dat do oken.

Mihai Vlase, Dan Munteanu, and Adrian Istrate: Improvement of K-means Clustering Using Patents Metadata.

Prezentace se zabývala novou technikou pro zpracování metadat, který využívá tzv. Custom Patern Model.

Godfried Toussaint and Constantin Berzan: Proximity-Graph Instance-Based Learning, Support Vector Machines, and High Dimensionality: An Empirical Comparison.

Přednáška se zabývala otázkou jaký graf pro znázornění blízkosti instancí je vhodnější používat v rámci algoritmu Support Vector Machine.

Mariusz Borawski and Dariusz Frejlichowski: An Algorithm for the Automatic Estimation of Image Orientation.

Přednáška představila nový účinný algoritmus pro automatické odhadování orientace fotografie založený na nalezení oblohy.

Luca Piras, Giorgio Giacinto and Roberto Paredes: Enhancing image retrieval by an Exploration-Exploitation approach.

Příspěvek se týkal nové metody vylepšující úspěšnost získání informací z obrazových datasetů.

P. Forczmanski and D. Frejlichowski: Classification of Elementary Stamp Shapes by Means of Reduced Point Distance Histogram Representation.

Autoři představili algoritmus pro rozpoznávání standardních tvarů razítek na dokumentech. Byli schopni detekovat jednotlivé tvary i u neúplných otisků nebo u otisků razítek přes jiný text.

Alberto Diez and Alberto Carrascal: A Multiclassifier Approach for Drill Wear Prediction.

Přednáška se zabývala provedenou studií týkající se využití více klasifikátorů a hlasovacích kritérií pro snížení opotřebování hlavy vrtné soupravy.

Carlos Herrera-Yagüe and Pedro J. Zufiria: Prediction of telephone user attributes based on network neighborhood information.

Příspěvek pojednával o možnosti získání informací o vlastnostech osob podle informací týkající telefonické komunikace s jinými známými osobami.

Další prezentace (panely, demonstrace, postery)

V rámci konference MLDM 2012 byly prezentovány také následující postery: Classification of Arrhythmic ECG Data Using Machine Learning Techniques, Fuzzy Logistic Model Tree for Ordinal Data Prediction, Educational News Articles Classification Based on Kolmogorov Complexity Theory and Machine Learning Approach, A Class Method of Blind Source Extraction Using Second-order Statistics, Fast Nystrom for Low Rank Approximation, Machine Learning Approach for Study Course Comparison, Soccer team formation recognition using hierarchical fuzzy clustering, Possibilistic Clustering and Feature Weighting based on Robust Modeling of Finite Generalized Dirichlet Mixture.

Shrnutí konference (perspektivní témata apod.)

Na konferenci byly prezentovány jak teoretické příspěvky z oblastí klasifikace, klastrování, asociačních pravidel a patern miningu, tak algoritmy a aplikované výsledky na datech obsahujících texty, obrázky, video či webové stránky.

Navázání kontaktů

Na konferenci jsem neformálně projednával vědecké problémy s následujícími účastníky: Petra Perner, Michael Glodek.

Fotografická dokumentace



Tomáš Kühr s hlavní organizátorkou kongresu Petrou Perner sledují úvodní přednášku Grahama Williamse.

Přílohy

Program konference