A framework for mining the university student behaviour based on social network

To cite this article: M K M Nasution et al 2019 J. Phys.: Conf. Ser. 1235 012017

View the article online for updates and enhancements.
A framework for mining the university student behaviour based on social network

M K M Nasution¹, M A Muchtar¹, T H Nasution².
¹Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Sumatera Utara, Padang Bulan 20155 USU, Medan, Indonesia
²Fakultas Teknik, Universitas Sumatera Utara, Padang Bulan 20155 USU, Medan, Indonesia
E-mail: mahyuddin@usu.ac.id

Abstract. The intersection principle has opened the idea of developing multiple frameworks for mining the meaning any information, including those from the database. The concept of intersection is the foundation to build a social network or a community, so community behaviour can be meant from the available database including information on academics. Meaningful of any data can also be used to explore their data for the related behaviour. This paper explores some of the meanings formulated in the formulation or rules. The concept used to build meaning frameworks and to explore the behaviour of specific social communities, i.e. a framework based on Bayes theorem.

1. Introduction
Every nation needs human resources that have the competence, so that the nation can compete with other nations well [1]. General description of competitiveness can be seen from the existence of universities [2], while the potential of universities can be seen from the behavior of academics including students [3, 4].

In universities there is an academic information system (AIS) that contains a variety of information about the activities of academics [5]. Data stored in the database only stored and used as necessary, without any improvement in the function to set strategies in improving the quality of human resources [6, 7]. However, the data mining based on social networks of the AIS needs to be done to give appropriate meaning. Therefore, this paper intends to formulate some of order meaning to academic data.

2. A review and problem definition
Social network of: academic social actors, or students, can be built on the intersection concept or \( \cap \) which means there are the same things that function as similarities between social actors [8, 9]. Thus, social networks will group social actors with similar behavior into adjacent structures [10], so that if it is declared in reverse, the data coming from the same source will also behave the same [11, 12].

To obtain common ground, several approaches can be developed, such as conditional probabilities, or involving Bayes theorem [13]. This theorem is stated to classify issues that need to be predicted as an interpretation, by which the behavior of something can be recognized
The conditional probability $P(a|D)$ used to declare a social actor event $a \in A$ may occur against another attributes $D$ in the database,

$$P(a|D) = \frac{P(a \cap D)}{P(D)},$$  \hspace{1cm} (1)

with which $P(a|D)$ is the probability $a \in A$ occurs if given $D$ [15]. $A$ is a set of social actors [16]. While,

$$P(D|a) = \frac{P(a \cap D)}{P(a)},$$  \hspace{1cm} (2)

with which $P(D|a)$ is the probability that $D$ occurs if given $a \in A$ is observed [15]. Therefore,

$$P(a \cap D) = P(a)P(D|a).$$  \hspace{1cm} (3)

This means that the conditional probability usage approach is generally enabled to forensic the interconnected data [17]. In this case, the social network mining approach is to forensic the available data to illustrate the behavior of entities related to data, i.e. social actors [18]. In other words, this may reveal matters related to the data, including explaining from where the origin of the data source be the original behavior [19]. The origin behavior is as the cause that sought in the forensic which involves social networks.

3. An approach

The disclosure of academic social actor behavior is based on recent developments on the data mining related to social networks, which is an overlap principle that allows equality to be expressed [20]. In this case, the premise of the Bayes Theorem reveals an event occurring with the origin degree, and then new information about the degree of trust is formed as well. Two degrees are represented by the prior probability $P(a|D)$ like in Eq. (1) and the posterior probability $P(D|a)$ like in Eq. (2) and by substituting Eq. (3) so that they are coupled by [21, 22]

$$P(a|D) = \frac{P(a)P(D|a)}{P(D)}.$$  \hspace{1cm} (4)
Bayes’ decision rules reveal that based on posterior probabilities it is possible to assign any attribute in the database to a class with a probability [23].

To develop the meaning of the data available in the database, a framework is designed to involve data mining based on social networks. To help this design, a table that can determine position parameters and the giving of meaning based on probabilities, see Fig. 1(a): a number of students $t = x_1 + x_2 + x_3$ while a probability of $y_1$ is $P(y_1) = x_1/t$ or $P(y_1) \times 100\%$, etc., where $x_i$ are a number of students in a study program [24]. For example, involves attributes such as the identity of the student, study program, and entry year to realize the number of students on the study program for a particular year, the percentage of students per study program can be calculated against the total number of students within the university [25]. In other cases, this can also be expressed as a conditional probability $P(a|D)$ for $D = \{\text{entry year, study program}\}$ and $a_i \in A, i = 1, \ldots, n$ are students as social actors. This design can then be expanded by adding columns to the available table, but different data attributes, see Fig. 1(b), where different selections per study programs are compared. Based on the attributes to built communities that connect each student with other students based on attributes inherent such as the entry year, the study program, the senior high school and so on. Based on the scenario it is possible to express the meaning related to the improvement of human resource development, namely the improvement of the quality of education coming from human resource development and the increasing competitiveness of learners after accepted in university [26].

4. The framework design

In AIS there usually is attributes from the same entities, i.e. about the candidates or the students (those who have been accepted in university): either the senior high school ($shs$) or the student of university ($sou$). The attributes about $shs$ are identity ($id_{shs}$), name ($n_{shs}$), accreditation ($ac_{shs}$), address ($ad_{shs}$), and category ($ct_{shs}$), while the attributes for $sou$ are identity ($id_{sou}$), name ($n_{sou}$), and level ($l_{sp}$) about study program ($sp$), and identity ($id_{sp}$), entry year ($ey_{sou}$), grade point ($gp_{sou}$), selection ($s_{sou}$), and single tuition of students ($st_{sou}$), see Fig. 2. Whereas for the operation between attributes as development with Bayes theory involved is shown in Fig. 3, with the following attributes: number of students ($n(sou)$), $l_{sp}$, where the information of student entity is $id_{shs}$, $id_{sou}$, $ey_{sou}$, $st_{sou}$, program’s identity ($id_{sp}$), $gp_{sou}$, and $s_{sou}$ [27].

Data mining to derive meaning from a database is based on attributes that may produce meaning. For example, the entry opportunities for each study program involve the attributes like $sp$ and $ey_{sou}$, and the data mined is $n(sou)/sp/ey_{sou}$. Sometimes these opportunities can be synchronized with the passing grade of a study program, which serves to give meaning to the level of human resource capability that goes into each study program. Passing grade is derived.
Table 1. Level of the single tuition and its selection

| Level of \(st_{sou}\) | %     | Selections                        |
|------------------------|-------|----------------------------------|
| High level             | 20%   | invitation, test, independent, international |
| Medium level           | 30%   | invitation, test                  |
| Low level              | 30%   | invitation, test                  |
| Lowest level           | 20%   | invitation, test                  |

from the average score for a number of candidates accepted in a study program [1].

Increased human resource development can be traced from the compliance of each educational unit to the prevailing provisions or following existing standards. Increase in the form of predictions of student units, i.e. the classification of each student into a particular group according to the attribute in \(D\) gives different probabilities. Student search based on \(st_{sou}\) gives different meaning. Assessment of honesty of prospective students is expressed through \(st_{sou}\) and origin school \(shs\). \(st_{sou}\) consists of the lowest rank to the highest, see Table 1. The number of students with the lowest \(st_{sou}\), but comes from one school classified high cost, will reveal that the candidates have manipulated their data by themself. The number of student with the lowest \(st_{sou}\) under the provisions should only be about 20% of the maximum number of students in a university for each entry year. This is done to maintain the balance, so the number of students with the highest \(st_{sou}\) is 20% minimum. So the student opportunity to pay the lowest \(st_{sou}\) is \(P(a|st_{sou} < 20\%) \leq 0.20\) while the student probability pay the highest \(st_{sou}\) is \(P(a|st_{sou} > 20\%) \geq 0.20\).

In addition, either low or high quality of education can be expressed from the comparison of passing grade and average of \(gp_{sou}\) from one generation of alumni to the next generation of alumni. Schools \(shs\) as a source of candidates being educated in a university will be seen to have quality if the number of alumni of the school many accepted in universities and then they able to complete the lecture on time. Therefore, for the set of attributes \(D = \{lsp, id_{shs}, id_{sou}, gysou, st_{sou}, id_{sp}, gp_{sou}, s_{sou}\}\) we will be obtain conditional probabilities \(P(a|d_i \in D)\) with which \(d_i, i = 1,\ldots,n\) represent each attributes exist. Of course, to measure the effectiveness of \(s_{sou}\) required some of cores from data or the summary of items. It is assumed that there are 4 (four) selections to goes into the university: invitation \((ju)\), test \((jt)\), independent \((jm)\), and international \((ji)\). They have quotas 30%, 30%, 20%, and 20%, respectively, for example. Thus the chances of a candidate entering through independent path.
$jm$ is $P(a|jm) = 0.2$. So the probability the candidate has entered with via another path is $P(a\neg jm) = 1 - P(a|jm) = 0.8$. Suppose it is known that in total the students entering the independent path pay highest $st_{sou}$, so its probability is $P(st_{sou} = 100\%|jm) = 0.25$, whereas the non-independent path pay the highest $st_{sou}$ at least 20% or $P(st_{sou} \geq 20\%|\neg jm) = 0.20$. Based Eq. (4) we obtain

$$P(jm|st_{sou} \geq 20\%) = \frac{P(st_{sou}=100\%|jm)P(a|jm)}{P(st_{sou}=100\%|jm)P(a|jm)+P(st_{sou} \geq 20\%|\neg jm)P(a|\neg jm)} \tag{5}$$

$$= \frac{(0.25)(0.20)}{(0.25)(0.20)+(0.20)(0.80)} = 0.21$$

Normally it is generally known that students who enter a university with different $s_{sou}$ must have $gp_{sou} \geq 2.75$ is 32%. Thus we obtain $P(gp_{sou} \geq 2.75|jm) = 0.32$, while $P(gp_{sou} \geq 2.75|\neg jm) = 0.32$. Based Eq. (4) we obtain

$$P(jm|gp_{sou} \geq 2.75) = \frac{P(gp_{sou} \geq 2.75|jm)P(a|jm)}{P(gp_{sou} \geq 2.75|jm)P(a|jm)+P(gp_{sou} \geq 2.75|\neg jm)P(a|\neg jm)} \tag{6}$$

$$= \frac{(0.32)(0.20)}{(0.32)(0.20)+(0.32)(0.80)} = 0.20$$

Eq. (5) and Eq. (6) can be used as the rules to measure the effectiveness of the selection $s_{sou}$ used to filter candidates into university. This is done by expressing the database for the actual percentage of student who enter by paying the highest $st_{sou}$, and then the percentage of alumni with $ip_{sou} \geq 2.75$ when graduated from university. By using the concept based on Eq. (5) or Eq. (6) we can develop the rules to mining meaning. This is the concept for developing a framework for mining the university student behaviour based on social network where the student community is formed from the database in AIS.

5. Conclusion
One step to dig something is to get a framework for something. By involving the concept of Bayes theorem which involves the principle of intersection, a framework is built to mine the meaning of the database. Next work we will use this concept to build meaningful formulations involving relational databases rules.

Acknowledgement
We are grateful for Universitas Sumatera Utara, Medan, Indonesia to support research funding with entitle "PENAMBAHAN PERILAKU MAHASISWA BERBASIS JARINGAN SOSIAL (Mining of the student behaviour based on the social network)" at 2016.

References
[1] Nasution M K M 2016 Menggali potensi ujian nasional Harian Waspada, Opini.
[2] Nasution M K M 2018 No research without publication: early mining Journal of Physics: Conference Series 978(1)
[3] Nasution M K M, Elveny M, Syah R, Noah S A 2015 Behavior of the resources in the growth of social network Proceedings - 5th International Conference on Electrical Engineering and Informatics: Bridging the Knowledge between Academic, Industry, and Community, ICEEI 2015.
[4] Nasution M K M 2018 Indonesia knowledge dissemination: a snapshot IOP Conf. Series: Journal of Physics: Conf. Series 978.
[5] Indrayani E 2013 Management of academic information system (AIS) at higher education in the city of Bandung Procedia - Social and Behavioural Sciences 103.
[6] Nurjanah S, Santoso H B, Hasibuan Z A 2017 ICT adoption for education: A proposed framework Second International Conference on Informatics and Computing (ICIC).
[7] Nasution M K M, Nuradi T E, Syah R 2017 SumutSiana: A framework for applying ICT to preserve the cultural heritage of Sumatera Utara Indonesia *Journal of Telecommunication, Electronic and Computer Engineering* 9(2-4).

[8] Nasution M K M, Noah S A 2011 Extraction of academic social network from online database 2011 *International Conference on Semantic Technology and Information Retrieval, STAIR 2011*.

[9] Mahyuddin K M N, Sitompul O S, Nasution S, Ambarita H 2017 New similarity *IOP Conference Series: Materials Science and Engineering* 180(1).

[10] Nasution M K M, Syah R, Elveny M 2017 Studies on behaviour of information to extract the meaning behind the behaviour *Journal of Physics: Conference Series* 801(1).

[11] Wahyudin A, Hasibuan Z A 2015 Research classification in strategic information system planning development: A critical review *International Conference on Science in Information Technology (ICSTTech)*.

[12] Kasamsap K 2018 Multifaceted applications of data mining, business intelligence, and knowledge management *Intelligent Systems: Concepts, Methodologies, Tools, and Applications*.

[13] Nasution M K M, Noah S A 2011 Probabilistic generative model of social network based on Web features *Interior*.

[14] Purbolaksono M D, Widiastuti K C, Mubarok M S, Adiwijaya, Ma'ruf F A 2018 Implementation of mutual information and bayes theorem for classification microarray data *J. Phys.: Conf. Ser.* 971.

[15] Akbar M D, Mizoguchi Y, Adiwijaya 2018 Class dependency of fuzzy relational database using relational calculus and conditional probability *J. Phys.: Conf. Ser.* 971.

[16] Nasution M K M 2014 New method for extracting keyword for the social actor *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8397 LNAI(PART 1).

[17] Kasamsap K 2018 Multifaceted applications of data mining, business intelligence, and knowledge management *Intelligent Systems: Concepts, Methodologies, Tools, and Applications*.

[18] Hasibuan M S, Nugroho L E, Santosa I P 2017 Learning style model detection based on prior knowledge in e-learning system *Journal of Accounting Research* 21(1).

[19] Abdou S, Scordilis M S 2004 Beam search pruning in speech recognition using a posterior probability-based confidence measure *Speech Communication* 42(3-4).

[20] Lubis I, Nasution M K M, Maulina M 2018 Basic framework of urban design based on natural resources *IOP Conf. Ser.: Earth and Environmental Science* 126.

[21] Mahyuddin 1995 Pengantar Sistem Pengelolaan Basis Data - DBMS (Database Management System) USU Press, Medan.