

APPLICATION OF SOCIAL COMPUTING ANALYSIS VIA ATTITUDES OF MEDIA USERS

Turhan CİVELEK

Department of Software Engineering, The University of Kirkkareli Turkey

Abstract

Analyzing and modeling of social platforms and media data are rather developed and still developing today. Social behaviors are analyzed and characterized with social and artificial intelligence, statistical methods, analytical approaches and evolutionary algorithms. Data mining and machine learning are used in the social computing. In this study, social computing included, two surveys used a questionnaire with 43 different questions handed over in random sample to 537 different high school students in Istanbul. Social media attitudes of students were evaluated using J48 algorithm. According to J48 algorithm, which constitutes machine learning, learning took place for 14 questions out of 43. The analysis produced a Kappa Statistic (K) value of 0.446 and F-Measure of 0.701. A decision tree was created to support decision-making.

Keywords: Social Computing, Machine Learning, J48 Algorithm, Decision Tree.

Özet

Günümüzde sosyal platform ve medya verilerinin analiz ve modellenmesi gelişmektedir. Sosyal zeka, yapay zeka, istatistiksel yöntemler, analitik yaklaşımlar ve evrimsel algoritmalarla sosyal davranışlar analiz ve karakterize edilmektedir. Veri madenciliği ve makine öğrenmesi sosyal bilgi işlem analizinde kullanılmaktadır. Sosyal bilgi işlem içerikli bu çalışmada İstanbul'da rastgele seçilen farklı liselerden 537 lise öğrencisine iki anket yapılmıştır. Anketler sonucunda elde edilen veriler J48 algoritması ile analiz edilmiştir. Makine öğrenmesi gerçekleştiren J48 algoritması 43 sorunun 14 sorusunda öğrenme gerçekleştirmiştir. Analiz sonucunda kappa değeri (K) 0.446 ve F değeri 0.701 bulunmuştur. Karar vermeyi desteklemek için karar ağacı elde edilmiştir.

Anahtar Kelimeler: Sosyal Bilgi İşlem, Makine Öğrenmesi, J48 Algoritması, Karar Ağacı.

Corresponding author. Tel: +90-288-2140514; fax: +90-288-2140516; e-mail:

turhancivelek@klu.edu.tr

1. INTRODUCTION

Social computing is a new and developing paradigm, which includes modeling and analysis of social behaviors [1]. It is defined as a knowledge and communication technique design and usage tool as well as a tool for calculation simplification of human social dynamics and works placing social content in focus [2]. It produces results and applications for intelligent and interactive practices of multi-disciplinary approaches by analyzing and modeling various data from media and social platforms. It supports various areas such as human-computer interaction, computer, sociology, social psychology, and organization theory [3].

Social computing, as it is introduced, as a new concept based on artificial systems, is advancing beyond processing of just social data but towards social intelligence such as artificial intelligence [4, 5]. Computer simulations, using methods such as artificial intelligence, complex statistical methods, analytical approaches and evolutionary algorithms model social behaviors [6].

Social computing encompassing calculation paradigms and technology applications, helps scientists in evaluation of individual and organizational behavior and the management and implementation of emergency situations [7]. The social data processing is composed of modeling a community with artificial systems, analyses and evaluation using computational tools, management and control of a real community using parallel execution.

Similar studies in different areas, such as in socio-economic systems [8], e-commerce [9], emergency rescues [10], operations management against the risk of environmental accidents [11], determination of amount of passengers at stations in real time [12], safety efficiency in rail transport systems [13], personal and social group behaviors, social networking sites [14, 15, 16, 17] have been reported previously.

Machine learning techniques can effectively be used in social process modeling and decision-making [18, 19, 20, 21]. It evaluates the data and performed improvements in decision-making in, among others, medical diagnostics, image and face recognition, signal processing, optimization of network, artificial intelligence, robotics and psychology [22]. Machine learning algorithms can be used for reporting individual and social behavior, which constitutes one supporting point of the computational analysis and evaluation process based on social data processing. Machine learning is preferred in development of computer learning techniques, and

production of new solutions based on analysis of the data stacks of past learning information [23].

J. Ross Quinlan developed renewable J48 algorithm, based on information gain theory. It is a decision tree algorithm, which has C4.5 algorithms in the background. It is one of the best algorithms, which performs the best learning from the samples [24, 25]. The fundamentals of it are placed on divide and conquer rule. It constructs a decision tree by dividing and choosing related samples using IF-THEN conditions [26]. J48 algorithm produces membership function sets output and decision tree interfaces by dividing samples from the point where information gain, according to IF-THEN rule, is the best. Decision trees are tools that facilitate understanding sampled data, support decision-making and have influential accuracy for graphical presentation [23]. Data is divided after selecting its best parameter in the data domain. Decision tree is structured top-down, beginning from the best root parameter.

J48 algorithm can effectively prune meaningless branches. The reason for this is that the decision tree does not explore data but it creates a simple classification model of them. Furthermore, decision trees prevent finding more quality rules because they use only heuristic search. They follow solely a single path and thereby miss alternatives including better rules. This paper intends to remedy this through proposing a decision tree algorithm to explore patterns of useful business intelligence that it is made by a high-profit-divide method as in [27].

2. DATA and METHOD of ANALYSIS

This research was carried out to identify approaches of high school students' attitudes in connection with the social media usage and to their loneliness [28]. Two surveys used questionnaires with 23 and 20 different questions that were handed over in random sample to 537 high school students in Istanbul. Obtained data from the survey were transformed to .arff format and analyzed by using WEKA 3.7.12 (Waikato Environment for Knowledge Analysis) which is a popular suite of machine learning software. It contains a collection of visualization tools and algorithms for data analysis and predictive modeling with graphical user interfaces for easy access to these functions [29, 30]. J48 and ZeroR classification algorithms from the WEKA library evaluated the data from the surveys. In this endeavor, the success rate of each questionnaire item by algorithms performed-classification was targeted. The machine-learning

algorithm J48 performed the most success in the respond of “*I do not allocate enough time to the other social activities because of social media sites*”

It is desired to know the success rates of machine learning algorithms in the analysis they are used. Accuracy rate, precision, recall and F-measure are used to compare success rates of the algorithms [31]. The class of data is predicted true or wrong when evaluated by a classification algorithm. WEKA produces a confusion matrix which is a digital output summary of made predictions.

Accuracy rate is used to measure model performance. It is the ratio of correctly classified samples number (TP +TN) to total number of samples (TP+TN+FP+FN).

$$\text{Accuracy rate} = \frac{(TP + TN)}{(TP + FP + FN + TN)} \quad (1)$$

Error Rate is the complementary part, which completes the accuracy rate to 1. It is the ratio of incorrectly classified number samples (FP+FN) to total number of samples (TP+TN+FP+FN).

$$\text{Error Rate} = \frac{(FP + FN)}{(TP + FP + FN + TN)} \quad (2)$$

Precision is the ratio of true positive samples number (TP) whose class is estimated to be 1 to total samples number (TP+FP) whose class is estimated to be 1.

$$\text{Precision (P)} = \frac{TP}{(TP + FP)} \quad (3)$$

Recall is the ratio of the correctly classified positive samples number to total positive samples number (TP+FN).

$$\text{Recall } (R) = \frac{TP}{(TP + FN)} \quad (4)$$

Precision and recall are not enough alone to extract meaningful results in inter-observer comparisons. *F-measure*, which is harmonic average of precision and recall, is defined to solve this problem [32]. The output of F-measure value has better learnt the relationship between algorithm classes that are close to 1. F-measure is calculated as follows:

$$\text{F-measure } (F) = \frac{2DK}{(D + K)} \quad (5)$$

Kappa value (K) determines the concurrence reliability of comparison between observers [33, 34]. The interpretable interval of Kappa-value is between -1 and 1. K is 1 when there is an exact concurrence comparison between observers. Negative values of kappa ($K < 0$) are meaningless in terms of reliability. The range of 0.41 and 1.0 is acceptable value for K [35]. If P_o is acceptable rate and P_c is the expected rate, K is calculated as follows:

$$\text{Kappa value } (K) = \frac{(P_o - P_c)}{(1 - P_c)} \quad (6)$$

3. FINDINGS

In Table 1, While 329 ((%61.27) of 537 samples were classified as true with ZeroR primitive algorithm, 208 (%38.73) of them were classified as false in WEKA evaluation as shown in Table 2. Kappa Statistic value was 0. While 385 (%71.69) of 537 samples were classified as true with J48 classification algorithm, 152 (%28.31) of them were classified as false. Kappa Statistic value was 0.446.

Table 1 Confusion Matrix (J48 and ZeroR)

J48 Confusion Matrix					ZeroR Confusion Matrix				
a	b	c	←classified as		a	b	c	←classified as	
290	17	22	I	a=1	329	0	0	I	a=1
36	21	21	I	b=2	78	0	0	I	b=2
44	12	74	I	c=3	130	0	0	I	c=3

The true classification number of J48 algorithm is bigger than true classification number of ZeroR algorithm. This indicates that J48 algorithm accomplishes learning. If J48 kappa Statistic value (0.45) is between 0.41 and 0.60, the success of learning of J48 algorithm is moderate and acceptable [36].

Table 2 Results of Analyzing Survey Data with WEKA Program

Tipi	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
J48	0.717	0.270	0.694	0.717	0.701	0.800
ZeroR	0.613	0.613	0.375	0.613	0.466	0.495

In Table 2, It is seen that $F(J48) > F(ZeroR)$ in Table 3. The more F-Measure value of J48 algorithm is close to 1 the more successful machine learning is. Analysis of survey data by WEKA proves that, J48 algorithm makes the best learning. The decision tree in Fig. 1 was obtained according to learning style, which was realized by J48 algorithm in WEKA program.

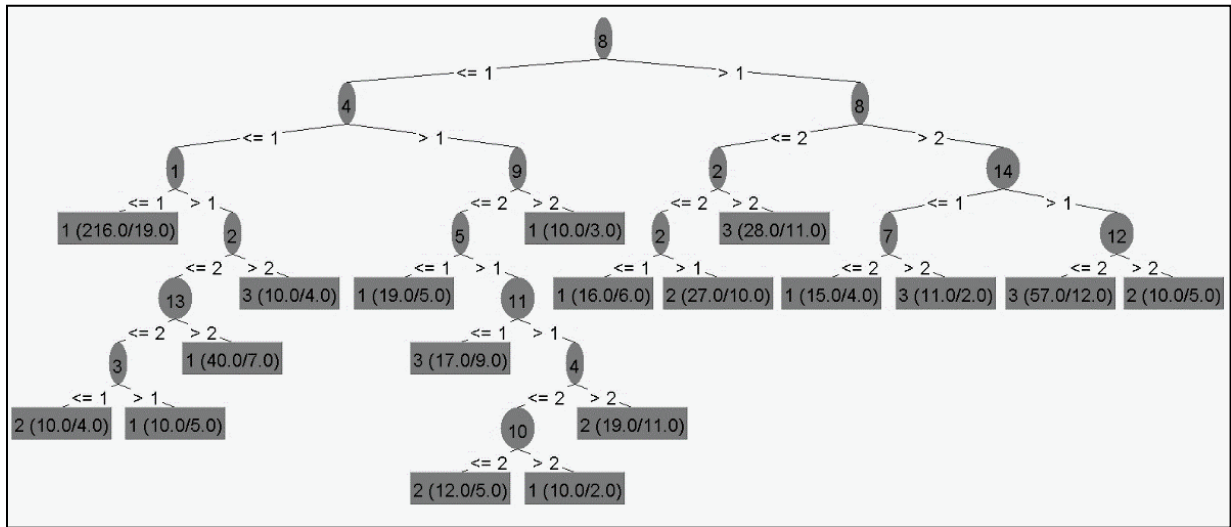


Figure 1 J48 Decision Tree

The machine learning algorithm J48 was used and the basic question, “*I do not allocate enough time to the other social activities because of social media sites*” was asked to the participants to determine whether learning took place or not. Learning took place successfully for the following 14 expressions in 43 based on the learning questions. They were:

1. I feel I gain a new personality by means of social media sites,
2. I think the social media sites break off me from my family,
3. I think I get rid of loneliness by means of social media sites,
4. I do not allocate enough time to my family because of social media sites,
5. I like the spread of my social media messages among friends,
6. I do not allocate enough time to the other social activities because of social media sites,
7. I express my feelings more to the friends of mine to whom I have special interests comfortably by the agency of social media sites,
8. I do not allocate enough time to my friends because of social media sites,
9. I do not have anybody who can help,
10. I have very many things in common with people around me,
11. I can find friends when I want,
12. I am unhappy due to the fact that I am withdrawn into my shell to such a high degree,
13. I have people to whom I can talk,

14. I have people to whom I can tell my concerns.

The algorithm made a simple classification of the data set. It divided the data by choosing the best root variable, and built the decision tree from top to down.

The best root variable was placed to top, and according to gain condition, other root variables were aligned downwards. Therefore, the graphical presentation of the decision tree created with high accuracy as shown in Figure 1. Decision tree, built as explained above, begins by choosing the best root variable and data. Attributes are established as points of nodes, and leaves, thus produced class labels.

The question of if social media users allocated enough time to social activities was investigated using machine learning performed by J48 algorithm in WEKA. J48 accomplished high degree of learning for question 8. Using social media consumers, that machine learning of J48 algorithms was accomplished regarding the answer, *'I can't spare enough time for other activities because of social media sites'* of persons who allocated enough time to their friends;

- Two hundred sixty-four (%49.2) persons allocated time to social activities. From others seventy-one persons did not allocate time for social activities,
- Thirty-two (%6) persons occasionally allocated time or did not allocate time their families allocated time social activities,
- Of persons who allocated time for their families within persons who allocated time for their friends;
 - Two hundred thirty-five (%43.8) persons allocated time for social activities,
 - One hundred ninety-seven (%36.7) persons, who did not acquire new personality, allocated time for social activities,
 - Thirty-three persons (%6.1) of those who acquired new personality and considered themselves that they were not detached from their families or felt sometimes that they were and they had persons in their contact circles to talk declared that they allocated time for social activities.
- Of persons who did not spare time for their friends but had people around or occasionally had people around with whom they could talk about their concerns,
 - Fifty (%9.3) persons occasionally allocated time or did not allocate time for social activities,

- Forty-five (%8.4) persons who were occasionally unhappy or were unhappy, due to have been closed to outer world, did not allocate time for social activities, have been learnt by the J48 algorithm correctly.

4. DISCUSSION and RESULTS

In this study, individual effects of social media on media users were evaluated, and social attitudes of social media users were characterized by machine learning. The findings obtained here supported previous studies which were carried out by Macy & Willer (2002) and those by Mao & Wang (2012). The result of data analysis of the survey with WEKA program proved that J48 algorithm made the best learning and a decision tree was obtained as function of learning styles by J48 algorithm in WEKA program.

It was observed that those of social media users who allocated time for their families allocated time their friends as well. J48 algorithm performed best extraction based on the best root variable and presented it graphically. Results and the presentation of decision tree based sports activities supported pervious analogous studies which reported by Sadeh, Hong, Cranor, Fette, Kelley, Prabaker, Rao, Danezis (2009).

On the whole, the WEKA analysis accomplished learning correctly regarding the fact that 290 (%54) persons allocated time for social activities, 21 persons (%13.78) allocated occasionally and that 74 (%13.78) persons did not allocate time. It did not obtain learning about the social activity period of 152 (%28.31) persons.

Social-oriented data processed by machine learning evaluation can support related research. They can develop strategies with the guidance of gains and outputs based on individuals and organizations which they reported by King, Li, & Chan, (2009). Statistical values and decision trees' graphical interfaces can support decision-making processes about social computing. Obtained results supported pervious analogous studies which reported by Fang &Lefevre (2010), Xie, Knijnenburg & Jin (2014). As it is done in this research, machine-learning algorithms can be used in computational analysis and evaluation, which constitutes one part of the social data processing, to analyze and document behaviors of individuals and groups.

REFERENCES

- [1] F.-Y. Wang, «Toward a paradigm shift in social computing: The acp approach,» *cilt 22*, no. 5, pp. 65-67, 2007.
- [2] F.-Y. Wang, D. Zeng, K. M. Carley ve W. Mao, «Social computing: From social informatics to social intelligence,» *IEEE Intelligent Systems*, cilt 22, no. 2, pp. 79-83, 2007.
- [3] I. King, J. Li ve K. T. Chan, «A Brief Survey of Computational Approaches in Social Computing,» *IJCNN'09 Proceedings of the 2009 international joint conference on Neural Networks*, NJ,USA, 2009.
- [4] F.-Y. Wang, «Computational theory and method on complex system,» *China Basic Science*, cilt 6, no. 5, p. 3–10, 2004.
- [5] F.-Y. Wang ve S. T. Tang, «Artificial societies for integrated and sustainable development of metropolitan systems,» *IEEE Intelligent Systems*, cilt 19, no. 4, pp. 82-87, 2004.
- [6] M. W. Macy ve R. Willer, «From Factors to Actors: Computational Sociology and Agent-Based Modeling,» *Annual Review of Sociology*, no. 28, pp. 143-166, 2002.
- [7] W. Mao ve F. Wang , *Advances in Intelligence and Security Informatics*, Oxford, UK: Academic Press, 2012.
- [8] D. Wen, Y. Yuan ve X. R. Li, «Artificial Societies, Computational Experiments, and Parallel Systems: An Investigation on a Computational Theory for Complex Socioeconomic Systems,» *IEEE Transactions on Services Computing*, cilt 6, no. 2, pp. 177-185, 2015.
- [9] F. -y. Wang, D. -j. Zeng ve Y. Yuan, «An ACP-based Approach for Complexity Analysis of E-commerce System,» *Complex Systems and Complexity Science*, cilt 5, no. 3, pp. 1-8, 2008.
- [10] J. Sifeng, X. Gang, F. Dong ve H. Chunpeng, «Study on the emergency rescue decision support system of petrochemical plant based on ACP theory,» *in Proceedings of the 6th Management Annual Meeting*, 2011.
- [11] G. Xiong, X. Dong, D. Fan ve F. Zhu, «Parallel Bus Rapid Transit (BRT) operation management system based on ACP approach,» *Networking, Sensing and Control (ICNSC), 2012 9th IEEE International Conference on*, 2012.
- [12] B. Ning, F.-y. Wang, H.-r. Dong , R.-m. Li, D. Wen ve L. Li, «Parallel Systems for Urban Rail Transportation Based on ACP Approach,» *Journal of Transportation Systems Engineering and Information Technology*, cilt 10, no. 6, pp. 23-28, 2010.
- [13] F. -Y. Wang, J. Zhao ve S. -X. Lun, «Artificial power systems for the operation and management of complex power grids,» *Southern Power System Technology*, cilt 2, no. 3, pp. 1-11, 2008.
- [14] P. Singla ve M. Richardson, «Yes, There is a Correlation From Social Networks to Personal Behavior on the Web,» *WWW '08 Proceedings of the 17th international conference on World Wide Web*, New York, 2008.
- [15] B. He, C. Macdonald ve I. Ounis, «Ranking opinionated blog posts using opinionfinder,» *SIGIR '08 Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information*, New York, 2008.

- [16] A. Sun, M. Hu ve E. -P. Lim, «Searching blogs and news: a study on popular queries,» *SIGIR '08 Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, New York, 2008.
- [17] J. L. Elsas, J. Arguello, C. Jamie ve J. G. Carbonell, «Retrieval and feedback models for blog feed search,» *SIGIR '08 Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, New York, 2008.
- [18] N. Sadeh, J. Hong, L. Cranor, I. Fette, P. Kelley, M. Prabaker ve J. Rao, «Understanding and capturing people's privacy policies in a mobile social networking application,» *Personal and Ubiquitous Computing. Springer*, cilt 13, no. 6, pp. 401-412, 2009.
- [19] G. Danezis, «Inferring privacy policies for social networking services,» *AISec '09 Proceedings of the 2nd ACM workshop on Security and artificial intelligence*, New York, 2009.
- [20] L. Fang ve K. LeFevre, «Privacy wizards for social networking sites,» *WWW '10 Proceedings of the 19th international conference on World wide web*, New York, 2010.
- [21] J. Xie, B. P. Knijnenburg ve H. Jin, «Location sharing privacy preference: analysis and personalized recommendation,» *IUI '14 Proceedings of the 19th international conference on Intelligent User Interfaces*, New York, 2014.
- [22] E. Alpaydin, «Introduction to Machine Learning,» *Introduction to Machine Learning*, USA, Barnes&Noble, 2004, pp. 350-368.
- [23] D. -A. SITAR-TĂUT ve A. -V. SITAR-TĂUT, «Overview on How Data Mining Tools May Support Cardiovascular Disease,» *Journal of Applied Computer Science & Mathematics*, cilt 8, no. 4, pp. 57-62, 2010.
- [24] L. Gaga, V. S. Moustakis, Y. Vlachakis ve G. Charissis, «ID+: Enhancing medical knowledge acquisition with machine learning,» *dx.doi.org*, cilt 10, no. 2, pp. 79-94, 1996.
- [25] J. R. QUINLAN, «Induction of Decision Trees, Machine Learning,» *Kluwer Academic Publishers*, cilt 1, pp. 81-106, 1986.
- [26] K. Mollazade, H. Ahmadi, M. Omid ve R. Alimardani, «An Intelligent Combined Method Based on Power Spectral Density, Decision Trees and Fuzzy Logic for Hydraulic Pumps Fault Diagnosis,» *World Academy of Science, Engineering and Technology International Journal of Mechanical, Aerospace, Industrial, Mechatronic and Manufacturing Engineering*, cilt 2, no. 8, p. 986, 2008.
- [27] J. Li, A. W. Fu ve P. Fahey, «Efficient discovery of risk patterns in medical data,» *Efficient discovery of risk patterns in medical data, Elsevier*, cilt 45, no. 1, pp. 77-89, 2009.
- [28] F. S. Argin, *Ortaokul ve Lise Öğrencilerinin Sosyal Medyaya İlişkin Tutumlarının İncelenmesi*, İstanbul: Yeditepe Üniversitesi, 2013.
- [29] L. H. Witten, E. Frank ve M. A. Hall, *Data mining: practical machine learning tools and techniques*, New York: Morgan Kaufmann. Elsevier., 2011.
- [30] «<http://www.cs.waikato.ac.nz/ml/weka/>,» 20 Aralık 2015. Available: <http://www.cs.waikato.ac.nz/ml/weka/>.
- [31] E. K. Aydoğan, C. Gencer ve S. Akbulut, «Churn Analysis And Customer Segmentation Of A Cosmetics Brand Using Data Mining Techniques,» *Journal of Engineering and Natural Sciences*, cilt 26, no. 1, pp. 42-56, 2008.

- [32] J. Han ve M. Kamber, Data Mining Concepts and Techniques, Francisco, CA: Morgan Kaufmann, Elsevier inc, 2006.
- [33] J. Cohen, «A Coefficient of Agreement For Nominal Scales,» *Educational and Psychological Measurement*, cilt 20, pp. 37-46, 1960.
- [34] J. . L. Fleiss, «Measuring nominal scale agreement among many raters,» *Psychological Bulletin*, cilt 76, no. 5, pp. 378-382, 1971.
- [35] J. R. Landis ve G. G. Koch , «The measurement of observer agreement for categorical data,» *PubMed* , cilt 33, no. 1, pp. 159-74, 1977.
- [36] P. Brennan ve A. J. Silman, «Statistical Methods for Assessing Observer Variability in Clinical Measures,» *BMJ Clinical Research*, cilt 304, no. 6840, pp. 1491-4, 1992.