



HAL
open science

A Communication Model that Bridges Knowledge Delivery between Data Miners and Domain Users

Scarlett Kelly

► **To cite this version:**

Scarlett Kelly. A Communication Model that Bridges Knowledge Delivery between Data Miners and Domain Users. HICSS 2018 - 51th Hawaii International Conference on System Sciences, Jan 2018, Hawaii, United States. pp.192-198. hal-01651737

HAL Id: hal-01651737

<https://hal.science/hal-01651737>

Submitted on 1 Dec 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Communication Model that Bridges Knowledge Delivery between Data Miners and Domain Users (Author version)

Scarlett Kelly

Dalhousie University; Institut national de recherche en informatique et en automatique (Inria)
scarlett.kelly@dal.ca

Abstract

Findings generated from data mining sometimes are not interesting to the domain users. The problem is that data miners and the domain users do not speak the same language, so human subjectivity towards the domain users' own knowledge fields affects the understanding of knowledge generated from data mining. This paper proposes a communication model based on the reference services model in the field of library science in order to bridge the communications between data miners and domain users. The creation of a data liaison specialist role in the data mining team aims at understanding the subjectivity as well as the thinking process of both parties in order to translate knowledge between the two fields and deliver findings to domain users. Through five steps—data interview, pre-mid evaluation, post-mid evaluation, knowledge delivery, and follow up—the data liaison specialist can achieve effective knowledge synthesis and delivery to the domain users.

1. Introduction

In the time of big data, information and data are ubiquitous and the amount and complexity present an increasing and cross-boundaries trend [1]. Data mining aims at making sense of big data through generating interesting findings or generating new knowledge from datasets [1] [2]. However, there are significant gaps between knowledge synthesis from datasets and knowledge delivery to the domain users. Current attempts, including data visualization [2], domain user engagement [2] [9], and the refinement of technology, all have limitations that inhibit the effective improvement in understanding the knowledge generate from data mining for the domain users.

The fundamental problem is human subjectivity due to different knowledge basis [14] [15] [16]. The source of the problem is that there are two different fields of knowledge between data miners and domain users. In

other words, there are communication problems when two different fields speak different languages. Based on identifying this problem, this paper takes a unique social science perspective by focusing on the human factors in order to understand human subjectivity and prevent the subjectivity from reducing interesting findings, while not eliminating subjectivity. As this problem involves surfacing information, and translating information needs, the reference interview process is considered. Using the existing and successful reference services model in the library science field, the paper proposes a communication model between data miners and domain users with the creation of a data liaison specialist role in data mining teams. Instead of focusing on communicating knowledge after the findings are generated, which is the time that the problem of miscommunication appears, the communication model is designed throughout the data mining process in order to detect subjectivity, generating more interesting findings based on subjectivities, and explain uninteresting findings in plain language that can potentially increase their interestingness.

2. Background: data mining process and domain users

Data mining is a process that discovers knowledge from large amounts of data [3]. Only interesting patterns that are discovered from the datasets represent knowledge [3]. Interesting knowledge includes patterns that are easy to understand by the domain users, confirming a hypothesis for the domain users, valid with some degree of certainty, potentially useful, and novel [3]. Since it is unrealistic and inefficient for data mining to generate all possible patterns, data miners desire to generate only interesting patterns [3]. However, such interestingness is highly vulnerable to subjectivity. For example, measuring subjective interestingness can be based on domain users' beliefs in looking for unexpected, expected, or actionable data [3]. Yet there is no standard of what interestingness is to different domain users. A quantitative study of 13 participants

found that participants were most interested to see unexpected results [4]. However, when comparing correlations between individual users and the wider populations of users, the measures based on comparisons of correlations is no longer effective for identifying interesting information [4]. The uncertainty of interestingness can result in the fact that sometimes an interesting finding in the eyes of data miners are not interesting in the eyes of domain users.

Disagreements between data miners and domain users on the definition of interestingness can result in serious consequences. Such disagreement-resulted mutual influence may not have immediate short-term effect, but long-term changes in both data miners' and domain users' behaviours are inevitable. On the one hand, facing uninteresting results, the domain users can feel dissatisfaction and doubt about data mining technology in general. On the other hand, data miners may tend to find patterns based on the domain users' definition of interestingness in the future. Interesting findings may be discarded only because they do not meet the domain users' expectations. Therefore, in the long run, subjectivity induced disagreement in interestingness will affect the data mining outcomes and its growth as an industry.

3. The nature of the problem and why it exists?

Humans are subjective in nature [17]. Such subjectivity does not only reflect in data mining but also in various fields. For example, decision-making in governments has typically followed a top-down hierarchical process and has been a highly subjective activity that is solely based on the decision-makers' knowledge [6]. E-government provides a platform for citizens to contribute ideas and opinions, so the decision-making process becomes more horizontal [6]. However, e-government is not a solution because governments are still the ultimate decision-makers and they have the option to engage citizens' input or not, even though citizens' opinions are in the Cloud [6]. It is safe to say that so far there is not a single model that can eliminate human subjectivity in the realm of social science. Data mining is a different field: it is a science that relies on scientific methods to extract knowledge from datasets. However, it faces the same subjectivity issue because mathematical tool and feature selections are done by humans [2]. In other words, the data mining tools are not context-aware, so data mining depends on the humans to find interesting patterns by asking the right questions and using the right tools [2]. In this sense, even though the methods are objective and scientific, the choice of the methods can be subjective.

The source of subjectivity in the field of data mining comes from the fact that two different fields do not

speaking the same languages. Data mining uses mathematical methods to generate interesting findings, but the methods are not the best at explaining the findings [2]. For example, Neural Network is a method that is great at finding patterns, but it is not great at explaining how the findings are generated [2]. For domain users, receiving a list of findings without the necessary explanations makes it difficult for them to visualize or interpret these interesting findings, just like a photo without metadata is not going to provide explanations on where and when the photo is taken or who is in the picture [18]. Moreover, sometimes data miners are not mathematicians or computer scientists [2]. Therefore, even though they are familiar with the mathematical tools that they use, they may not be proficient enough in explaining the rationale behind using these tools or the outcomes generated from these tools [2]. In this way, data mining speaks the language that is technical and lacking in explanations, which must be translated from mathematical methods to natural language that is easy to understand by domain users.

For domain users, they are from a variety of fields. The possibility of them knowing data mining language is very small. Therefore, if the findings are not translated to the languages in their fields, it is difficult for them to understand and interpret the findings from the data mining language. In this sense, direct communications between data miners and domain users are unachievable but necessary.

4. The limitations of the current attempts to find solutions

There are a few possible solutions that the current data mining practice is exploring in order to facilitate better communication with the domain users. To date, there have been three possible solutions using descriptive approaches or modifying the mining process, including 1) data visualization [2], 2) domain user engagement [2] [9], and 3) the refinement of technology.

Visualization techniques, such as plotting, are the conventional ways of describing the findings in order to make sense of the results for the domain users [2]. The advantage is that visualized data can function as a universal language between data miners and domain users. However, there are two considerations. First, data miners are not necessarily data visualization experts. Introducing data visualization experts into the data mining process is facing the same problem, which is that data mining and visualization are two fields and they do not speak the same language. Therefore, it runs the risk of not solving the problem of the communication between different fields but adding more complexity. Second, datasets are getting much bigger due to the low costs of preserving data in the cloud environment [2]. Therefore, there are increasing dimensions of large

datasets [2]. Facing hundreds or thousands of dimensions in a dataset, finding the right samples to explore has become a problem [2]. Visualizing data with multiple dimensions will face the possibilities that the domain users cannot understand the complex and multi-dimensional data visualization.

Inserting domain users to the data mining process in order to receive ongoing feedback is another possible solution [2] [19]. The benefit of achieving such interactive data mining process is that the domain users can be a part of the procedure and help identify uninteresting results at an early stage. However, the risk is that new and unexpected knowledge may be discarded at an early stage only because users do not find them interesting [7]. For example, when unexpected results appear in the process due to noisy data and the existence of outliers, the domain users may request more experiments or updates [7]. However, cleaning the data too much in the process runs the risk of eliminating real interesting results at the end. Moreover, if the data miners know about the domain users' expectations of data, the data miners may be influenced and try to meet the expectations of the domain users. If the domain users know about the data miners' work during the process, the domain users may offer too many suggestions and influence the data miners. Uninteresting findings may be discarded during the process before they become interesting findings. Even though the domain users should not intentionally avoid any unexpected result, human subjectivity based on prior knowledge can play a significant part and guide data mining away from the unexpected results [7]. Therefore, inserting domain users into the data mining process could increase the risks of subjectivity and decrease the quality of the overall outcome of data mining.

Data mining is a technology-oriented subject. Therefore, some studies focus on improving technologies in hope of solving human problems. For example, SIREN is an interactive tool that removes redundant results—redescriptions—that do not convey significant new information and require filtering [8]. In this way, SIREN induced a redescription mining that improves the descriptive approach of interesting results of data mining [8]. However, such a method only focuses on the descriptive approach that delivers the interesting results, not the predictive power that generates interesting results [8]. Moreover, technologies are designed by computer scientists or data miners. For example, the creation of artificial users aims to examine the discovered patterns in the data mining process in order to test the interestingness of findings [20]. However, data miner's subjectivity can reflect in the design of the user. Moreover, such interactive process in data mining may increase subjectivity, as the user's background distribution changes and becomes

conditioned on the presence of the newly revealed pattern to the user [5]. Therefore, using technology does not necessarily decrease the human subjectivity towards the datasets: the human subjectivity issue remains.

5. Considering the creation of a data liaison specialist role

When two fields do not speak the same language, the subjectivity towards the knowledge in their own fields increases [15] [16]. A communication model must be built in order to enable communications between the two languages and bridge the understanding of the knowledge generated from data mining. Just like when one person only speaks English and the other only speaks French, a translator must be placed between the two people. However, data mining is already a complex field that requires interdisciplinary knowledge of data science, programming language, algorithms, and statistics. The domain of data sources can also demand high level of knowledge, especially in the field of medicine and biology. Therefore, expecting a translator to speak both languages in data mining and another domain and translate them is unattainable. This requires a new way of thinking of the problem and the creation of an unconventional model to solve the problem.

To solve the problems between communications between data miners and domain users and understand human subjectivity from both fields, humans' involvement is inevitable. Since engaging domain users are not achievable as explained earlier, engaging other human actors can help advise on data interestingness and usefulness to data miners and provide explanations to domain users. For example, Creedo provides a system that supports real users to participate and perform certain data analysis tasks [9]. The advantage is that such arrangements are scalable and repeatable [9]. In this way, Creedo involves humans, who have no previous knowledge of the subjects of data mining and the datasets, as both test participants and evaluators [9]. Even though Creedo significantly increases administrative burden to data mining in terms of study design, multi-users communications and task distributions, and user workload control, the idea behind—engaging the human components into data mining—is highly valuable.

Based on this idea, how to effectively engage the human component into data mining without significant administrative burdens become a question. Social science in the field of library science provides a reference services model. The reference service provides a link between the vast amount of knowledge (data mining) and the knowledge seeker (domain users). A reference librarian does not need to have any previous knowledge in the field that s/he provides references services. For example, a legal librarian does not need to

hold a law degree. However, the reference librarian must have the ability to capture what the knowledge seeker is saying, ask the right questions to get at what the person is not saying, and understand and interpret the person's needs and present the knowledge in a way that meets these needs [24]. In the field of data mining, the reference librarian's role can be transferred to a data liaison role with certain modifications.

The data liaison role can play an important part of data mining as the knowledge synthesis and delivery specialist. As a part of the data mining team (not the domain user's company) serving as the middle-person between data miners and domain users, the data liaison specialist needs to initiate conversations with the two parties from the beginning of the data mining process in order to achieve a holistic understanding of what both parties are looking for from the raw data, the findings/expectations in the process, and the meanings of the interesting findings at the end. In other words, understanding the data mining process, why the findings are interesting from the data miners' perspective, and the possible subjectivities that make the domain users find the findings uninteresting are keys to objectively understand the potential subjectivities and achieve the communications when knowing what the two parties think. A detailed role design / operation is as below.

6. The operation of the Data Liaison Specialist role in data mining

The operation of the data liaison specialist role follows the principles of reference services in a library setting in professional behaviours and people-oriented interactions. In the library science field, reference services must follow the "guideline for behavioral performance of reference and information services providers" criteria — "visibility/approachability," "interest," "listening/inquiring," "searching," and "follow-up"—as well as relevant theories on information seeking and retrieving, including "uncertainty principle," "hierarchical relationship of information," "relevance," and "information representation" to measure the strengths and weaknesses of the reference interaction [10] [11]. For a data liaison specialist, this means that the person needs to be approachable, showing interest to the fields of knowledge around data mining and the datasets being mined, asking questions and listening, interpreting the dialogues and finding potential subjectivities, and continuing the dialogues during the data mining process. The data liaison specialist should be uncertain about what two parties have in mind in order to eliminate the specialist's own subjectivity, understanding human

behaviour and the cause-and-effect relations between behaviour and information seeking and using, asking relevant and wide-ranged questions, and interpret received information in order to detect subjectivities. All these require the data liaison specialist to have strong communication, interpersonal, qualitative research, and knowledge translating skills.

Just like reference services, which are non-linear service delivery but requires the exchanges of dialogues and ideas in order to understand information needs, the communications between the data liaison specialist and the two parties—data miners and domain users—need to happen in a non-linear fashion at different stages of the data mining process. However, the timing and the procedure of the communication need to consider the different mining processes of different datasets as well as the availabilities of data miners and domain users. In general, the communication model should include five stages: data interview, pre-mid evaluation, post-mid evaluation, knowledge delivery, and follow up, as demonstrated in the graph on the next page.

In the first data interview stage, the data liaison specialist functions as an interviewer, listening to the thoughts of both parties, and learning about their uncertainties and certainties about the datasets [10] [12]. These dialogues are crucial because they enable the data liaison specialist to understand the data miner's plan with the datasets as well as the domain user's initial subjectivity in terms of what his/her expectation is towards the findings and definitions of interesting findings. Such knowledge on what both parties think will help the data liaison specialist understand what subjectivity is around the certain datasets before the data mining process begins.

In the second pre-mid evaluation stage, the data liaison specialist can take the opportunity to monitor the data mining progress and learn about the initial findings. At this stage, the data miner has developed a sense of the data quality and what initial findings can be generated. Interestingness from the data miner's perspective can be compared with the interestingness from the domain user's perspective from the data interview stage. It is very unlikely that the recognitions of interestingness perfectly match. At this point, it becomes important that the data liaison specialist to communicate necessary information to both parties without influencing their subjectivities. For example, if the data miner has conducted several outlier removals in order to achieve better results in clustering, but the domain user expects to see some abnormal detection, it is important for the data liaison specialist to ask the data miner to perform tasks on outliers, though detailed

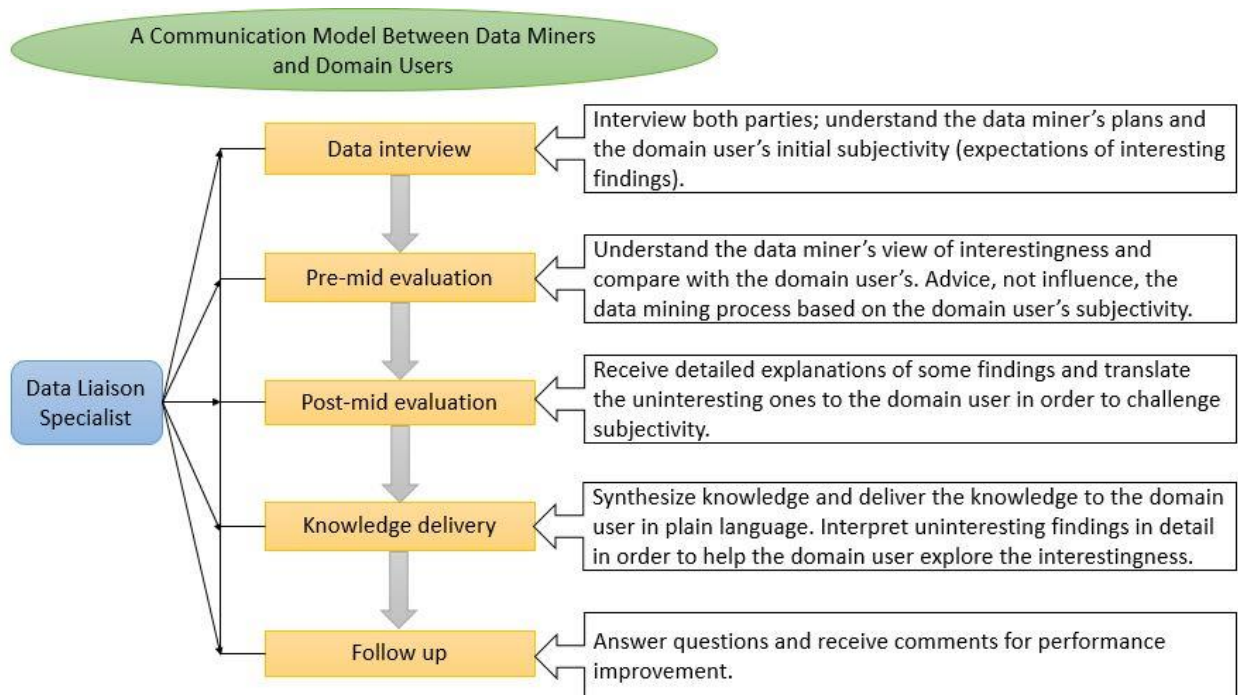


Figure 1. A communication model between data miners and domain users

mechanisms need to be determined by the data miner. This does not mean that the data liaison specialist should advise the data miner to change the mining directions. Rather, the data liaison specialist should advise the data mining process in a way that prevent the domain user's subjectivity from influencing the data miner. Only then data mining can be performed with the considerations of both the data miner and the domain user's subjectivities, and generate new, interesting, and unexpected findings.

In the third post-mid evaluation stage, the data liaison specialist can learn more about the data mining progress and more findings. Since the mining of the datasets is reaching completion, the data liaison specialist can receive detailed explanations from the data miner and deliver some findings that is challenging the domain user's subjectivity. This is a stage that prepares the domain user to learn about the final findings that may not be what the user expects.

In the fourth knowledge delivery stage, the data liaison specialist functions as a knowledge filter, who synthesizes knowledge from the data miner and delivers the synthesized knowledge to the domain user. This requires the data liaison to have the ability to explain abstract findings with plain natural language to the domain user. In this way, some potential uninteresting results can become interesting if the explanation is in detail and easy-to-understand.

In the fifth follow-up stage, the data liaison will continue to function as the bridge between the data miner and the domain user. Any further questions or

concerns from the domain user should come through the data liaison specialist. In this way, the specialist can translate the knowledge from the data miner and deliver to the domain user in plain language. Through the follow-up, any comments and feedback on the data liaison specialist's work performance will contribute to the development of the new role.

7. Case Study

As early as 2003, M. Hofmann and B. Tierney recognized the importance of involving human resources in large scale data mining projects [21]. Their paper introduced a few key human positions, such as business analyst, data analyst and knowledge engineer [21]. The paper also pointed out some key competencies that these positions should have, including leadership, customer relations, as well as risk and change management [21]. However, since the paper was published, such involvement of human resources in data mining team has been informal. This is the reason that the repeated search with changing search terms only led to the conclusion that the real-life case studies virtually do not exist in literature, not to mention any statistical evidence on the cost of a knowledge synthesis role that is similar to the data liaison specialist role. This is consistent with what presented in section 4 of this paper, which stated that the current solutions to the human subjectivity problem lacks the application of the human components. Two qualitative studies, though not directly relate to the operations of a data liaison

specialist role in the data mining team, focused on the communications in the data mining process.

The first case identified the roles and skillsets of the business analyst and the systems analyst roles [22]. After interviewing eight semi-structured interviews in a domain user company, the research concluded that the business analyst required to have the role components of, in descending order, requirements elicitation, mediation, solution designer, and technical specialist, while system analyst required the same but in reversed order [22]. The findings on business analyst can contribute to the development of the data liaison specialist role. The second case followed meetings between domain experts and data miner experts, gained a more in-depth understanding of the collaborative process in data mining, and proposed a new model for the meetings [23]. Even though it did not involve the introduction of a data liaison specialist role, the case showed that improving the communications between data miners and domain users is necessary and achievable.

8. The evaluation criteria, benefits, and limitations of the communication model

The minimized impact of one party's subjectivity on the findings of data mining is key to indicate the usefulness of the communication model and the data liaison specialist role. There are three components to evaluate the success of the communication and the data liaison specialist. First, the findings of data mining are not reduced due to human subjectivity but expanded because the findings are meeting both the expectations of interestingness of the data miner and the domain user. Second, the data miner has the liberty to explore the potentials of the datasets and detect the maximized numbers of meaningful new findings. Third, the domain users accept the unexpected new findings instead of rejecting them only because they are interesting to the data miners but not to the domain users. All these indicators of success reflect the benefits of the communication model.

The biggest benefit of the communication model and the creation of the data liaison role is that the model does not limit or change findings but expand findings based on both the data miner and the domain user's subjectivity by explaining in a way that the domain user can understand and potentially appreciate uninteresting findings. In other words, uninteresting but potentially valuable findings will not be discarded due to the miscommunications between the two fields. It is also a new way of thinking: by adding the external human components into data mining, the subjectivity of data miners and domain users are not eliminated but understood so that the negative effects of subjectivity on data mining can be minimized. Other benefits include

the continuity of the data liaison role in the data mining team without the restraints of disciplines and the relatively low administrative burden in human resources and costs (only one additional employee and salary).

The most obvious limitation of the communication model is that it is still at the theoretical level. Without the validation of other literature and the application to real cases, the usefulness of the model receives no support from real-life evidence. Another limitation is that the model focuses on the subjectivity of the domain user, but the data miner's subjectivity is not addressed enough. Data miner's subjectivity can root at the very beginning of the data mining process, such as sampling (the data selection stage that determines what data is relevant to the analysis tasks) [3]. The choices of sampling can directly affect later clustering and pattern mining, and eventually findings. The data liaison specialist can advise the data miner to perform more sampling in the second pre-mid evaluation stage based on the domain user's subjectivity, but it runs the risks of lengthening the mining project and increasing costs.

9. Conclusion and the future development

Data mining requires the collaboration of different expertise. Human subjectivity exists due to the different fields of knowledge [16]. Such subjectivity is impossible to eliminate in the data mining that requires the use of different knowledge. Therefore, instead of attempting to eliminate subjectivities, the communication model aims to expand the collaboration and communications between data miners and domain users. The creation of a data liaison specialist role can bridge communications throughout the data mining process. Most importantly, by understanding the subjectivities of both data miners and domain users, the data liaison specialist can understand the thinking process of both parties and synthesize and deliver findings in plain language that can potentially increase the levels of interestingness. In this way, the communication model prevents from removing important information only because they do not seem interesting in the eyes of the domain users [13].

For future development, it is crucial to apply the communication model to real cases so that its benefits and limitations can be further examined. Facing the lack of real life case studies, a case study that applies the data liaison specialist role into the real-life data mining process is in planning and will be carried through once funding is in place. Detailed cost-benefit analysis with statistic evidence can be developed from this future case study. More detailed mechanism of the model can be developed based on the different datasets. However, even though the proposed data liaison specialist role is theoretical, the formalization of the idea can make the data mining teams that have already informally applied

such practice examine their practice and potentially conduct case studies on their practices. In this sense, the impact of the idea of the data liaison specialist can be significant in the data mining practice and knowledge sharing. The collaboration of domain users is also important in order to minimize the effects of subjectivity from a single individual [7]. With a more mature communication model, further collaboration in the data mining process will become possible.

Acknowledgement

I would like to thank Mitacs Globalink Research Award for providing me funding to conduct this research at Inria LACODAM team, as well as Dr. Alexander Termier, Dr. Sandra Toze, Dr. Ryan Whalen, Ms. Lindsay McNiff, Mr. Clement Gautrais, and Ms. Marie-Noëlle Georgeault for their advice, generous help, and ongoing support.

References

- [1]. S. Russell and I.S. Moskowitz, "Human information interaction, artificial intelligence, and errors", *Association for the Advancement of Artificial Intelligence*, 2016, pp. 33-39.
- [2]. S.A Amraii, M. Lewis, R. Sargent, and I. Nourbakhsh, "Explorable Visual Analytics Knowledge Discovery in Large and High-Dimensional Data", *KDD 2014 Workshop on Interactive Data Exploration and Analytics (IDEA)*, New York, 2014, pp. 26-34.
- [3]. Han, J., M. Kamber, and J. Pei, *Data mining: Concepts and Techniques*, Morgan Kaufmann, Boston: Elsevier, 2012.
- [4]. S.L. Jones and R. Kelly, "Finding 'Interesting' Correlations in Multi-Faceted Personal Informatics Systems", *SIGCHI Extended Abstracts on Human Factors in Computing Systems 2016*, California, 2016, pp. 3099-3106.
- [5]. T. De Bie. "Formalising the subjective interestingness of a linear projection of a data set: two examples", *KDD 2014 Workshop on Interactive Data Exploration and Analytics (IDEA)*, New York, United States, 2014, pp. 47-51.
- [6]. Kelly, S., *Digital information revolution changes in Canada: E-government design, the battle against illicit drugs, and health care reform*, Lammi Publishing Inc., Alberta: Coaldale, 2016.
- [7]. P. Miettinen. "Interactive Data Mining Considered Harmful*(If Done Wrong)", *KDD 2014 Workshop on Interactive Data Exploration and Analytics (IDEA)*, New York, 2014, pp. 85-87.
- [8]. E. Galbrun, and P. Miettinen. "SIREN: An interactive tool for mining and visualizing geospatial redescription". *Proceeding of the 18th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'12)*, 2012, Beijing, China, pp. 1544-1547.
- [9]. M. Boley, M. Krause-Traudes, B. Kang, and B. Jacobs. "Creedo—Scalable and Repeatable Extrinsic Evaluation for Pattern Discovery Systems by Online User Studies", *KDD 2015 Workshop on Interactive Data Exploration and Analytics (IDEA'15)*, Sydney, Australia, 2015.
- [10]. B. J. Jansen, and S.Y. Rieh, "The seventeen theoretical constructs of information searching and information retrieval", *Journal of the American Society for Information Science and Technology*, pp. 1517-1534.
- [11]. RSS Management of Reference Committee, "Guidelines for behavioral performance of reference and information service providers. Reference and User Services Association (RUSA)", *American Library Association*, Chicago, 2011.
- [12]. Bopp, R.E., and L.C. Smith, *Reference and information services: An introduction*, CA: Libraries Unlimited. Santa Barbara, 2011.
- [13]. P. Fule and J.F. Roddick, "Experiences in Building a Tool for Navigating Association Rule Result Sets", *Australian Computer Society, Inc. Dunedin, New Zealand*, 2004, pp. 103-108.
- [14]. T. De Bie. "Maximum entropy models and subjective interestingness: an application to tiles in binary Databases", *Data Mining Knowledge Discovery (DMKD)*, Springer, 2011, pp. 407-446.
- [15]. T. De Bie. "Subjective interestingness in exploratory data mining", In *Advances in Intelligent Data Analysis XII - 12th International Symposium, IDA 2013*, London, UK, 2013, pp. 19-31.
- [16]. B. Crémilleux, M. Plantevit, and A. Soulet. "Preference-based Pattern Mining", *14th International Conference on Formal Concept Analysis*, Rennes, France, 2017, pp. 1-171.
- [17]. Hockenbury, S.E., S.A. Nolan, and D.H. Hockenbury, *Psychology*, Worth Publishers, United States: New York, 2015.
- [18]. A.J. Gilliland, "Setting the Stage", in M. Baca (Ed.), *Introduction to metadata*, Getty Publications, Los Angeles, United States, 2008.
- [19]. T. De Bie, "An Information Theoretic Framework for Data Mining", *KDD 2011 Workshop on Interactive Data Exploration and Analytics (IDEA)*, San Diego, United States, 2011.
- [20]. V. Dzyuba, "Mine, Interact, Learn, Repeat: Interactive pattern-based data exploration", unpublished thesis, 2017.
- [21]. M. Hofmann, and B. Tierney. "The involvement of human resources in large scale data mining projects", *1st international symposium on Information and communication technologies*, Trinity College Dublin, 2003, pp. 103-109.
- [22]. A. Vonsavanh, and B.R. Campbell. "The Roles and Skill Sets of Systems vs Business Analysts". *19th Australasian Conference on Information Systems*, Christchurch, 2008, pp. 1059-1068.
- [23]. J. Ma, and C.G. Drury. "Analysis of Collaborative Meetings in Developing Data Mining Models". *the Human Factors and Ergonomics Society Annual Meeting*, Los Angeles, CA, 2005, pp. 686-689.
- [24]. Krier, L. and Strasser, C.A. *Data management for libraries: a LITA guide*, American Library Association, U.S: Chicago, 2014