

Employing data mining to explore association rules in drug addicts

F. Zahedi^{1*} and M. R. Zare-Mirakabad²

1. Department of Engineering, College of Computer Engineering, Yazd Science and Research Branch, Islamic Azad University, Yazd, Iran.
2. School of Electrical and Computer Engineering, Department of Computer Engineering, Yazd University, Yazd, Iran.

Received 18 January 2014; accepted 15 February 2014
*Corresponding author: f.zahedi@rvu.ac.ir (F. Zahedi).

Abstract

Drug addiction is a major social, economic and hygienic challenge that impacts on all the community and needs serious threat. Available treatments are only successful in short-term unless underlying reasons making individuals prone to the phenomenon are not investigated. Nowadays, there are some treatment centers which have comprehensive information about addicted people. Therefore, given the huge data sources, data mining can be used to explore knowledge implicit in them; their results can be employed as knowledge-based support systems to make decisions regarding addiction prevention and treatment. We studied 471 participants in such clinics, where 86.2% were male and 13.8% were female. The study aimed to extract rules from the collected data by using association models. Results can be used by rehab clinics to give more knowledge regarding relationships between various parameters and help them for better and more effective treatments. The finding shows that there is a significant relationship between individual characteristics and LSD abuse, individual characteristics, the kind of narcotics taken, and committing crimes, family history of drug addiction and family member drug addiction.

Keywords: *Drug Addiction, Data Mining, Association Rules, Rules Discovery.*

1. Introduction

Nowadays, most societies face the serious challenge of drug addiction, which is often associated with social problems. Yet, no sure method has been proposed to treat it though much research has been conducted in this regard. Drug addiction is closely related to cultural, religious, economic, social, and historical aspects of a community and is affected by different factors such as biological, psychological, and social ones[1].

It affects individuals' physical, social, and mental performance and increases the number of hepatitis and AIDS victims. In particular, the youth has been heavily influenced by this phenomenon, which can lead to mental and behavioral deviations [2]. Considering that in many cases addiction-related treatments are not successful, identifying factors influencing drug addiction is of paramount importance for preventive purposes. That is, preventive steps and appropriate treatments reduce risk behaviors and

consequently, mental and physical impacts of the behaviors and treatment costs are lessened [3].

On the other hand, using data mining to explore health-related issues proposes a lot of potentials and advantages and it has been shown that data mining algorithms dramatically diminish diagnosis costs and health-related risks. It should be noted that the success of data mining in health systems is a function of employing and having access to valid and comprehensive data. Then, efficient procedures for collecting, storing, and processing the data are vital [4].

In addition, using machine learning techniques in medicine and psychiatry offers a totally new research avenue for monitoring, diagnosing, and categorizing psychiatric disorders such as schizophrenia, psychosis, depression, anxiety, and alcohol abuse [5]. According to [6], qualitative questionnaires and statistical procedures are dominantly used in social studies. The authors suggest that data mining can be introduced into

social sciences such as psychology and cognitive sciences as well.

A large number of research projects have been carried out on the factors influencing drug addiction in many fields of study such as medicine, psychology, and psychiatry.

In most of these research projects, specific statistical procedures have been used to analyze data and the usual way of hypothesis testing has come to be statistical procedures.

This is where data mining comes in since numerous variables can be investigated, through which meaningful, recurrent, and implicit patterns can be extracted without the need for formulating a hypothesis. The results of data mining studies can be applied as the knowledge source of backup systems to prevent addiction or make decision treatment.

2. Literature review

Evidence coming from the available literature suggests that data mining has different applications in different fields of study such as medical diagnosis and treatment options. Authors of [7] applied associative rules to youth drug addiction databases and indicated that there is a relationship between parents' behaviors, abnormal behaviors of the youth, peer group addiction, and taking illegal drugs.

Also, [8] aimed to analyze data and quantify the extent to which personal, environmental, and familial factors could predict cannabis use among the youth by using the ZINB model, decision tree, and association rules techniques. Also, [9] used a combination of k-medoids and hierarchical clustering procedures and launched genetic analyses in order to identify genes responsible for triggering addiction-related risks.

The present study aimed to use available data on drug addicts in order to reveal meaningful relationships beneficial to decision making regarding prevention, addiction diagnosis and treatment, and educating families. The ultimate goal was to use the data to ameliorate the current situation.

Results of the study can help educational system authorities and be included in preventive programs so as to raise the consciousness of the youth and promote well-being.

Moreover, findings of the study are useful for other domestic centers such as the Welfare Organization, Ministry of Health and Medical Education, and other affiliated organizations.

3. Data mining

Data mining is simply defined as extracting knowledge implicitly exists in a given database [10]. That is, data mining refers to the process of studying and analyzing a huge amount of data in order to extract underlying, meaningful patterns, relationships, and laws. In recent years, increasing importance has been given to data mining since databases are inoperable unless turned into knowledge. New patterns can be investigated for descriptive and predictive purposes. The former aims to unveil patterns in a way that is understandable and interpretable. On the other hand, the latter focuses on identifying variables or characteristics of a database and is used to predict the amount or future behavior of some variables [11].

Exploring and generating knowledge is a multi-staged process; its quality is affected by the accurate implementation of every single stage, which is as follows: 1) understanding a given problem and data, 2) collecting, analyzing and preprocessing relevant data, 3) selecting a method from data mining functionalities, 4) interpreting and evaluating the results in order to validate them and calculate accuracy of the model, and 5) utilizing the unveiled knowledge. Following some data mining techniques exploited in this study are surveyed.

3.1. Data mining techniques: Association rules

Different data mining techniques are employed to examine databases and result in various kinds of discovered knowledge. One of the important data mining techniques is the association rule techniques, which are used to extract correlations, recurrent patterns, causal relationships, and associations among a colossal number of items in a database, and then, the results form the basis of decision-making [12].

Association rules are simple, clear, intuitive, intelligible, and practical tools which expose the relationships among various variables in a descriptive way and aim to mine data so as to identify patterns without having background information on a given reality [7]. Perhaps, the main weakness of association rules is the number of produced rules, which require operators to do a lot of filtering. Also, it should be noted that extracting an association rule does not necessarily mean that it is efficient [7].

Finding attractive association rules helps decision-making process [13]. The attractiveness of a rule can be determined objectively or subjectively. Support and confidence coefficients are the

criteria based on which attractiveness of a rule can be assessed. The coefficients show efficiency and determination of the generated rules, respectively [14]. Applying the coefficient may result in a large number of rules from which some operators may find uninteresting. Therefore, support-confidence framework can be complemented by the lift correlation so as to reduce the number of produced rules and find more meaningful ones. In the case of $A \Rightarrow B$, if the lift is lower than 1, then the occurrence of A is negatively correlated with the occurrence of B. That is, occurrence of one result in the absence of another, from which it can be concluded that a negative correlation renders the association rule meaningless though it may have a high support coefficient. If the outcome of the association rule lift is higher than 1, then A and B are positively correlated, meaning that the occurrence of one proceeds the other one. If the outcome of the association rule is 1, then, A and B are independent meaning that there is no correlation between them [14].

Producing association rules is usually divided into two steps [12]:

- Applying the minimum support in order to find all frequent items in a database.

- Applying the minimum confidence to the frequent item sets to generate rules.

4. Introducing the proposed system

A successful treatment of drug addicts is very difficult and challenging. Hence, in order to save costs and time, it is preferably required to identify factors affecting drug addiction in an attempt to prevent addiction, something which can be fulfilled through data mining. Thus, the present study focused on discovering the patterns existing in the data related to people referring to rehabilitation clinics to undergo treatment.

4.1. Participants of the study (Dataset)

Target population of the study was all the individuals referring to rehab centers in Yazd, Iran, which included 471 patients. Of the participants, 86.2% were male and 13.8% were female. The participants were selected from an exclusively female rehab center, a prison clinic, Yazd Narcotics Anonymous Society, and other clinics across Yazd, Iran. Table 1 shows the dataset schema collected in this study.

Table 1. Conceptual schema of dataset.

No.	Items	Data category
1	sex, birth year, address, number of children, income, education, residence, marriage status, current job	Personal Information
2	number of family members, parents job, parents education, family history of drug addiction	Family Information
3	first substance, starting age of use and addiction, job, Marital Status, bidder, location, feeling, Risk factors, Parental reactions	First Use Information
4	Treatment count, Treatment type, Treatment Reason, Recurrence reason	Treatment Information
5	crime, number of arrests, arrest time	Crime Information
6	Physical Disease · Mental Disease	Disease Information
7	used substance, time of consumption	Drug Use Information

4.2. Understanding and preparing the data

In order to get familiar with the subjects studied, an interview was conducted with experts in the rehab center. Then, histograms and distribution diagrams were used to display the distribution of variables.

After the initial data understanding, the data was rounded up and the number of variables, their amounts, and columns, as the case may be, were reduced. Not only is the quality of results not damaged by this, but it also improves the results of data mining. Columns reduction is performed by human analysts or a number of automated processes, based on an understanding of the domain of use and purposes of the exploration.

After the preparatory stage, the association rules model was used to expose relationship existing in the data.

4.3. Discovering relationships

In order to discover association rules hidden in the data, a priori model was employed. However, due to the multiplicity of features and to avoid generating meaningless rules, specific features were nominated. That is, before modeling, useful features were selected and used as the input and then, support and confidence coefficients were extracted. In order to make sure of the validity of the generated rules, the lift correlation was applied, too. Table 2 shows the most valid rules

generated after applying rule pruning. Obviously, some of the generated rules can evidence the validity of the relationships that experts discover through experimental procedures. On the other

hand, some are more speculative and can be employed to take preventive and rehabilitating measures against drug addiction.

Table 2. Some generated association rules with interestingness measures.

Number	Association rules	Support	Confidence	Lift
1	If the person is young and unmarried and has started substance use since he was a teenager, he consumes LSD.	5.52	57.69	3.19
2	If the person feels hatred when he uses substance for the first time, he gives up the addiction.	1.69	25	2.35
3	If one of the members of a family is addicted, the woman gets addicted at home if she is suggested to use substance.	7	84.68	4.41
4	If a depressed young person uses different kinds of drugs, he'll commit drug sale crime.	5.14	56.66	4.23
5	If a low-income depressed person uses different kinds of drugs, he'll commit the robbery.	5.52	57.69	5.54
6	If a low-income middle aged man , is not resident in Yazd, he'll commit drug traffic crime.	5.3	36	3.94
7	If a middle-aged woman is a housewife, her addiction will return due to physical, mental or family problems.	5.94	39.28	2.66

As an example, according to rule 1, 57.69% of single young people experiencing drug early in their adolescence are using LSD. In other words, the likelihood of LSD abuse is 18%. Being celibate, adolescent, and an early drug abuse raises the amount to 57.69% meaning that those who are single and abuse drug early in their adolescence are three times more likely to turn to LSD abuse. According to rule 2, a feeling of hatred following the first drug abuse raises the likelihood of rehabilitation up to 2.35 times. Rule3 indicates that for women a family history of drug abuse results in home drug abuse. Rule 4 suggests that young addicts who suffer from depression and simultaneously take different narcotics such as LSD and Heroine are approximately four times more likely to commit crimes.

5. Conclusion

The present work aimed at exploring drug addiction-related databases from different rehab centers in order to find underlying patterns using association rules mining. Some of the extracted rules provided insights into the following issues: the role of individuals' feelings after the first drug abuse in the success of following treatments, characteristics of LSD abusers, main reason of addiction among different age groups, and the influence of family history on individuals' addiction. Findings of the study can help different organizations in making decisions about addiction.

References

- [1] Goreishizade, M. & Torabi, K. (2002). A study of co morbidity of substance abuse and psychiatric disorders in self-referred addicts to Tabriz welfare center. *Medical journal of Tabriz University of medical sciences*, vol. 55, pp. 49-53.
- [2] Pourmovahed, Z., Yassini Ardakani, M., Ahmadih, M., Deghani, K., & Kalani, Z. (2010). Evaluation of the Knowledge of Rural High School Students in Yazd About Drugs. *SSU_Journals*, vol. 18, no. 3, pp. 179-183.
- [3] Dastjerdi, G., Ebrahimi Dehshiri, V., Kholasezade, G. & Ehsani, F. (2010). Effectiveness of Methadone in Reduction of High Risk Behaviors in Clients of MMT Center. *SSU_Journals*, vol. 18, no. 3, pp. 215-219.
- [4] Padhy, N., Mishra, P. & Panigrahi, R. (2012). The Survey of Data Mining Applications And Feature Scope. *International Journal of Computer Science, Engineering and Information Technology (IJCSEIT)*, vol. 2, no. 3, pp. 43-58.
- [5] Diederich, J., Al-Ajmi, A. & Yellowlees, P. (2007). *EX-ray: Data mining and mental health*. *Applied Soft Computing*, vol. 7, no. 3, pp. 923-928.
- [6] Liao, S. H., Chu, P. H. & Hsiao, P. Y. (2012). Data mining techniques and applications –A decade review from 2000 to 2011. *Expert Systems with Applications*, vol. 39, no. 12, pp. 11303-11311.
- [7] Garcia, E. G., Blasco, B. C., López, R. J. & Pol, A. P. (2010). Study of the factors associated with substance use in adolescence using Association Rules. *Adiciones*, vol. 22, no. 4, pp. 293-299.
- [8] Gervilla, E., Cajal, B. & Palmer, A. (2011). Quantification of the influence of friends and antisocial behaviour in adolescent consumption of cannabis using

the ZINB model and data mining. *Addictive behaviors*, vol. 36, no. 4, pp. 368-374.

[9] Sun, J., Bi, J., Chan, G., Oslin, D., Farrer, L., Gelernter, J. & Kranzler, H. R. (2012). Improved methods to identify stable, highly heritable subtypes of opioid use and related behaviors. *Addictive Behaviors*, vol. 37, no. 10, pp. 1138-1144.

[10] Yun, L. & Xiang-sheng, L. (2010). The data mining and knowledge discovery in biomedicine. 5th International Conference on Computer Science and Education (ICCSE), IEEE, China, pp. 1050-1052.

[11] Chhieng, D. C. (2007). *Data Mining and Clinical Decision Support Systems*. Clinical Decision Support Systems, Springer New York, pp. 44-63.

[12] Divya, R. & Vinodkumar, S. (2012). Survey on AIS, Apriori and FP-Tree algorithms. *International Journal of Computer Science and Management Research*, vol. 1, no. 2, pp. 194-200.

[13] Khare, N., Adlakha, N. & Pardasani, K. R. (2010). An Algorithm for Mining Multidimensional Association Rules using Boolean Matrix. 2010 International Conference on Recent Trends in Information, Telecommunication and Computing (ITC), Kochi, Kerala, India, pp. 95-99.

[14] Han, J., Kamber, M. & Pei, J. (2006). *Data mining: concepts and techniques*. Morgan Kaufmann Publishers.

کشف روابط در داده‌های افراد معتاد به مواد مخدر با داده کاوی

فرزانه زاهدی^۱ و محمدرضا زارع میرک‌آباد^۲

^۱دانشکده فنی مهندسی، دانشگاه آزاد اسلامی، پردیس تحصیلات تکمیلی علوم و تحقیقات یزد، یزد، ایران.

^۲دانشکده مهندسی برق و کامپیوتر، گروه مهندسی کامپیوتر، دانشگاه یزد، یزد، ایران.

ارسال ۲۰۱۴/۰۱/۱۸؛ پذیرش ۲۰۱۴/۰۲/۱۵

چکیده:

اعتیاد به مواد مخدر یکی از مهم‌ترین مشکلات اجتماعی، اقتصادی و بهداشتی است که عوارض ناشی از آن تهدیدی جدی برای جامعه بشری محسوب شده و موجب رکود اجتماعی در زمینه‌های مختلف می‌شود؛ و تا زمانی که به علل گرایش بیمار توجه نشود، درمان جسمی و روانی فقط برای مدتی نتیجه‌بخش خواهد بود و فرد معتاد دوباره گرفتار مواد اعتیادآور می‌گردد. امروزه مراکز برای درمان اعتیاد وجود دارد که اطلاعات کاملی راجع به افراد معتاد در اختیار دارند، بنابراین با وجود حجم بالای داده، می‌توان با استفاده از ابزار داده‌کاوی به کشف دانش در داده‌ها پرداخت و از نتایج مدل‌های داده‌کاوی به عنوان پایگاه دانش سیستم پشتیبان تصمیم‌گیری در زمینه پیشگیری و درمان اعتیاد، استفاده نمود. جامعه آماری مورد استفاده در این پژوهش، مراجعین مراکز درمان اعتیاد شهر یزد می‌باشند. نمونه موردنظر شامل ۴۷۱ مراجعه‌کننده (%۸۴/۲ مرد و %۱۳/۸ زن) است. هدف از این مطالعه، استخراج قوانین مفید از داده‌های جمع‌آوری شده است که از طریق مدل Apriori صورت گرفته و روابطی بین ویژگی‌های فردی و مصرف شیشه، ویژگی‌های فردی و مواد مصرفی با انواع جرم‌ها و سابقه اعتیاد در خانواده با اعتیاد فرد کشف شده است. نتایج این مطالعه می‌تواند در مراکز درمان اعتیاد و سازمان‌های مرتبط استفاده شود.

کلمات کلیدی: اعتیاد، مواد مخدر، داده‌کاوی، قوانین انجمنی، کشف روابط.