Proceedings of the 2014 Federated Conference on
Computer Science and Information Systems pp. 1391–1396
DOI: 10.15439/2014F289
ACSIS, Vol. 2

Intelligent Association Rules for Innovative SME Collaboration

Gulgun Kayakutlu *
Industrial Engineering Dept.
Istanbul Technical University
Macka 34367
Istanbul, Turkey
Email:{kayakutlu@itu.edu.tr}

Irem Duzdar
Industrial Engineering Dept.
Istanbul Arel University
Türkoba Mahallesi Erguvan Sokak
34537, Tepekent İstanbul-Türkiye
Email:{iremduzdar@arel.edu.tr}

Eunika Mercier-Laurent
IAEUniversity Lyon 3
6, cours Albert Thomas
Lyon,France
Email:{eunika@innovation3d.fr}

***Abstract*—SMEs are encouraged to collaborate for research and innovation in order to survive in tough global competition. Even the technology SMEs with high knowledge capital have the fear to collaborate with other SMEs or bigger companies. This study aims to illuminate the preferences in customer, supplier and competitor collaboration within industry or inter industry. A survey is run on more than 110 companies and Machine Learning methods are used to define the association rules that will lead for success.**

***Index Terms*—Collaborative Innovation, Association Rules, SVM, SOM**

I. Introduction

KNOWLEDGE based SMEs need to construct successful alliances in order to have sustainable business in a competitive environment. Global experiences with randomly chosen collaborators have shown failures that caused the fear of new collaborative work. Causes of failure based on the culture and the type of collaboration are studied [1]. Alliance in new product development has been the focus of industrial researchers [2][3][4].

This study aims to provide a pre-analysis of the path for successful alliances that will lead improvements in innovative power. Both qualitative and quantitative analysis of the SME alliances is realized to find the conditions causing failures and supporting the success in innovation. Support Vector Machine and Self Organized Maps are used to define the most frequent patterns that will give the support and confidence to identify the relationships. Association rules achieved will determine the optimal use of resources.

This paper is so organized that the literature review will be given in the second section and the methodology definition will follow. The fourth section will be reserved for presenting the survey and the results. The conclusion will be given in the fifth and last section.

The implication of the study is generic enough to help any SME or research organization or large business to reduce risks in future alliances.

II. Background

The first research on Association Rule Mining and Methods is found in 1996 [5] trying to find the most frequent occurrences of events to support the linked processes. The research in the field followed the timeline shown in Figure 1.

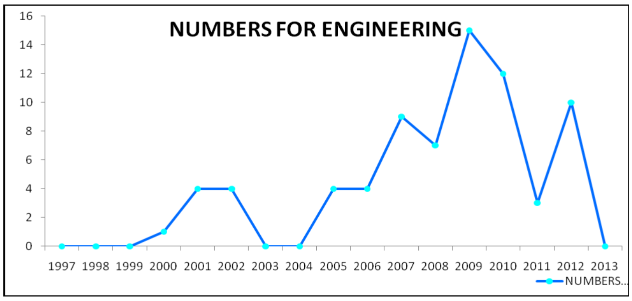

Figure 1. Research timeline on Association Rules

Post et Al. showed the fact that SMEs would like to collaborate only for developing new products [6]. However, Kakabadse et Al. showed that using improved communication and information technologies will improve the SME collaboration[1]. New product based collaboration has evolved fast [7]. Corporation and competition are found as flaming collaboration types that feed the SME improvements in innovation [8]. Association rules defined for the failure types have opened a new dimension for the research on failure of collaboration [9]. The first study on mining the SME innovation by Wang et Al has found some patterns for allocating the R&D resources [10]. Suh & Kim have detailed the R&D collaboration in service industries detected the positive relations of technology and the product or process innovation[11]. Swarnkar et Al. analyzed when and how the collaboration strategies will be used in virtual organizations[12]. Wiltsey et Al. claimed that extent, nature or impact of R&D programs are studied rarely. The interactions among the influences must be given in multiple levels and fidelity and changes must be observed in time [13]. Woodland & Hutton introduced the social dimension on the collaborative success [14]. Both the fear issues and the success causes studied by Bouncken et Al defined technology influencers, sharing the knowledge and learning from the partner as the main influencers [2].

Knowledge management and data mining overviews [15] and Knowledge Management performance studies [16] realized recently do not show any association rule study for the collaborative innovation success and failure.

METHODOLOGIES

A. Association Rules

Given a set of transactions, rules are defined that will exhibit that the occurrence of an item based on the occurrences of other items in the transaction. This is the association analysis. It is useful to explore the interesting relations, which are embedded in the huge data sets. These hidden interactions can be stated in the form of association rules[17]. The strength of an association rule is measured with its support and confidence values. Support shows the how often that rule is applicable to a given dataset. The Confidence is the occurrence frequency of the item in that transaction [5].

Support (s) is the fraction of transactions that contain an itemset

$$\text{Support, } s(X \longrightarrow Y) \;=\; \frac{\sigma(X \cup Y)}{N}; \tag{1}$$

Confidence (c) measures how often items in Y appear in transactions that contain X

$$\text{Confidence, } c(X \longrightarrow Y) \;=\; \frac{\sigma(X \cup Y)}{\sigma(X)}. \tag{2}$$

The item set patterns are found in various methods which could be apriori or aposteriori. The overview of all the methods used in association srule studies are given in Figure 2.

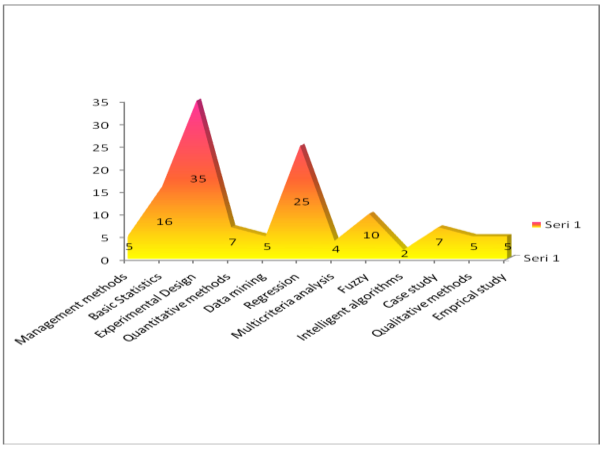

Figure 2. Overview of Association rule techniques

B. Support Vector Machines

It is a machine learning technique, which is mainly introduced for classification in two classes [18] but further used in clustering[19].

It can be analyzed as an optimization problem as in equation 3 [20] relaxed with Lagrange multipliers in objective function as in equation 4.

$$\begin{gathered} \min z = \frac{1}{2}\|w\|^2 \\ \text{s.t.} \\ r_i\left(w_i x_i + w_0\right) \geq 1 \end{gathered} \tag{3}$$

Data is separated with a hyper-plane multiplied by -1 or +1.

$$L_p = \frac{1}{2}\|w\|^2 - \sum_i \lambda_i r_i\left(w_i x_i + w_0\right) + \sum_i \lambda_i \tag{4}$$

Using Using a Gaussian Kernel as defined in Eq. 5. will increase the reliability on dissimilarities [22].

$$K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2}\right) \tag{5}$$

C. Self Organized Maps

Self-Organizing Map (SOM) is a widely used artificial neural network technique in clustering with unsupervised learning algorithm. This technique clusters according to the similarities to the input data [23]. SOMs structure the output with individual node similarity as well as cluster center distance. This technique is based on competitive learning, where the output nodes are made of the winning node activated by one input node. The output nodes would have scoring values using a function, most commonly Euclidean distance between the inputs and weights. For each input vector x, and for each output node j, the value D (w_j, x_n) of the scoring function. Euclidean distance function is shown in Eq. 6

$$Dw_j{,}x_n = (w_{ij} - x_{ni})^2 \tag{6}$$

The winning node therefore becomes the center of a neighborhood of excited nodes. In self- organizing maps, all nodes in the given neighborhood share competition. Therefore, even if the nodes in the output layer are not connected directly to the input layer, they tend to share common features, of the neighborhood [24]. The nodes in the neighborhood of the winning node participate in adaptation, which is, learning. The weights of these nodes are adjusted to improve the weights defined in Eq. 7., until a threshold is reached.

$$w_{ij}\ new = w_{ij}\ current + \alpha\ x_{ni} - w_{ij}\ current \tag{7}$$

In Eq 7. α is the learning rate. If it is necessary, the learning rate and neighborhood size are adjusted.

APPLICATION

A survey is run with the technology firms sited in Techno-parks of linked , 5 are about competences and 4 four the technologychoices. 130 firms responded but only 105 are included in the analysis. 14% of the companies were medium size and 37 % of them were aged more than 10 years. They have chosen the type of collaboration among the SME and Big firms as well as among the customers, suppliers and competitors as shown in Figure 3.

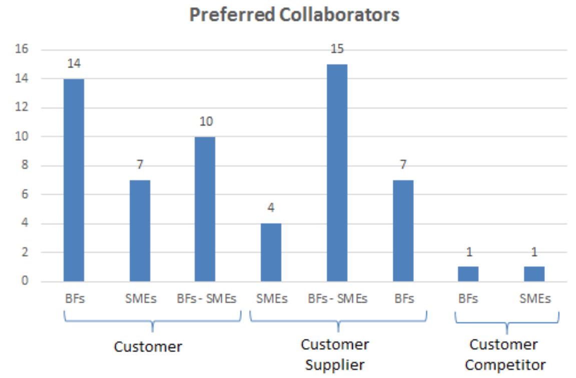

Figure 3. Choice of collaboration according to size and relation

The reason for innovative collaboration is stated as shown in Figure 4.

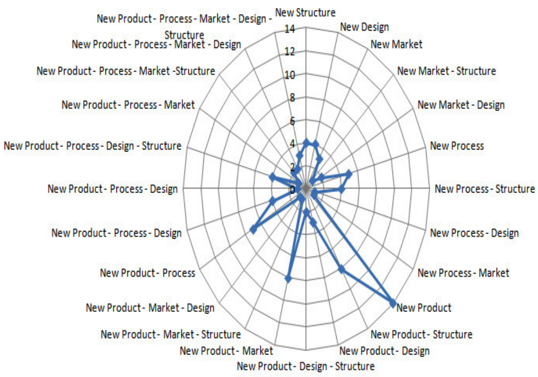

Figure 4. Innovation causes in collaboration

SVM is applied to classify the collaboration made specifically for innovation, validated by Cronbach alpha, that resulted as more than 0.7. SOM is used to cluster those to show the supporting frequencies of choices. Then the cross tables are achieved as a basis for the association rules. The first achievement was micro companies would like all the collaborators to have innovation culture which is less important for the bigger companies. Some other samples of cross tables are given as below.

Table 1 gives the fact that majority of responders prefer for at least one collaborator to have design competence.

Table 1. Business Understanding and Design Competence

INNOV			DESIGN All	DESIGN At least one	DESIGN Negligе
Collaboration for others	Bus-Under	All	4	14	
		At least one	2	15	
Collaboration for innovation	Bus-Under	All	14	26	4
		At least one	7	13	2
		Neglig.	1	1	2

In the firms collaborating for innovation for 1 to 3 years for the success of innovation it is necessary understanding the market requirements by all the firms together with the well-developed innovation culture.

Table 2. Age-Innovation Culture and Market Requirements Relation

COLLAB_Age			INNO_CULT All	INNO_CULT At least one	INNO_CULT Neglige
0	UNDERSTD_REQ	All	13	8	
		More than one	6	6	
		At least one	0	1	
<1 year	UNDERSTD_REQ	All	9	1	0
		More than one	3	1	1
1-3 years	UNDERSTD_REQ	All	13	1	
		More than one	1	1	
		At least one	4	2	
3-5 years	UNDERSTD_REQ	All	6	1	
		More than one	3	2	
		At least one	0	1	
>5	UNDERSTD_REQ	All	8	2	0
		More than one	3	2	2
		At least one	2	2	0

RESULTS

The Reliability analyzes have been done on the results of the questionnaires, the Cornbach's Alpha value is 0.602; this value is in the acceptable range.

In the logistic regression model, the determined significance level is 0.100. The values of attributes of innovation are neglected, since they are greater than 0.100. Innovation related criteria in this study have no significance based on firm size and collaborator types of these firms.

In other words, firms care the technological features but ignore the innovative attributes.

In the table below, the significance values are shown for the other 3 attributes and the results of analyzes for features of the questions. The significance values less than 0.100 are used here (Figure 5).

Significance Table				
		Clusters		
		Finance	*Technology*	*Management*
Properties of Firms				
Firm Size	Firm Size	-	-	-
	Employee Size	0.000	0.000	0.000
	Firm Age	0.007	0.003	0.045
	Collaborate for Innovation	0.111	0.068	0.399
	Collaboration Duration	0.019	0.019	0.027
Employee Size	Firm Size	0.000	0.000	0.000
	Employee Size	-	-	-
	Firm Age	0.152	0.193	0.074
	Collaborate for Innovation	0.497	0.441	0.083
	Collaboration Duration	0.560	0.513	0.013
Firm Age	Firm Size	0.019	0.000	0.000
	Employee Size	0.176	0.002	0.009
	Firm Age	-	-	-
	Collaborate for Innovation	0.203	0.002	0.030
	Collaboration Duration	0.000	0.000	0.000
Collaborate for Innovation	Firm Size	0.006	0.019	0.041
	Employee Size	0.054	0.076	0.084
	Firm Age	0.010	0.025	1.000
	Collaborate for Innovation	-	-	-
	Collaboration Duration	0.000	0.000	0.000
Collaboration Duration	Firm Size	0.099	0.001	0.000
	Employee Size	0.624	0.045	1.000
	Firm Age	0.000	0.000	0.043
	Collaborate for Innovation	0.000	0.000	0.000
	Collaboration Duration	-	-	-

Figure 5. Significance values for each clusters based on demographic properties

After the multinomial logistic regression analysis coefficients for all attributes are obtained. The statistically significant attributes are used. Coefficients of financially related criteria are shown in Figure 6.

Parameter Estimates

SIZE		B	Std. Error	Wald	df	Sig.	Exp(B)	90% Confidence Interval for Exp(B) Lower Bound	Upper Bound
2	Intercept	-6,395	2,530	6,387	1	,011			
	[ComAge_A=0]	2,079	1,020	4,156	1	,041	7,994	1,494	42,771
	[ComAge_A=1]	0			0				
	[ComAge_B=0]	1,725	,758	5,186	1	,023	5,614	1,615	19,522
	[ComAge_B=1]	0			0				
	[ComAge_C=0]	,652	,680	,920	1	,338	1,920	,627	5,877
	[ComAge_C=1]	0			0				
	[ComAge_D=0]	-,752	,732	1,056	1	,304	,472	,142	1,571
	[ComAge_D=1]	0			0				
	[Cap_A=0]	,055	,611	,008	1	,928	1,057	,387	2,889
	[Cap_A=1]	0			0				
	[Cap_B=0]	,145	,718	,041	1	,840	1,156	,355	3,768
	[Cap_B=1]	0			0				
	[Innov_Op_A=0]	,184	,716	,066	1	,797	1,203	,370	3,906
	[Innov_Op_A=1]	0			0				
	[Innov_Op_B=0]	,741	,714	1,075	1	,300	2,098	,648	6,794
	[Innov_Op_B=1]	0			0				
	[Price_A=0]	-,504	,927	,296	1	,587	,604	,132	2,775
	[Price_A=1]	0			0				
	[Price_B=0]	,473	,973	,237	1	,627	1,606	,324	7,957
	[Price_B=1]	0			0				
	[Export_A=0]	2,347	,910	6,659	1	,010	10,459	2,342	46,706
	[Export_A=1]	0			0				
	[Export_B=0]	1,839	,792	5,389	1	,020	6,289	1,709	23,142
	[Export_B=1]	0			0				

Figure 6. Regression Coefficients & Significancies

The cross relation tables have been constructed to define the rules obtained from the model.

Finance Related Criteria :

RULE 1:

IF

(Firm Size = “Micro” AND Firm Age ≤ “1”)

THEN

(Innovation operation expenditure = “proportionally shared” AND Price = “important”)

MEANING:

✓ Preferences for the initiating micro SMEs emphasize the innovation operation expenditures according to the collaborator sharings and market value (price) of the innovated product (Figure 7).

Finance Related Criteria		Capital_More than Average of Sector (A)	Capital_Average of Sector (B)	Innovation Operation Expenditure_Balance (A)	Innovation Operation Expenditure_More is more (B)	Price_Important (A)	Price_Only motivation (B)	Exportation_All collaborators (A)	Exportation_Only one (B)
Firm Size	Firm Age								
MICRO (1)	≤ 1 YEAR OLD (A)								
	1-3 YEARS OLD (B)								
	3-5 YEARS OLD ©								
	5-10 YEARS OLD (D)								
SMALL (2)	< 1 YEAR OLD (A)								
	1-3 YEARS OLD (B)								
	3-5 YEARS OLD ©								
	5-10 YEARS OLD (D)								
MEDIUM (3)	< 1 YEAR OLD (A)								
	1-3 YEARS OLD (B)								
	3-5 YEARS OLD ©								
	5-10 YEARS OLD (D)								

Figure 7. Cross – relations Table: Firm Size – Firm Age – Finance Related Criteria

RULE 2:

IF

(Firm Size = “Small” AND Collaborate = “Large Firms”)

THEN

(Capital = “more than average” AND Exportation facilities = “only one firm”)

MEANING:

✓ Small SMEs emphasize that capital of the collaborators are to be more than the sector average and exportation facilities are done by only one collaborator for innovation (Figure 8).

Finance related Criteria		Capital_More than Average of Sector	Capital_Average of Sector	Innovation Operation Expenditure_Balance	Innovation Operation Expenditure_More is more	Price_Important	Price_Only motivation	Exportation_All collab	Exportation_Only one
COLL	Firm Size								
Cust – LFs	Micro								
	Small								
	Medium								
Cust – SMEs	Micro								
	Small								
	Medium								
Cust-Supp – LFs	Micro								
	Small								
	Medium								
Cust-Supp – SMEs - LFs	Micro								
	Small								
	Medium								
Cust	Micro								
	Small								
	Medium								
LFs	Micro								
	Small								
	Medium								
Cust-Supp-SMEs	Micro								
	Small								
	Medium								

Figure 8: Cross – relations Table: Collaboration – Firm Size – Finance Related Criteria

RULE 3:
IF
(Firm Age = "> 10 years old" AND Collaboration Duration = "3-5 years")
THEN
(Capital = "more than average" AND Price = "motivation")

MEANING:

- ✓ 10 years (and more) old SMEs emphasize capital of the collaborators have to be more than the sector average whereas market value (price) of the innovated product is only a motivation.

Technology Related Criteria :

IF
(Firm Size = "Small" AND Firm Age = "3-5 years old")
THEN
(Connectivity = "All possible ways" AND Change Management = "All Collaborators")
MEANING:

- ✓ SMEs with age 5 to 10 years old emphasize change management is to be applied by all collaborators and use all possible connectivity possibilities.

Technology Related Criteria		MIS_Least one	MIS_All	Comm_If tech exist	All tech	ChngMng_individual	ChngMng_Together	Connect_only one type	Connect_All ways
COLL	Firm Size								
Cust-LFs	Micro								
	Small								
	Medium								
Cust – SMEs	Micro								
	Small								
	Medium								

Figure 9. Cross – relations Table: Collaboration – Firm Size – Technological Criteria

Figure 9 shows the cross – relations between collaborator type (customer – supplier – SMEs – large firms) and firms' size for the technology related criteria. Also this mentioned model makes sense statistically significant as a result of the logistic regression analysis.

RULE 2:
IF
(Firm Size = "Small" years old" AND Collaborate = "Customers" and "LFs")
THEN
(Communication technologies = "All opportunities" AND Change Management = "Individual")
MEANING:

- ✓ Small SMEs who are collaborating with the customers which are large firms (LFs) emphasize usage of all communication technologies opportunities and change management is can be individual choice (Figure 9).

Management Related Criteria :

RULE 1:
IF
(Firm Size = "Medium" AND Firm Age = "5-10 years old")
THEN
(Professionalism = "Motivation" AND Organizational Structure = "Effective" AND Cooperation & Coordination = "All" AND Leadership = "Only one")
MEANING:

- ✓ Medium size SMEs of age 5-10 years prefer to collaborate with companies which have effective organizational structure with both cooperation and coordination attitude; in the collaboration a single leader is preferred and professionalism can be taken only as the motivator.

RULE 2:
IF
(Firm Size = "Micro" AND Collaborate = "Customers" and "SMEs")
THEN
(Professionalism = "All" AND Business Experience = "All" AND Leadership = "All")
MEANING:

- ✓ The micro SMEs collaborating with the customers emphasize professionalism, business experience for the problem solving and they prefer all the collaborators to have the leadership features.

CONCLUSION

This study investigates the most preferred conditions for a successful collaboration for innovative SMEs. SVM and SOM are used to construct the basis for creating the association rules. As the result of a survey in Turkey, there are hundreds of relations depicted in the analysis.

The achievements are interesting enough to show that the technology companies are confused in differentiating the technology and innovation concepts. It was interesting to observe micro and young companies not willing to collaborate with the big and overwhelming companies. Everybody asks for full communication technology, but only small SME with 5-10 years of experience ask for the collaborators to have effective organization and full professionalism.

The validation by logistic regression on the same data is in process. All the results achieved using logistics regression will be cross-validated with machine learning application results. Future survey will be aiming to improve the innovation concept of the technology firms in detail.

References

[1] Kakabadse, N.K.; Kakabadse, A.; Ahmed, P.K.; Kouzmin, "The ASP phenomenon: an example of solution innovation that liberates organization from technology or captures it?," *Eur. J. Innov. Manag.*, vol. 7, no. 2, pp. 113–127, 2004.

[2] S. Bouncken, Ricarda B.; Kraus, "Innovation in knowledge-intensive industries: The double-edged sword of coopetition," *J. Bus. Res.*, vol. 66, no. 10, pp. 2060–2070, 2013.

[3] D. R. Gnyawali and B. R. Park, "Co-opetition and technological innovation in small and medium sized enterprizes A Multilevel Conceptual Model," *J. Small Bus. Manag.*, vol. 47, no. 3, pp. 308–330, 2009.

[4] R. Narula, "R&D collaboration by SMEs: New opportunities and limitations in the face of globalisation," *Technovation*, vol. 24, no. 2, pp. 153–161, 2004.

[5] R. Agrawal, "Parallel mining of association rules," *IEE Tran. Knowl. Data Eng.*, vol. 8, no. 6, pp. 962–969, 1996.

[6] G. J. J. Post, L. Hop, and J. E. van Aken, "Indicators for establishing SME product development networks," *J. Sci. Ind. Res.*, vol. 60, no. 3, pp. 264–276, 2001.

[7] R. Narula, "R&D collaboration by SMEs: New opportunities and limitations in the face of globalisation," *Technovation*, vol. 24, no. 2, pp. 153–161, 2004.

[8] D. R. Gnyawali and B. R. Park, "Co-opetition and technological innovation in small and medium sized enterprizes A Multilevel Conceptual Model," *J. Small Bus. Manag.*, vol. 47, no. 3, pp. 308–330, 2009.

[9] Z. Z. Z. Zheng, Z. L. Z. Lan, B. H. Park, and A. Geist, "System log pre-processing to improve failure prediction," *2009 IEEE/IFIP Int. Conf. Dependable Syst. Networks*, 2009.

[10] C. H. Wang, Y. C. Chin, and G. H. Tzeng, "Mining the R&D innovation performance processes for high-tech firms based on rough set theory," *Technovation*, vol. 30, pp. 447–458, 2010.

[11] Y. Suh and M.-S. Kim, "Effects of SME collaboration on R&D in the service sector in open innovation," *Innovation: Management, Policy & Practice*, vol. 14, no. 3. pp. 349–362, 2012.

[12] R. Swarnkar, A. K. Choudhary, J. A. Harding, B. P. Das, and R. I. Young, "A framework for collaboration moderator services to support knowledge based collaboration," *Journal of Intelligent Manufacturing*, vol. 23. pp. 2003–2023, 2012.

[13] S. Wiltsey Stirman, J. Kimberly, N. Cook, A. Calloway, F. Castro, and M. Charns, "The sustainability of new programs and innovations: a review of the empirical literature and recommendations for future research," *Implementation Science*, vol. 7. p. 17, 2012.

[14] R. H. Woodland and M. S. Hutton, "Evaluating Organizational Collaborations: Suggested Entry Points and Strategies," *American Journal of Evaluation*, vol. 33. pp. 366–383, 2012.

[15] H. H. Tsai, "Knowledge management vs. data mining: Research trend, forecast and citation approach," *Expert Syst. Appl.*, vol. 40, no. 8, pp. 3160–3173, 2013.

[16] I. B. Tae Hun Kim, Jae-Nam Lee, Jae Uk Chun, "Understanding the effect of knowledge management strategies on knowledge management performance: A contingency perspective," *Inf. Manag.*, vol. 51, pp. 398–416, 2014.

[17] J. Jackson, "Data mining: A conceptual overview," *Commun. Assoc. Inf. Syst.*, vol. 8, pp. 267–296, 2002.

[18] C.Cortes and V. Vapnik, "Support Vector Networks," *Mach. Learn.*, vol. 20, pp. 273-297, 1995.

[19] T. Finley and T. Joachims, "Supervised clustering with support vector machines," in *Proceedings of the 22nd International Conference on Machine learning (ICML)*, 2005, pp. 217–224.

[20] S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice-Hall, New Jersey, 1999.

[21] E. Alpaydın, Machine Learning, Massachusetts Institute of Technology, USA, 2004.

[22] V. Cherkassky, Y. Ma, Practical selection of SVM parameters and noise estimation for SVM regression, Neural Networks 17 (2004) 113–126.

[23] E. Leopold, M. May, and G. Paaß, "Data Mining and Text Mining for Science & Technology Research," in *Handbook of Quantitative Science and Technology Research - The Use of Publication and Patent Statistics in Studies of S&T Systems*, H. F. Moed, W. Glänzel, and U. Schmoch, Eds. Springer, 2004, pp. 187–213.

[24] Larose, D.T. 2005. Discovering Knowledge in Data: An Introduction to Data Mining. New Jersey: John Wiley & Sons, Inc.