10th International French-speaking conference on Knowledge Discovery and Management (EGC2010): conference report

Sadok Ben Yahia
Faculté des Sciences de Tunis
Campus Universitaire
1060, Tunis, Tunisia
sadok.benyahia@fst.rnu.tn

Jean-Marc Petit
Université de Lyon – CNRS
INSA Lyon – LIRIS UMR5205
7 av, Jean Capelle
69621 Villeurbanne Cedex, France
jean-marc.petit@insa-lyon.fr

ABSTRACT

This conference report gives an overview of the 10th International French-speaking conference on Knowledge Discovery and Management (KDM) held in Hammamet, Tunisia, January 26-29, 2010. Since 2001, this annual conference has been supported by EGC[1]. EGC is a non-profit French association which aims at promoting KDM both in academia and industries. The conference is the main annual event of the association EGC. Even if the official language is French, the selection process of the conference follows high-quality international standard, e.g. at least 3 reviews per paper, blind reviews, author feedbacks ...
At a glance, EGC 2010[2] may be summarized in some key figures: 139 submitted papers, 32 accepted long papers (25% selection ratio), 30 short papers, 31 posters and 12 software demonstrations; 7 joined workshops and 2 tutorials; 4 keynote talks; a best paper contest rewarded by 2 prizes. More than 275 participants from 23 countries attended this edition celebrating the 10th birthday of the conference.
The next EGC conference will be held in Brest, France, in January 25-28, 2011.

1. INTRODUCTION

In the last decade, the French-speaking scientific community has been very active in the new trends of research around KDD, data mining and knowledge management. Knowledge Discovery and Management (KDM) is concerned with different scientific areas, e.g. Data Mining, Knowledge Discovery, Database, Machine Learning, Statistics, Business Intelligence, Knowledge Engineering, to cite but a few. Since 2001, the French-speaking association EGC aims at promoting the exchanges between researchers and practitioners. The association offers a website[3] and a moderated mailing list registering more than 1000 members, supports an annual International French-speaking conference, and delivers awards to the best research and application papers (2 prizes of 1500 euros each). Previous editions of the conference took place, respectively, in: Nantes (2001), Montpellier (2002), Lyon (2003), Clermont-Ferrand (2004), Paris (2005), Lille (2006), Namur (Belgium) (2007), Sophia-Antipolis (2008) and Strasbourg (2009).

It is worth mentioning that the paper selection process meets high-quality international standard: at least 3 reviews per paper, blind reviews, author feedbacks period ... An important novelty has been introduced this year in the aim of setting up a better biding of papers to reviewers: We experimented the introduction of *moderators* over the four major themes of the conference :

Theme 1 Foundations of KDM (moderated by Gilbert RITSCHARD and Marc BUI),

Theme 2 KDM in emergent domains (moderated by Pascal PONCELET and Georges HÉBRAIL),

Theme 3 Methodological aspects of KDM (moderated by Pierre GANÇARSKI and Stéphane LALLICH),

Theme 4 Experiences and/or systems in KDM (moderated by Fabrice CLÈROT and Jean-Gabriel GANASCIA).

2. SELECTION AND PROCEEDINGS

The selection policy of EGC consists in both selecting high quality papers, and giving a chance to preliminary works. Hence, three formats of acceptance distinguish the contributions: long papers (12 pages) for advanced and mature works, short papers (6 pages) for intermediate results and posters (2 pages) for new ideas still at a preliminary stage of development.

The program committee gathered more than 120 members from all over the world, e.g. USA, Japan, Belgium, Germany, Tunisia, Romania, Switzerland. More than 20% of the committee members came from well established companies doing business in the KDM field, such that Orange Labs, Alcatel-Lucent, EDF R&D.

Papers written in English can be submitted to the conference if one of the co-authors is a non-native French speaker.

We used the free EasyChair[4] on-line submission system to keep track of 158 submitted abstracts (38 for Theme 1, 30 for Theme 2, 55 for Theme 3 and 35 for Theme 4). 139 full papers were finally submitted. We got 453 referees reports:

---

[1] EGC stands for 'Extraction et Gestion des connaissances' which can be translated as 'Knowledge Discovery and Management'

[2] http://www.projets.rnu.tn/egc2010/

[3] http://www.egc.asso.fr/

[4] http://www.easychair.org

79 papers with 3 reviews, 51 with 4 reviews and only 6 with 2 reviews. After a physical meeting of the program committee members in Paris, we got: 32 long papers (25% selection ratio), 35 short papers and 31 posters. In addition, 12 software demonstrations were retained for presentation by the "demos committee" of EGC'2010. The conference proceedings include the whole selected papers completed with the invited talks. We would like to take this opportunity to acknowledge the tireless cooperation and support of the reviewers.

This year, EGC 2010 proceedings where published in the RNTI, E-19, Cèpaduès Editions. The conference is indexed by DBLP. In cooperation with Springer-Verlag, a post-conference proceeding is edited in English from the best papers of the current edition with an additional round of reviews.

3. PROGRAM

The program is composed of a large variety of events: tutorials, workshops, regular talks, invited talks, demonstrations, posters, two best paper awards, and social events.

3.1 Honorary Chair and Invited Talks

Since 2004, an Honorary Chair is attributed to a confirmed and internationally recognized researcher on the topics of the conference. The Honorary Chair of EGC 2010 was Mohamed J. ZAKI from the Rensselaer Polytechnic Institute, New York (USA). The EGC 2010 program included four invited talks:

1. **Francisco de Assis Tenrio de Carvalho** (Universidade Federal de Pernambuco, Recife, **Brasil**): The talk titled RECENT ADVANCES IN PARTITIONING CLUSTERING ALGORITHMS FOR INTERVAL-VALUED DATA began by reviewing partitioning clustering algorithms for interval-valued data having as reference the dynamic clustering algorithm. For each clustering algorithm, it is given the clustering criterion, the best prototype of each cluster, the best distance associated to each cluster (if any) as well as the best partition in a fixed number of clusters. Moreover, various tools for the partition and cluster interpretation of interval-valued data furnished by these algorithms were also discussed. Finally, in order to show the usefulness of these algorithms and the merit of the partition and cluster interpretation tools, experiments with real interval-valued data sets were presented.

2. **Zbigniew W. Ras** (University of North Carolina, Charlotte, North Carolina, **USA**): In the talk titled ACTION RULES AND META-ACTIONS, the speaker introduced Action rules as possible transitions of objects in a decision system S from one state to another with respect to the decision attribute. Classification attributes in S are partitioned into stable and flexible. Meta-actions are defined as actions which trigger changes of flexible attributes in S either directly or indirectly because of correlations among certain attributes in the system. In medical area, taking a drug is an example of a meta-action since some test results (values of attributes in S) for a given patient will get changed. In his presentation, the speaker detailed how meta-actions can be of use to identify interesting action rules and action rules of the lowest cost

3. **Marie-Christine Rousset** (LIG, Grenoble, **France**): The talk titled COMBINING LOGICAL AND NUMERICAL APPROACHES FOR RECONCILIATING DATA AND ONTOLOGIES MAPPING stresses on the fact the reconciliation between data and ontologies mapping is of paramount importance in the sake of information integration. The sighting fact of this talk is to show that combining a logical approach and a numerical one can be of benefit for getting out better mappings as well as better scaling that heavily relies on the use of the logical inference to prune useless numerical computations.

4. **Mohamed J. Zaki** (Rensselaer Polytechnic Institute, Troy, New York, **USA**): The talk titled PATTERN MINING: THE PAST, PRESENT, AND FUTURE offered a thorough tour of the past and the present research landscape in the area of pattern mining (from subsets, to subsequences, subtrees, and subgraphs). The speaker stressed on the fact that as far as the complexity of the pattern types increases, one discovers potentially more informative patterns. Some thoughts on directions for the future of pattern mining concluded the talk.

3.2 Tutorial and workshops

The first day of the conference was dedicated to two tutorials and seven workshops. It met a great success with about 200 participants. The workshops addressed a large variety of topics: Mining Complex Data, Personalized web information retrieval, quality of data and knowledge, Extraction & visualization of knowledge, social web, Assessment of knowledge extraction methods and graphs & alignment of complex objects. The first day program also included two tutorials respectively dedicated to link prediction in social networks, and parallel and distributed Data Mining. The full proceedings of the tutorial and workshop sessions are published by the EGC Association in a paper version. A full electronic version could also be downloaded from the conference and association web sites.

3.3 Regular sessions

The program was composed of 8 sessions for regular papers (short and long), addressing a wide set of topics: Sequential patterns, information retrieval and Web, Clustering, Image mining and augmented reality, Social networks, Pattern mining, Data Streams, Concise representations of patterns and queries, Functional patterns, New mining applications, Annotation, Analysis and visualization of data, Management of Data and Knowledge, Ontologies and applications, supervised classification Techniques, Text mining. The regular sessions were completed with a plenary session devoted to three minutes presentation of each poster and demonstration abstract.

3.4 Best Paper Awards

The EGC Association awards 2 prices of 1500 euros each to the best research paper and to the best application paper. In a first stage, the Program Committee chair nominated 12 papers from accepted long papers. The moderators were also involved in this task. Then, the EGC Steering Committee selected a best research paper[5] and the best application

---

[5] *Extraction of distinctive itemsets from data streams* authored by C. Zhang and F. Masseglia.

paper[6].

3.5   Social events

The organization committee did an excellent job: everything went well during the conference. The conference was held in a five-star hotel resort in Hammamet at a very low price (all-inclusive formula). Moreover, they organized a Tunisian typical social event: Those who attended the gala diner are not ready to forget it !

ACKNOWLEDGMENTS

About the authors

Dr. Sadok Ben Yahia (1971) is an Associate Professor at the Computer Sciences department of the Faculty of Sciences of Tunis since October 2002. He obtained his Ph.D. in Computer Sciences from the Faculty of Sciences of Tunis in September 2001. Currently, He is leading a small group of researchers in Tunis, whose research interests include efficient extraction of informative and compact covers of association rules, visualization of association rules and soft computing. He has served in the program committees of over eight international conferences and has been technical reviewer of different international journals. He has also served is co-guest editor of three international journals special issues. He is currently member of the steering committee of the international conference on Concept Lattices and their Applications (CLA).

Dr. Jean-Marc Petit (1968) is professor at INSA Lyon and LIRIS (UMR 5205 CNRS) since 2005. INSA Lyon belongs to the University of Lyon and is one of the top engineering school in France. From september 1997 to august 2005, he was an associate professor at the University Blaise Pascal, Clermont-Ferrand. Since 2008, he leads the database group (more than 30 members) and he is director at INSA of the master by research program in Computer Sciences since 2007. He belongs to the steering committee of the Association EGC since 1994. His main research interest concerns database and data mining.

---

[6] *Towards Optimal Ad Display on Web Portals in a Realistic Context* authored by V. Gabillon, J. Mary and P. Preux.