Data Mining Applications in Healthcare: Research vs Practice

Olegas NIAKŠU and Olga KURASOVA
Vilnius University, Institute of Mathematics and Informatics,
Akademijos str. 4, LT-08663, Vilnius, Lithuania
E-mail: Olegas.Niaksu@mii.vu.lt , Olga.Kurasova@mii.vu.lt

Abstract. The paper interrogates the commonly accepted belief that data mining is widely used in medicine [8] by comparing academic advances with practical achievements in the field. The paper discusses practical usage and potential gains of data mining in healthcare facilities along with the growing number of publications indicating increasing interest to the topic in the scientific society. In order to evaluate the practical use of data mining in healthcare, a survey of tertiary hospitals in 5 countries has been conducted. The countries from diverse economic development regions were selected to cover 7 tertiary hospitals with unlike economic potential. Quantitative analysis of publications in the area of data mining applications in healthcare was made in the period of the last 8 years.

Keywords. Data mining applications, medical information systems, medical informatics

Introduction

The healthcare domain is known for its ontological complexity and variety of medical data standards and variable data quality [3, 4, 5]. Adding to this privacy consideration, making an effective and practically usable medical knowledge discovery is an open subject for the last decades. Modern clinical practices also undertake transformation not only in diagnosis, and treatment methods, but also in understanding of health and illness concepts [18].

Though data mining (DM) methods and tools have been applied in different domains already for more than 40 years, their applications in healthcare are relatively young. R. D. Wilson et al. [17] have started to classify and collect medical publications where knowledge discovery and DM techniques were applied or researched from 1966 till 2002. According to their study results "…some authors refer to DM as the process of acquiring information, whereas others refer to DM as utilization of statistical techniques within the knowledge discovery process." In fact, this mix of different concepts makes research complicated and less reliable. Therefore we decided to complement typical detailed analyses of scientific and commercial publications with surveying of large healthcare facilities, which conduct scientific and commercial research studies.

Aiming to avoid misinterpretations, the concept of data mining was defined and explained to the survey participants as follows: "Data mining, as part of knowledge discovery process, is a set of data analysis methods using statistical methods and heuristics, which are used for prediction, classification, clustering tasks or finding hidden patterns and correlations in raw data". Following examples of typical DM usage in healthcare were provided: "patient diagnostics, prediction of patient condition, prediction of post-operational complications…".

Gathering the information from hospitals allows us to put academic effort and practical usage side by side and conclude on actual DM usage, and to understand if there is a gap between data analysis experts' community and healthcare practitioners and scientists.

In this paper, we intend to combine the quantified results of publication search, which contained details of DM applications in healthcare with the results of tertiary[1] hospitals' survey on the practical DM usage. The outcome of the combination of these different sources should help us formulate a hypothesis for a further more specific and larger scale survey on the magnitude of actual DM applications in the healthcare.

Starting from the 21st century many countries have chosen e-Health as a priority national program, which in essence proposes to benefit from the standardization, aggregation of patient's clinical information and healthcare services rendered by providing instant access to that information to healthcare professionals as well as patients themselves [6, 12, 19]. According to the report [20] from the National Center for Health Statistics of USA, adoption of Electronic Health Record (EHR) in the USA as the most prominent medical information system is shown in Figure 1. It illustrates a linearly raising amount of non-sparse, but continual data reflecting patients' clinical continuity together with the treatment which took place and medication being used.

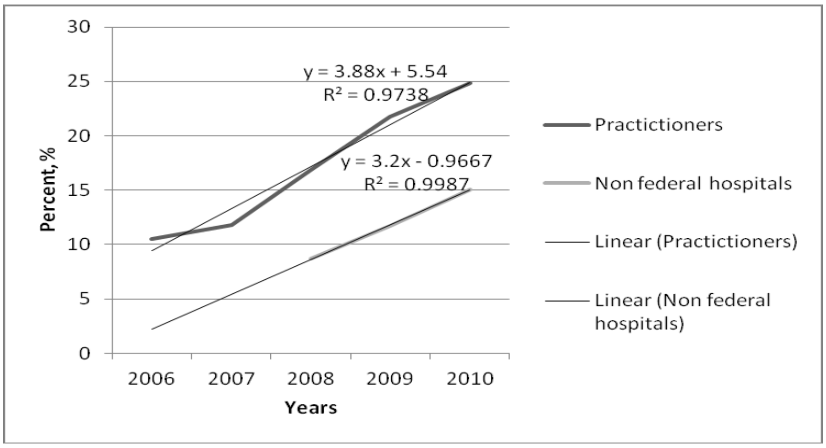

Fig. 1 EHR adoption in the USA

According to strategic plans of the EU member states, the USA and of many other nations from all continents, a considerable amount of investments is allocated to enable the global computerization of healthcare data. Taking a linear progression would mean that in 10 years all new medical encounters will be thoroughly digitalized in all developed countries. The exponential growth is doubtful mostly because of lack of governance structures, data protection and patient privacy issues and resistance from inside of medical community [12]. But even considering a conservative scenario, it is

[1] Tertiary hospital – a major hospital, providing wide range of high level specialized medical services. Commonly tertiary hospitals are university hospitals combining medical and academic activities.

becoming obvious that, for the first time in the history, research community is going to get a full set of a person's medical history from the birthdate till he or she passes away. And that is not for a small specific group limited by longitudinal research study, but the whole regions, nations, countries and even continents. This anticipated scenario forecasts tremendous potential for machine learning and in particular for DM applications in healthcare.

1. Scope of Analysis

Thomson Reuters Web of Science [26], Google Scholar [22] and PubMed [25] databases were used to analyze the number and distribution of scientific publications related to DM in medicine in the last decade.

Tertiary hospitals were selected as a primary source for our survey. The main reason is that typically, tertiary hospitals are in the first line of healthcare institutions that implement clinical software systems, enabling to collect clinical and demographical patient data needed for DM applications. Historically hospital information systems developed starting from the exotic show cases of the economically well-established communities to the standardized practice of handling clinical data and workflows since mid-nineties in developed countries and from the first decade of the XXI century in the developing countries and emerging markets.

2. Scientific Relevance and Development

2.1. PubMed Database

PubMed database is comprised of more than 21 million citations for biomedical literature from MEDLINE, life science journals, and online books. PubMed is operated by National Healthcare Library of U.S. and indexes all publications classifying its content with the help of MESH structured vocabulary [24]. Using MESH vocabulary terms as a search parameter in PubMed database guaranties that not only search wording matching publications will be found, but also its matching synonymic wording or previously used terms. MESH term, classified as MESH heading "data mining" is mapped to other similar concepts like "text mining". "data mining" term was appended to the vocabulary only in 2010 and the former terms e.g. "Information Storage and Retrieval", previously used for the same or similar and related concepts, are mapped to the latest one. A simple search criterion "data mining" was used to retrieve a number of publications and books within the medical domain with assigned MESH heading "data mining". The first publication is dated 1984, however the second one appears only after 10-year interval in 1994. This search resulted in 3077 publications.

2.2. Thomson Reuters Web of Science Database

Web of Science has been providing access to more than 12,000 journals in all subject area. It also includes citations to conference proceedings. The advanced search filter allows the use of logical operations, search restricted to the selected subject areas, and the search scope (title of the publication or whole text). The following constraints have been chosen for our analysis purposes:

(TS=(data mining) AND TS=(medic* OR clinical OR healthcare)) AND Document Types=(Article OR Abstract of Published Item OR Proceedings Paper)

Refined by: [excluding] Web of Science Categories=(OPERATIONS RESEARCH MANAGEMENT SCIENCE OR TELECOMMUNICATIONS)

Timespan=1996-2012. Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH. Lemmatization=On

This search resulted in 2272 publications.

2.3. Google Scholar

Google Scholar provides a scholarly literature search service across many disciplines and sources, including theses, books, abstracts and articles. However it is not limited to scientific publications only. Google Scholar indexes content items published since 1993. Google search filter allows the use of logical operations AND, OR, NOT, a restricted search only in the selected subject areas, search scope (title of the publication or whole text). The following constraints have been chosen for our analysis purpose:

Search in the title: "data mining" AND (medical OR clinical OR medicine OR healthcare)

This search resulted in 478 publications.

The choice of searching in the whole article text was rejected due to a serious flaw: a huge amount of DM centric publications have keywords “medicine” or “healthcare” in the text with a purpose to illustrate DM usage. But this proved to be insufficient to indicate that a publication is focused on DM applications in medicine.

Google indicated the number of publications in the selected period of time approximately. And according to our observations, precision is increasing with a larger quantity of the relevant content items found.

2.4. Results

Distribution of publications found in Web of Knowledge, Google Scholar and PubMed databases starting from 1997 to 2011 is shown in Figure 2.

We can see mostly a linear growth in all the cases, with different line slope: in Google Scholar case it is a very symbolic growth $m \approx 3{,}38$, more significant in Web of Knowledge $m \approx 18{,}28$ and finally in PubMed database $m \approx 30{,}5$. As seen from the explanation of the queries searched in two databases, the results are not directly comparable and are shown here to illustrate a constant raising interest of the academic society in the topic of DM applications in medical domain.

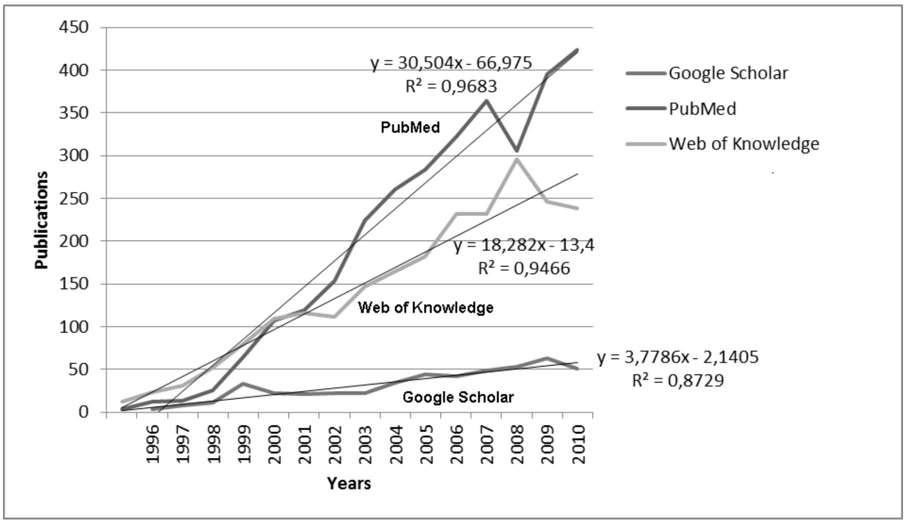

Fig. 2. Trend lines of DM applications in medicine related publications

2.5. Public Interest in the Topic of DM

Though it is difficult to estimate exact numbers, however we can get a fair understanding using a publicly available tool Google Trends [23], which analyzes all search queries executed in Google search engine worldwide. Google Trends service has been collecting and mining data since 2004, providing time-series analysis, reflecting the overall actuality of different search topics or "trends" as Google names it. The dimensions of geographical location, source and language are taken into consideration. We can get a better understanding by comparing the actuality of the term to other disciplines, like *artificial intelligence*, *machine learning*, or a broader concept like *computer science*. This type of analysis provides a unique source of unified data that combines search queries in different languages from different world locations.

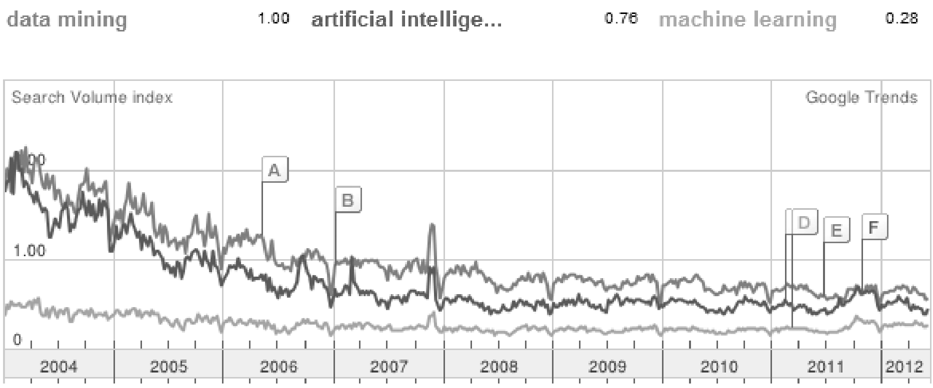

Fig. 3. Google Trends in DM, artificial intelligence and machine learning

Google trends is not providing absolute values on vertical axis, instead, chart's data is scaled to the average search traffic for "data mining" term (represented as 1.0) during the time period from 2004 till 2012.

As shown in Figure 3, there is a general correlation among all the concepts analyzed: *artificial intelligence, machine learning* and *computer science*. Addition to the analysis such disciplines as "mathematics" or "physics" will show the same correlation. A little decrease of the general public interest is noticeable in scientific or, we should say, scholarly topics over years.

Also, looking at Figure 4, which shows the trends in more detail for 2011, we can see a decrease of interest in all topics during summer months, which is most probably a direct indication of summer holidays in the academic society.

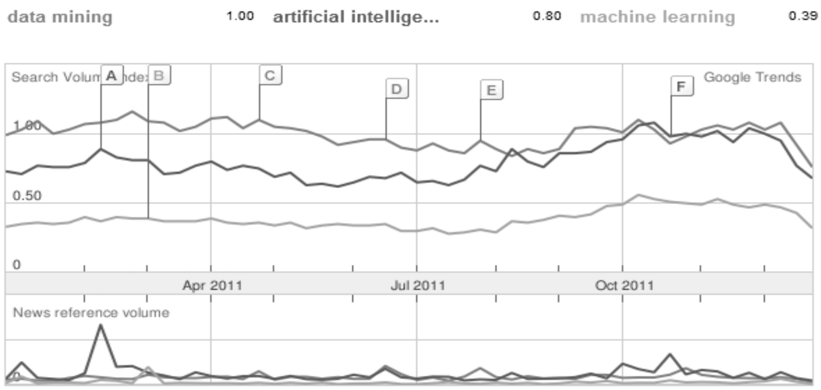

Fig. 4. DM Google trends during 1 year

Figure 3 illustrates that at least during last 8 years, DM topic has a more or less stable interest among the general public, but we assume this interest to be mostly related to scholarly activities. Another interesting outcome of Google Trends analyses was distribution of top 10 geographical locations of the searching query sources (Figure 5).

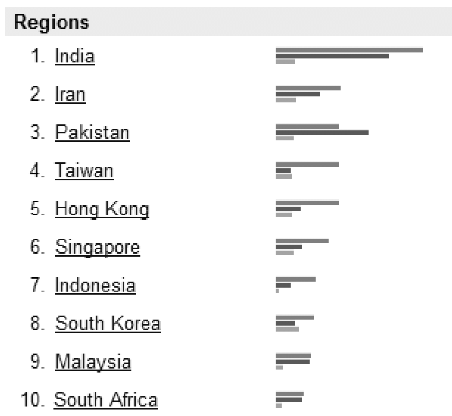

Fig. 5. DM Google Trends by regions

This distribution can be interpreted in different ways. However, it offers an additional perspective in understanding, which regions will be more active in the field in the nearest future.

Summarizing Google Trends results, we can conclude that the peak popularity of “DM” and “artificial intelligence” concepts finished by 2007 and afterwards it remains more or less stable.

Trying to find correlations between DM actuality in the academic world and general public would not be correct because the provided trends reflect different enquiries: scientific result creation versus generic interest in the topic without any obligation or intention to create any sort of result out of it.

3. Surveying DM Applications in Healthcare Facilities

As shown above, the volume of medical related DM research increases from year to year. Naturally, one can suppose that DM usage penetration is increasing accordingly. However, a data analyst working in the field will agree that a large number of research studies remains academic and has no clinical follow up and even rarely goes beyond the institutions which were directly involved in the research. And this already generates reasonable questioning and doubts on the rational and measurable outcome of the research effort. Undoubtedly, a number of proven DM niche applications counts award winning successes, like in radiology imaging or genetics analysis. But this cannot be said about hundreds of specific clinical DM research. Up till our research date, we could not find an example of a systematic approach in an attempt to understand history or the current situation of DM utilization by healthcare institutions. And that can be considered as a blurring factor, preventing the scientific society from concentration on correct ways of the knowledge discovery process tailored for healthcare, which would score the maximum benefit for the clinicians as end users of DM tools and methods and finally patients as beneficiaries and final added value recipients.

Due to the fact that the healthcare sector is very diverse and its entities as well as actors have different objectives and fields of activities, they, employ different methods and tools in their operations [1, 2, 11, 13, 14, 15, 16]. Therefore it was initially agreed to define the scope of this research as DM applications in the healthcare providers’ institutions. However, for this scope the statistically valid representation would require a significant number of different type of institutions like General Practitioners offices, private or public clinics, local and specialized hospitals, regional and, finally, tertiary hospitals. The initial experience of interviewing healthcare institutions suggested that the highest probability of DM usage will score in tertiary hospitals, which have tight relations with the academic society and participate in different sorts of scientific and commercial research on a regular basis. That does not lead to the conclusion that DM applications in smaller institutions do not occur, but focusing on tertiary hospitals allowed us to estimate the upper range of DM penetration into healthcare providers sector. The next important surveying scope constraint is geographical spread of healthcare institutions (HCI). Setting the initial research objective, to understand the practical usage of DM techniques and tools in HCI across the globe puts a very ambitious but unrealistic target to survey thousands of healthcare facilities. It has been decided, that for a limited resources study we should select tertiary hospitals at least from different zones of economic development, having a different magnitude of electronically available patient related data for further analyses. Therefore it was extremely important to select well financed hospitals from the leading economies countries as well as relatively modestly financed hospitals from the developing countries. Hospitals from the following countries participated in the survey: South African Republic, Lithuania, Switzerland, Albania, and Germany. This survey cannot be treated as final as we plan to continue gathering information in upcoming years; however it reveals clear patterns which lead us to concrete conclusions and summarizations that might be useful for both communities of data analysts and clinicians.

3.1. Preparation and Conducting the Survey

Already in early stages it has become obvious, that there is a huge gap in understanding of DM concept by its intended end users - clinicians. Typically hospital's IT department has knowledge and is able to describe how DM is used in the hospital. On the contrary, medical personnel are usually minimally informed or knowledgeable about what exactly DM is and more specifically, how it is used in the hospital. Accordingly, we are in a situation when we cannot ignore either the first class of respondents or the second one. And it was important to get both types of answers for later analysis. Afterwards, we have analyzed the answers classifying both respondents' classes separately and summing them up together. Taking this diverse interviewing audience into consideration, questions were formulated in a comprehensible way for a broader range of respondents with a medical or IT background. See the summarized questions below.

Usage of statistical data analyses, DM and clinical decision support systems:

1. Have you heard about practical applications of DM in medicine?
2. Do you know any research projects in your hospital using DM methods?
3. Have you or your colleagues been involved in DM research project, aiming to identify new patterns or finding new rules for patient diagnostics, prediction of treatment results or other. If yes, please provide a brief summary of research aim and the results.
4. If DM methods have been used, was your experience successful? Please comment
5. Has the clinical decision support IT system been used in your hospital?
6. Please specify which clinical specialties could benefit by using DM methods on collected patient clinical data in your hospital (choose from the list)
7. What type of clinical research your hospital is involved in?
8. Are you or your colleagues potentially interested in the benefits which DM could provide to you?

Availability of Electronic Patient Data for Research:

9. How many years have the patient data been collected in IT systems in your organization?
10. Please specify what clinical patient information is stored in IT systems (HIS, EHR, EMR, RIS, etc.). Select from the list: Observations, Lab results, Radiology reports, Anamnesis, Surgery reports, Discharge summary, Visit summary, Nursing data (vitals), Medication used (for inpatients).
11. Mark medical IT systems used in your organization. Select from the list: EMR / EPR, HIS, RIS/PACS, LIS, Specific clinical information systems, Emergency IS, OP clinic information system, Blood bank information system, Clinical decision support system, Pathology information system.

12. Specify what standard nomenclature is used in your organization (e. g. ICD9, ICD10, SNOMED-CD, LOINC). Select from the list: Patient diagnosis, Pathologic diagnosis, Procedure coding, Laboratory coding.

    Interest in DM:
13. Are you interested in international clinical DM research projects?
14. Specify the clinical specialty or problem you are interested in.

3.2. Method of Survey

The survey was conducted according to methodical guidelines of the Centre for Health Promotion of University of Toronto [21]. A call for survey was openly published in the eHealth news portal eHealthServer.com [27]. The survey was prepared in an online questionnaire and offline forms. The need for an offline version was pointed by some institutions with a limited or no internet access. Hospitals were asked that at least 2 respondents from each institution should fill the questionnaire; a person in charge for medical services, e.g. medical superintendent, director of medicine, head of the clinical department and a person in charge for Information Technology e. g. chief of the IT department. In parallel direct enquiries were sent to the officials of hospitals in 8 countries. Complete interviewing took five months instead of two months planned due to very little or no reaction from the respondents, especially from medical representatives.

The survey's questions allowed crosschecking correctness of the information provided. E. g. question #4 asks explicitly if DM tools are used and question #5 asks if a clinical decisions support system is in use. Typically a clinical decision support system would incorporate a few DM algorithms as well as statistics.

The aim of questions in the section "Usage of statistical data analyses, DM and clinical decision support systems" is to clarify the eligibility of the institution for DM, awareness of DM concept and known applications of DM.

Questions in the section "Availability of Electronic Patient Data for Research" help to figure out the potential of DM in the institution, based on the amount of electronically available data, the medical information system, and standardized medical nomenclature being used.

The aim of questions in the section "Interest in DM" is to define what the interest of the respondents is in possible future DM research projects.

4. Analysis of Survey Data

Out of 14 respondents 12 have confirmed that they had heard about practical applications of DM. However, after the quality validation and answer crosschecking, only 9 positive answers could be qualified. But even out of the remaining 9 respondents with positive answers only 4 are familiar with practical examples of such usage, making up 29% of the whole. Another aspect provided by data validation, is that the majority of medical respondents would have no information about DM research initiatives and applications in their own facilities. It is difficult to specify the overall level of awareness in terms of this survey; however, the selected method of surveying 2 and more representatives from each facility has proved that typically medical specialists, not related to the DM project in their own HCI, have no information about it.

Table 1. Summary of survey answers

Summarized questions	Hospitals of developing countries		Hospitals of emerging countries		Hospitals of western countries	
	IT rep	Clinical rep	IT rep	Clinical rep	IT rep	Clinical rep
Understanding, practical usage and interest						
Good understanding of DM concept	Yes	Differs	No	No	Yes	Differs
Awareness of practical use	No	No	No	No	Differs	No
Hands on DM applications	Differs	No	No	No	Yes	No
Interest in the topic	Yes	Yes	Yes	Yes	Yes	Yes
Clinical specialties	All	All	All	All	All	All
Availability of electronic data for research						
Number of years data is electronically captured	4-13 years		1-3 years		5-15 years	
Variety of medical information systems used to capture and operate with patient related data	patient demographics, radiology images, partly lab results, partly detailed clinical data, billing data		patient demographics, limited radiology images, partly billing data		patient demographics, radiology images, lab results, detailed clinical data, billing data	

Summarized answers, grouped by hospitals with alike economical situation are provided in 1 table. Country groups are represented as follows:

- developing countries – Albania;
- emerging countries – Lithuania and South African Republic;
- western countries - Switzerland and Germany.

Evaluating the benefits of gained DM experience, 50% of respondents, who declared a personal involvement in DM projects, were satisfied with the results achieved and 50% had a neutral opinion on the project success.

Analysis which clinical specialties have the highest potential in DM usage was not successful. Validation of answers has showed that typically all the selected clinical specialties were relevant either to the clinical profile of the respondent or to the clinical profile of the hospital. Summarizing the answers provided, we can conclude, that all clinical specialties without an exception have a potential for DM.

The interest in additional information on potential DM benefits was expressed by 86% of respondents, regardless of their initial experience with DM.

The analysis of electronic data availability for DM purposes showed us a correlation between depicted years of clinical data collection in a facility with the level of the region's economic development (Figure 6). Data collection timeframe values spread from 1 to 15 years, with the mean value 8 years and median value 4 years. In terms of medical IT systems being used, 100% of respondents have defined that hospital information system is in use; electronic medical record systems are used in 60% of facilities and radiology imaging systems in 83%.

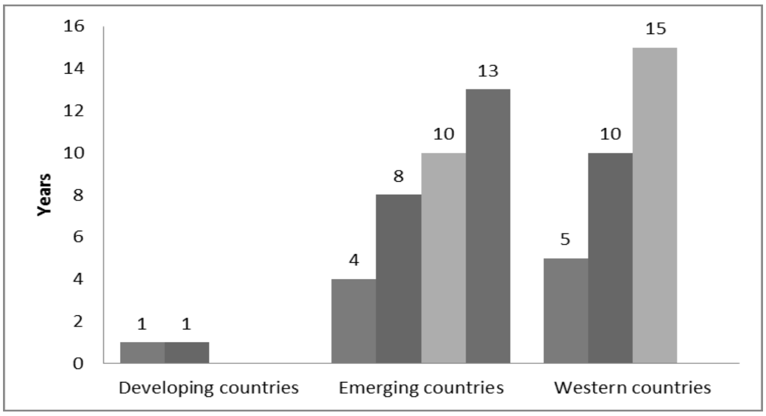

Fig. 6 Clinical patient data collected electronically in hospitals

The usage of standard terminology dictionaries varies depending on the originating country of the facility. The usage of ICD 9 and ICD 10 is very common for coding disease diagnoses. However, other nomenclatures, critical for DM applications and used to code procedure/intervention, laboratory tests, pathology diagnoses, are only partly implemented and at a different quality level.

93% of respondents expressed their will to participate in international clinical DM research projects.

5. Survey Findings

As it was presumed, understanding of DM as a concept as well as its potential depends on the background of the respondents. IT personnel of a hospital typically are well informed on the DM related research and usage inside of the hospital, scoring 100% of its surveyed IT department representatives. In addition, clinicians usually informed only if they were directly involved in such projects.

All the respondents have confirmed that they had heard about practical applications of DM in medicine. However, only 29% of respondents were able to provide any example of practical DM usage.

There is a noticeable confusion in differentiating DM and statistics concepts among healthcare professionals, and very rarely DM is treated by them as a practically valuable tool for clinical purposes.

The respondents from healthcare facilities with a relatively recent adoption of IT in the patient treatment process tend to mix statistical reporting and DM, hospital information systems, Electronic medical record systems and decision support systems.

Regardless of understanding and experience of DM, 86% of respondents expressed their interest in the DM topic and 93% would like to participate in international DM research projects as well as to be informed about utilization of DM techniques in the future.

6. Conclusions

The analysis of publications in the field of DM application in the medical domain has shown a steady growth since its accountable beginning till nowadays. The line slope of publications growth can be averaged to m ≈ 17 on the search conducted in PubMed, Web of Science, and Google Scholar databases. In the early 90'ties up to 5 publications were produced during one year and around 400 publications in 2011. We can conclude that a tremendous growth of interest and scientific advancement took place in the last decade.

On the DM value chain's side, survey revealed, that the greatest part of medical community of tertiary hospitals have either minimal or zero awareness of the DM practical usage and its potential possibilities. All the respondents from the largest university hospitals confirmed to be familiar with DM applications in healthcare, however only 29% of them were able to provide any example of practical DM usage. A huge gap in awareness and understanding of the DM potential was encountered even inside healthcare facilities splitting IT and the clinical personnel to different poles. If we interpolate these results to the smaller, less financed and less exposed to research projects healthcare providers, DM usage will be significantly lower. The survey identified a considerable potential for a further DM penetration due to an increasing amount of patient clinical data collected in HCI and interest declared by hospitals' clinical representatives: 86% of respondents expressed their interest in DM and even more would like to participate in international DM research projects.

However, the process of information digitalization in the developing countries is still in the early phases and the lack of electronically available data is a stopping factor for the spread of DM in a poor economic area.

Summarizing, we have showed that data mining perception and practical applications in healthcare is a way beyond its steady growth in the academic research field, which raises a hypothesis, that relatively a little percentage of academic research effort results in practical DM applications in healthcare, out of which we can conclude that the current interdisciplinary approach is not efficient enough. When considering the potential and benefits of knowledge discovery using DM tools in healthcare, it is clear that more attention should be paid to the domain specific problems of successful DM application in healthcare [4, 5, 10], emphasizing the usage of DM methods with self-explanatory models [7, 9, 16] in contrast to black-box methods.

Further research will be continued aiming to collect additional survey data from the USA, Middle East, Asia, and Australia to increase data representation and get more accurate results.

References

[1] R. Bellazzi and B. Zupan, Predictive data mining in clinical medicine: current issues and guidelines. *International Journal of Medical Informatics* **77** (2008), 81–97.

[2] P. Berka, J. Rauch, and D. A. Zighed, *Data mining and Medical Knowledge Management– Cases and Applications*. Idea Group Inc (IGI), 2009.

[3] O. Bodenreider, Ontologies for mining biomedical data. In: *IEEE International Conference on Bioinformatics and Biomedicine*, Philadelphia, Pennsylvania, 2008.

[4] H. Chen, S. Fuller, C. Friedman, and W. Hersh, editors, *Medical Informatics: Knowledge Management and Data Mining in Biomedicine*, Springer Science, 2005.

[5] K. J. Cios and G. W. Moore, Uniqueness of medical data mining, *Artificial Intelligence in Medicine* **26**(2002), 1–24.

[6] D. Castro, *Explaining International IT Application Leadership: Health IT*. The Information Technology @ Innovation Foundation, 2009.

[7] G. Dzemyda, O. Kurasova, and V. Medvedev. Dimension reduction and data visualization using neural networks emerging. In: I. Maglogiannis, K. Karpouzis, M. Wallace, J. Soldatos, editors, *Artificial Intelligence Applications in Computer Engineering* **160** (2007), IOS Press, 25-49.

[8] H. C. Kob and G. Tan. Data mining applications in healthcare, *Journal of Healthcare Information Management* **19**(2) (2005), 64-72.

[9] G. A. Miller, The magical number seven, plus or minus two: some limits on our capacity for processing information, *The Psychological Review* **63** (1959), 81-97.

[10] P. M. Pardalos, V. L. Boginski, and A. Vazacopoulos, editors, *Data Mining in Biomedicine*, Springer Science, 2007.

[11] D. Ruben and Jr. Canlas, *Data mining in Healthcare: Current Applications and Issues*, Thesis, Carnegie Mellon University, Australia, 2009.

[12] K. A. Stroetmann, J. Artmann, and V. N. Stroetmann. *European Countries on their Journey Towards National eHealth Infrastructures.* Final European Progress Report. European Commission, DG Information Society and Media, ICT for Health Unit, 2011.

[13] W. Stühlinger, O. Hogl, H. Stoyan, and M. Müller. Intelligent data mining for medical quality management. In: *Workshop Notes of the 14th European Conference Artificial Intelligence*, 2000, 55–67.

[14] V. Špečkauskienė and A. Lukoševičius. Methodology of adaptation of data mining methods for medical decision support: case study, *Electronics and Electrical Engineering* **2**(90) (2009), 25–28.

[15] P. Treigys, V. Šaltenis, G. Dzemyda, V. Barzdžiukas, and A. Paunksnis, Automated optic nerve disc parameterization, *Informatica* **19**(3) (2008), 403-420.

[16] S. Wasan, V. Bhatnagar, and H. Kaur, The impact of data mining techniques on medical diagnostics, *Data Science Journal* **5** (2006) 119–126.

[17] A. Wilson, L. Thabane, and A. Holbrook. Application of DM techniques in pharmacovigilance, *British Journal of Clinical Pharmacology* **57**(2), (2003), 127-134.

[18] H. R. Wulff, S. A. Pedersen, and R. Rosenberg, *Philosophy of Medicine an Introduction*, Blackwell Scientific Publications, Oxford, 1990.

[19] *Healthcare Information and Management Systems Society. Electronic Health Records. A Global Perspective*. White paper. HIMSS Enterprise Systems Steering Committee and the Global Enterprise Task Force, 2010.

[20] *Division of Health Care Statistics*. NCHS Health E-Stat Report. National Center for Health Statistics of US 2011.

[21] *Workbook. Conducting Survey Research*. Centre for health Promotion of University of Toronto, 1999.

[22] *Google Scholar* – Web search engine, indexes scholarly literature. Available from: http://scholar.google.com/.

[23] *Google Trends*. Web portal. Available from: http://www.google.com/trends/.

[24] *National Library of Medicine* – MeSH. Available from: http://www.nlm.nih.gov/mesh/meshhome.html.

[25] *PubMed – Database of References and Abstracts on Life Sciences and Biomedical Topics*. Available from: http://www.ncbi.nlm.nih.gov/pubmed/.

[26] *Web of Science* – Academic Citation Index Provided by Thomson Reuters. Available from: http://apps.isiknowledge.com/

[27] *eHealth Server news portal*. Available from: http://www.ehealthserver.com/research-and-development/935-survey-on-application-of-data-mining-to-support-clinical-decisions.