

Kullanıcı Merkezli İnteraktif Veri Madenciliği: Bir Literatür Taraması

Hacer KARACAN¹, Mehmet YEŞİLBUDAK²

Gazi Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Maltepe, Ankara¹
Gazi Üniversitesi, Teknik Eğitim Fakültesi, Elektrik Eğitimi Bölümü, Teknikokullar, Ankara²
hkaracan@gazi.edu.tr¹, mehmetyesilbudak@gmail.com²

Özet— Günümüzde birçok veri madenciliği modeli, veri tabanlarında gizli kalmış örüntülerin otomatik olarak elde edilmesi ve veri analiz sürecinin verimliliğine önem vermektedir. Ancak, örüntülerin sadece veri madenciliği algoritmaları ile keşfedilmesi doğru bir yaklaşım değildir. Çünkü keşfedilen örüntüler fazla, ancak kişiler için yararlı olmayabilir. Bununla birlikte bir kişi için faydalı olan özbilgi başka bir kişi için faydalı olmayabilir. Bu nedenlerden ötürü veri madenciliği sürecine kullanıcı merkezli interaktif bir yaklaşımın uygulanması gerekmektedir. Kullanıcı merkezli interaktif veri madenciliği modelleri aracılığıyla kullanıcılar ve bilgisayar sistemleri arasında uyarlanı ve etkin iletişim yapıları oluşturulmakta, kullanıcılara en uygun veri madenciliği algoritmasını tespit etme imkânı sağlanmaktadır. Böylece kullanıcılar için veri madenciliği sıkıntılı ve zorlu bir süreç olmaktan çıkarılmakta, kullanıcılara kendileri için en uygun özbilgileri keşfetme imkânı sağlanmaktadır. Bu çalışmada kullanıcı merkezli interaktif veri madenciliği ile ilgili literatür taraması yer almaktadır. Literatürde yer alan çalışmalar, bu çalışmalardan elde edilen sonuçlar, geliştirilen sistemlerin değerlendirmeleri ve öneriler sunulmuştur. Veri tabanlarında özbilgi keşfi için kullanılan veri madenciliğinin, interaktif bir yapıya ve çoklu görselleştirme tekniklerine sahip olması gerektiği ön plana çıkmıştır.

Anahtar Kelimeler— Veri madenciliği, interaktif sistemler, kullanıcı merkezli tasarım

User-Centered Interactive Data Mining: A Literature Review

Abstract— Nowadays many data mining models concentrate on efficiency of the data analysis process and finding the patterns remained hidden in databases autonomously. However, only the use of data mining models for recognition of the patterns is not an appropriate approach because the patterns obtained can be too many but not useful for human users. Moreover, the knowledge which is useful for one person may not be useful to another person. For these reasons, it is necessary to apply a user-centered interactive approach to data mining process. Adaptive and effective communications are occurred between human users and computer systems and human users are able to determine the most appropriate data mining algorithm through user-centered interactive data mining. Thus, data mining is not a heavy and hard process for human users anymore and human users are able to acquire the most appropriate knowledge which is useful for them. This paper includes a literature review related to user-centered interactive data mining. Studies in the literature, the results obtained from these studies, evaluations of the developed systems and recommendations are presented. It has been taken over that data mining used for knowledge discovery in databases should have an interactive structure and multiple visualization techniques.

Keywords— Data mining, interactive systems, user-centered design

1. GİRİŞ

Teknolojik gelişmelere paralel olarak verilerin dijital ortamlarda saklanması nedeniyle veri tabanlarının hacimlerinde olağanüstü bir artış meydana gelmiştir. Bu durum geleneksel sorgulama ve raporlama araçlarının dev veri yığınları karşısında etkisiz kalmasına yol açmıştır [1].

Bunun sonucunda veri tabanlarında özbilgi keşif sürecinin temelini oluşturan veri madenciliği önem kazanmıştır.

Veri madenciliği modellerinde kullanılan algoritma türüne göre; verilerin içerisinde gömülü olan belirli türde özbilgiler keşfedilmekte, farklı kullanıcı gereksinimleri karşılanmakta ve verilerin farklı yorumlamaları yapılmaktadır [2]. Varolan veri madenciliği modelleri incelendiğinde, genel olarak, veri madenciliği modelinin

edilmektedir. Şekil 1'de bu döngü yer almaktadır [3, 17, 23].

Kullanıcı merkezli interaktif veri madenciliği döngüsünden görüldüğü gibi, temel özbilgi keşif süreci; verilerin seçimi, temizlenmesi (önişlenmesi), dönüştürülmesi, örüntülerin keşfedilmesi (veri madenciliği), değerlendirilmesi ve yorumlanması aşamalarından oluşmaktadır [24, 25]. Kullanıcılar yaptıkları çıkarımlara dayanarak, bu temel sürecin herhangi bir anında sistemi değerlendirebilmekte ve böylece interaktif olarak süreci yönlendirebilmektedir.

Geliştirilen veri madenciliği sistemlerinin kullanıcı merkezli interaktif bir yapıya sahip olması için literatürde pek çok çalışma yapılmıştır. Zhao, Chen ve Yao kullanıcı merkezli interaktif veri madenciliği ile ilgili temel sorunları incelemiştir [2]. Geliştirilecek veri madenciliği sistemlerinin;

- Kullanıcılara çoklu veri madenciliği stratejilerini sunan,
- Çok katmanlı bakış açısına sahip olan,
- Kullanıcı davranışlarını modelleyen,
- Kullanıcı tercihleri için uyarlanabilir bir yapıya sahip olan,
- Elde edilen örüntüleri net bir şekilde açıklayan ve değerlendiren sistemler olması gerekliliği vurgulanmıştır.

Zhao ve Yao interaktif sınıflandırma analizine dayalı bir veri madenciliği uygulaması geliştirmişlerdir [3]. Geliştirilen uygulamada kullanıcı önceden var olan sınıflandırma kurallarını kullanabileceği gibi; yeni sınıflandırma kuralları da belirleyebilmektedir. Kullanıcı veri madenciliği sürecinin herhangi bir anında elde edilen örüntüleri değerlendirebilmekte ve buna göre sınıflandırma işlemini yönlendirebilmektedir. Geliştirilen sistem elde edilen örüntüleri tablo, ağaç görünümü, pasta grafiği ve çubuk grafik olarak sunabilmektedir. Geliştirilen sistem sayesinde, kullanıcı veri madenciliği sürecinde aktif olarak yer almakta ve kendisi için yararlı olan örüntüleri elde edebilmektedir. Guo, Peuquet ve Gahegan interaktif kümeleme [26]; Pigeot, Blauth ve Bry interaktif ilişkilendirme [27]; Pallez, Brisson ve Baccino interaktif evrimsel analiz modellerini uygulamışlardır [28]. Bu uygulamalarda geliştirilen etkileşimli veri madenciliği sistemleri sadece tek bir veri madenciliği tekniğini kullanmaktadır. Oysaki kullanıcılar farklı yeteneklere, zekâyâ, kavramsal seviyeye, hazır bulunuşluk düzeyine ve zihinsel özelliklere sahiptir. Bu nedenle her bir kullanıcı veriyeye farklı bir açıdan bakar ve farklı bir anlam yükler. Dolayısıyla geliştirilecek interaktif veri madenciliği sistemleri pek çok teoriyi ve metodu içermelidir [2].

Chen, Wu ve Zhu sınıflandırma, ilişkilendirme ve kümeleme analiz modellerinin kullanıldığı çevrimiçi interaktif veri madenciliği sistemi geliştirmişlerdir. Uygulanan veri madenciliği sürecinde kullanıcı ile

etkileşim kurulmakta ve sistem kullanıcıya en uygun veri madenciliği tekniğini seçmesi için yardım etmektedir. Böylece veri madenciliği süreci filtre edilmekte ve en etkin örüntüler elde edilebilmektedir. Geliştirilen sistem özellikle veri madenciliği ve yapay zekâyâ ilgi duyan öğrenciler ve üst düzey araştırmacılar için uyarlanmıştır [15].

Chmelar ve Stryka OLAM SE olarak adlandırılan çok seviyeli ilişkilendirme, tanımlama ve sınıflandırma algoritmalarının kullanıldığı interaktif ve sezgisel veri madenciliği platformu geliştirmişlerdir. Bu platform sayesinde kullanıcı veri madenciliği teknikleri hakkında bilgi sahibi olmasa bile ihtiyaç duyduğu örüntülere erişebilmiştir [29].

Han, Fu, Wang, Chiang, Gong, Koperski, Li ve Lu ilişkişel veri tabanlarında çok seviyeli özbilgilerin interaktif veri madenciliği ile elde edilebilmesi için DBMiner adlı bir sistem geliştirmişlerdir. DBMiner veri madenciliği tekniklerinden genelleştirme, tanımlama, ilişkilendirme, sınıflandırma ve tahmin etme fonksiyonlarını birlikte kullanmaktadır. Sistem uygulamada iyi bir performans sergilemiştir [30].

Ankerst veri madenciliği sürecine insanların etkin bir şekilde adapte edilebilmesi amacıyla her bir veri madenciliği tekniği için interaktif bir ara yüz oluşturulmasını ve elde edilen örüntülerin analizinde görselleştirme tekniklerinin kullanılmasını önermiştir [31].

Shouhong ve Hai interaktif veri madenciliği sistemlerinin birbirine bağımlı veri tabanı, veri derleme aracı, veri görselleştirme tekniği ve insan-bilgisayar etkileşimli ara yüz katmanlarından oluştuğunu savunmuşlardır [32].

Beale ve Pryke interaktif kümeleme analizi için genetik algoritmalar kullanmış ve elde edilen örüntüleri üç boyutlu olarak görselleştirmişlerdir. Kullanıcılar görsel sonuçlar yardımıyla elde edilen örüntülere daha geniş bir açıdan bakabilmiş ve kendileri için yararlı olmayan, detaylı örüntüler ile zaman kaybetmemişlerdir. Kullanıcı veri madenciliği sürecini yönlendirebilmiştir [33]. Veri madenciliği ve görselleştirme tekniklerinin bütünleştirilmesi, veri madenciliği alanındaki kullanıcı interaktifliğini arttırmıştır [34, 35].

Ankerst, Elsen, Ester ve Kriegel veri madenciliğinde karar ağaçlarının interaktif olarak oluşturulabilmesi için kompleks algoritmalar yerine çok boyutlu görselleştirme tekniklerinin ve interaktif yaklaşımların kullanıldığı interaktif görsel sınıflandırma sistemi geliştirmişlerdir. Bu sistem sayesinde kullanıcıların oluşturduğu karar ağaçlarının hem boyutu azalmış hem de anlaşılabilirliği ve kararlılığı artmıştır [36].

Hellerstein, Avnur, Chou, Hidber, Olston, Raman, Roth ve Haas veri analizinin online olarak etkin ve verimli bir şekilde yapılabilmesi için senaryoların, interaktif kullanıcı arayüzlerinin, veri kayıt sistemlerinin, veri görselleştirme

tekniklerinin, veri madenciliğinin online olarak kullanıldığı CONTROL adlı bir algoritma geliştirmişlerdir. Bu sayede insan-bilgisayar etkileşimi ve veri analizinin bütünleştirildiği hızlı bir veri madenciliği algoritması geliştirilmiştir [37].

Bhaskaran, Vijayshankar ve Hellerstein veri analiz sürecinin herhangi bir anında kullanıcının dinamik olarak elde edilen örüntüleri inceleyebildiği ve düzenleyebildiği P&S adlı bir algoritma geliştirmişlerdir. Bu algoritma aracılığıyla kullanıcılar veri analiz sürecini interaktif olarak kontrol edebilmiş ve böylece veri analiz süreci kısalmıştır [38].

Hübscher, Puntambekar ve Nye veri madenciliğine dayalı CoMPASS adlı eğitimsel hipermedya sistemi geliştirmişlerdir. Geliştirilen sistem öğrencilere dinamik içerik haritaları oluşturmuş ve öğrencilerin sistem içerisindeki navigasyonları kaydedilmiştir. Sistem sayesinde öğrencilerin davranışları ile öğrencilerin öğrenme stratejileri arasında söz konusu olan örüntüler elde edilmiştir. Geliştirilen sistemin en önemli avantajı; öğrencilerin çalışma alanına veya söz konusu olan problemlerine özgü veri madenciliği tekniğini sunmasıdır. Öğrencilerin yanlış varsayımlarda bulunduğu durumlarda ise, kullanılacak veri madenciliği tekniği de yanlış olacağı için gizli örüntüler elde edilemeyecektir. Bu ise sistemin dezavantajları arasında yer almaktadır [39].

Demsar, Zupan, Leban ve Curk Orange adlı açık kaynak kodlu bir platform geliştirmişlerdir. Bu platformda kullanıcılara, görsel programlama dili aracılığıyla yeni makine öğrenme algoritmaları ve yeni veri madenciliği uygulamaları geliştirebilme olanağı sunulmuştur. Platform özellikle deneyimli kullanıcılara ve araştırmacılara hitap etmektedir [40].

Hou, Gu, Che, Luo ve Jiang istatistiksel veri madenciliği için StatMine adlı interaktif bir sistem geliştirmişlerdir. Geliştirilen sistemde veri tabanlarındaki verilerin incelenmesi ve farklı türde özbilgilerin keşfi için özbilgi arama operatörleri kullanmıştır. Geliştirilen sistem veri tabanı yerine, bu veri tabanına ait özet tablolarından yararlanmıştır. Sistem istatistiksel özbilgilerin keşfinde kesin sonuçlar vermiştir [41].

5. SONUÇLAR

Kullanıcı merkezli interaktif veri madenciliği ile ilgili literatürde yer alan çalışmalar incelendiğinde, veri tabanlarında özbilgi keşfi için kullanılan veri madenciliğinin, interaktif bir yapıya sahip olması gerektiği ön plana çıkmıştır. Çünkü veri madenciliği gizli örüntüleri otomatik olarak elde ederken; interaktif veri madenciliği gizli örüntüleri kullanıcılarla etkileşimi neticesinde elde etmektedir. Dolayısıyla interaktif veri madenciliği sistemlerinden elde edilen örüntüler kullanıcılar için optimumdur. Bu sayede veri tabanlarında özbilgi keşif süreci hem kısalmış hem de etkin, adaptif bir yapıya sahip olmuştur.

Veri tabanlarında özbilgi keşif sürecinin herhangi bir anında, kullanıcıların, yaptıkları çıkarımlara dayanarak sistemi değerlendirmesi ve süreci yönlendirmesinin; elde edilecek örüntülerin kullanılabilirliğini arttıracığı vurgulanmıştır. Bu amaçla interaktif yapıya sahip sınıflandırma, kümeleme, ilişkilendirme, evrimsel analiz modelleri vb. uygulamalar geliştirilmiş fakat yapılan bu çalışmalar sadece tek bir veri madenciliği tekniğini kapsadığı için tüm kullanıcılara hitap edilememiştir. Geliştirilen veri madenciliği sistemlerinin interaktifliğini arttırmak için birkaç veri madenciliği tekniğinin bir arada bulunduğu ve kullanıcıya en uygun tekniği belirlemede yardımcı olan platformlar geliştirilmiştir.

Geliştirilen platformlar birden fazla interaktif veri madenciliği tekniğini içermesine rağmen, kullanıcılar elde edilen örüntülerin analizini tam anlamıyla gerçekleştirememiştir. Kullanıcıların elde edilen örüntülere daha geniş bir açıdan bakabilmesi amacıyla platforma veri görselleştirme teknikleri uygulanmıştır. Böylece kullanıcılar yararlı olmayan, detaylı örüntüler ile zaman kaybetmemişlerdir.

Gelecekteki kullanıcı merkezli interaktif veri madenciliği uygulamalarının tanımlama, ilişkilendirme, sınıflandırma, kümeleme, sıradışılık ve evrimsel analiz modellerinin tümünü içeren, çoklu görselleştirme tekniklerine sahip bir platformda geliştirilmesi gerekmektedir. Geliştirilecek platformun online olması; sistemin erişilebilirliğini, ve dolayısıyla kullanılabilirliğini arttıracaktır.

KAYNAKLAR

- [1] H. Akpınar, "Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği", Istanbul University Journal of the School of Business Administration, 29(1), 1-22, 2000.
- [2] Y. Zhao, Y. H. Chen, Y. Y. Yao, "User-Centered Interactive Data Mining", **5th IEEE International Conference on Cognitive Informatics (IEEE ICCI'06)**, Beijing, China, 457-466, 2006.
- [3] Y. Zhao, Y. Yao, "On Interactive Data Mining", **2nd Indian International Conference on Artificial Intelligence (IICAI'05)**, Pune, India, 2444-2454, 2005.
- [4] İnternet: İnsan Bilgisayar Etkileşimi, http://ceng.gazi.edu.tr/~hkaracan/BM515_H1.pdf, 2009.
- [5] İnternet: Data Mining: Concepts and Techniques, <http://www.ir.iit.edu/~dagr/DataMiningCourse/Spring2001/BookNotes/4lang.pdf>, 2001.
- [6] H. Sever, B. Oğuz, "Veritabanlarında Bilgi Keşfine Formel Bir Yaklaşım, Kısım 1: Eşleştirme Sorguları ve Algoritmalar", Information World, 3(2), 173-204, 2002.
- [7] A. S. Albayrak, Ş. K. Yılmaz, "Veri Madenciliği: Karar Ağacı Algoritmaları ve İMKB Verileri Üzerine Bir Uygulama", Süleyman Demirel University The Journal of Faculty of Economics and Administrative Sciences, 14(1), 1-52, 2009.
- [8] T. Altay, **Knowledge discovery in databases and data mining techniques: An applied study**, M.Sc Thesis, Marmara University, Institute for Graduate Studies in Pure and Applied Sciences, 2005.
- [9] K. Wang, S. Tong, B. Eynard, L. Roucoules, N. Matta, "Review on Application of Data Mining in Product Design and Manufacturing", **4th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD'07)**, Hainan, China, 613-618, 24-27 August, 2007.
- [10] C. Orlar, L. Wehenkel, "Data Mining", IEEE Computer Applications in Power, 12(3), 19-25, 1999.
- [11] R. Jin, G. Yang, G. Agrawal, "Shared Memory Parallelization of Data Mining Algorithms: Techniques, Programming Interface and

- Performance”, *IEEE Transactions on Knowledge and Data Engineering*, 17(1), 71-89, 2005.
- [12] A. Kalikov, **Veri madenciliği ve bir ticaret uygulaması**, Yüksek Lisans Tezi, Gazi Üniversitesi, Fen Bilimleri Enstitüsü, 2006.
- [13] M. Chen, J. Han, P. S. Yu, “Data Mining: An Overview From a Database Perspective”, *IEEE Transactions on Knowledge and Data Engineering*, 8(6), 866-883, 1996.
- [14] P. Booth, **An Introduction to Human-Computer Interaction**, Lawrence Erlbaum Associates, Hove, UK, 1989.
- [15] Q. Chen, X. Wu, X. Zhu, “OIDM: Online Interactive Data Mining”, **17th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems (IEA/AIE 2004)**, Ottawa, Canada, 66-76, 2004.
- [16] A. S. Filho, H. E. Liesenberg, R. M. Barros, “Designing User Interface for Web Interactive Systems”, **3rd IEEE Symposium on Application-Specific Systems and Software Engineering Technology (ASSET’00)**, Richardson, Texas, USA, 9-16, 2000.
- [17] Q. Luo, “Advancing Knowledge Discovery and Data Mining”, **1st International Workshop on Knowledge Discovery and Data Mining (WKDD’08)**, Adelaide, South Australia, 3-5, 2008.
- [18] F. V. Ficarra, M. C. Ficarra, “Interactive Systems, Design and Heuristic Evaluation: The Importance of the Diachronic Vision”, **New Directions in Intelligent Interactive Multimedia**, Springer, Heidelberg, 625-634, 2008.
- [19] Internet: Ten Usability Heuristics, http://www.useit.com/papers/heuristic/heuristic_list.html, 2005.
- [20] Internet: Güvenlik Sistemleri için Kullanıcı Arayüzü Tasarımı, <http://www.bilmuh.gyte.edu.tr/~ispinar/BIL673/Zsevklikull%20ayuz-tas.pdf>, 2004.
- [21] Internet: User Interface Design, http://ceng.gazi.edu.tr/~hkaracan/BM515_H5.pdf, 2009.
- [22] S. Zuffia, C. Brambillab, G. Berettac, P. Scalaa, “Human Computer Interaction: Legibility and Contrast”, **14th International Conference on Image Analysis and Processing (ICIAP’07)**, Modena, Italy, 241-246, 2007.
- [23] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, “From Data Mining to Knowledge Discovery in Databases”, *AI Magazine*, 17(3), 37-54, 1996.
- [24] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, “The KDD Process for Extracting Useful Knowledge from Volumes of Data”, *Communications of the ACM*, 39(11), 27-34, 1996.
- [25] A. Lew, H. Mauch, “Introduction to Data Mining Principles”, **Introduction to Data Mining and Its Applications**, Springer, Heidelberg, 1-20, 2006.
- [26] D. Guo, D. J. Peuquet, M. Gahegan, “ICEAGE: Interactive Clustering and Exploration of Large and High-Dimensional Geodata”, *Geoinformatica*, 7(3), 229-253, 2003.
- [27] I. Pigeot, A. Blauth, F. Bry, “Interactive Analysis of High-Dimensional Association Structures with Graphical Models”, *Metrika*, 51(1), 53-65, 1998.
- [28] D. Pallez, L. Brisson, T. Baccino, “Towards a Human Eye Behavior Model by Applying Data Mining Techniques on Gaze Information from IEC”, **3rd International Conference on Human Centered Processes (HCP’2008)**, Delft, Netherlands, 51-64, 2008.
- [29] P. Chmelar, L. Stryka, “Interactive Mining on Hierarchical Data”, **13th Conference STUDENT EEICT**, Brno, Czech Republic, 410-414, 2007.
- [30] J. Han, “DBMiner: Interactive Mining of Multiple-Level Knowledge in Large Relational Databases”, **International Conference on Management of Data (SIGMOD’96)**, Montreal, Canada, 50-59, 1996.
- [31] M. Ankerst, “Human Involvement and Interactivity of the Next Generation’s Data Mining Tools”, **International Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD’01)**, Santa Barbara, California, USA, 2001.
- [32] Internet: “Interactive Visual Data Mining”, <http://www.scribd.com/doc/15983592/FADLINREQ3BIdeagrouppublishingencyclopediaofDatawarehousingandMiningjul2005eBookLinG>, 2009.
- [33] A. Pryke, R. Beale, “Haiku: Interactive Comprehensible Data Mining”, **Workshop on Ambient Intelligence for Scientific Discovery (SIGCHI’04)**, Vienna, Austria, 2004.
- [34] M. Chen, Q. Zhu, Z. Chen, “An Integrated Interactive Environment for Knowledge Discovery from Heterogeneous Data Resources”, *Information and Software Technology*, 43(8), 487-496, 2001.
- [35] M. C. Oliveira, H. Levkowitz, “From Visual Data Exploration to Visual Data Mining: A Survey”, *IEEE Transactions on Visualization and Computer Graphics*, 9(3), 378-394, 2003.
- [36] M. Ankerst, M. Ester, H. P. Kriegel, “Visual Classification: An Interactive Approach to Decision Tree Construction”, **5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**, San Diego, California, USA, 392-396, 1999.
- [37] J. Hellerstein, R. Avnur, A. Chou, C. Hidber, C. Olston, V. Raman, T. Roth, P. Haas, “Interactive Data Analysis: The Control Project”, *IEEE Computer*, 32(8), 51-59, 1999.
- [38] V. Raman, B. Raman, J. M. Hellerstein, “Online Dynamic Reordering for Interactive Data Processing”, **25th International Conference on Very Large Data Bases**, Edinburgh, UK, 709-720, 1999.
- [39] R. Hübscher, S. Puntambekar, A. H. Nye, “Domain Specific Interactive Data Mining”, **11th International Conference on User Modeling (UM’07), Workshop on Data Mining for User Modeling**, Corfu, Greece, 81-90, 2007.
- [40] J. Demsar, B. Zupan, G. Leban, T. Curk, “Orange: From Experimental Machine Learning to Interactive Data Mining”, **Practice of Knowledge Discovery in Databases**, Springer, Heidelberg, 537-539, 2004.
- [41] W. Hou, Y. Gu, D. Che, C. Luo, Z. Jiang, “StatMine: An Interactive Statistical Data Mining System”, *International Journal of Computational Science*, 2(1), 122-140, 2008.

