ASI-MGK is an applied science tool available to researchers in all disciplines

Bakys Bruno Ralahady, André Totohasina

ASI-MGK IS AN APPLIED SCIENCE TOOL AVAILABLE TO RESEARCHERS IN ALL DISCIPLINES

RALAHADY Bruno Bakys[1]*, **TOTOHASINA André**[1]

[1]*Ecole Normale Supérieure pour l'Enseignement Technique, Université d'Antsiranana-Madagascar, BP O Antsiranana 201 - Madagascar.*

Keywords : Data analysis, ASI, ASI-MGK, Applied sciences, Confidence threshold.

Introduction

On the occasion of work in didactics carried out on the teaching of data processing in the Malagasy secondary and high schools, we used several multi-factorial experimental methods. To analyze the observed data, we used a non-symmetrical method of data analysis, called implicative statistical analysis (ISA), based on the measure of the intensity of implication. Statistical implicative analysis (ISA) is a non-symmetrical data analysis method designed by Régis Gras based on Gras's Intensity of implication measure [4]. For a large volume of data, Agrawal [1] and his team develop Apriori-type algorithms based on support and confidence deemed even less selective, less relevantand unrecognized the negative association rules. Sylvie Guillaume [5], in his thesis, proposes another more selective implicative quality measure M_{GK}. Totohasina and his teams continued this work, [8] defined these different mathematical properties justifying its relevance and developed its new non-subjective significance threshold. Ralahady, in his thesis [6,7], develops ASI-MGK, an implicative analysis tool based on the support, the M_{GK} and the confidence threshold.

Database

In a transactional database (quantifying the students' knowledge of computer science in their daily school environment) of this didactic research where the different fields correspond to the answers to a questionnaire counting 71 transactions and 50 answers. On the whole we have a Boolean matrix of dimensions 71 x 50.
We want to extract the valid association rules, of type A ⟹B "if a student understands a concept A, then it is very likely that he also understands a concept B", the most relevant.

College students during computer class

Methods

Using the ASI-MGK we performed a successive generation by modifying the minimum support and the confidence threshold.

Quality measure used

Guillaume Kentchaff Measure (M_{GK})

Let X and Y two patterns of a data mining context [6,7,8].

$$M_{GK}(X\Rightarrow Y) = \begin{cases} \dfrac{P_{X'}(Y') - P(Y')}{1-P(Y')} & \text{If } P_{X'}(Y') \geq P(Y') \\ \dfrac{P_{X'}(Y') - P(Y')}{P(Y')} & \text{If } P_{X'}(Y') < P(Y') \end{cases}$$

$$M_{GK}\text{critic}(X\Rightarrow Y, \alpha) = \sqrt{\frac{n_{Y^*}(n-n_X)}{n_{^*X^*}(n-n_Y)}}\chi_\alpha \quad \text{With X favors Y}$$

If $P_{X'}(Y') < P(Y')$ we study the negative rule $\overline{X} \Rightarrow Y$.

Generation of M_{GK}-valid rules

1) If $M_{GK}(Y\Rightarrow X) < M_{GK}(X\Rightarrow Y)$, then we retain the rule X⟹Y, respect val. Critical. In case of a tie, we have the rule X⟹Y.

2) If $M_{GK}\text{critic}(X\Rightarrow Y, \alpha) < M_{GK}(X\Rightarrow Y)$ at threshold b, then [7] the rule X⟹Y is validated at the confidence level (1-α)100%.

Pre-processing

Transcription and coding of responses

Codes	Signification	Support
F	Female sex	0.57
M	Male sex	0.43
BVR	Talkative during computer class unplugged	0.60
PTB	Disruptive	0.25
ACT	Actif	0.52
PRT	Participatory	0.52

Boolean variable

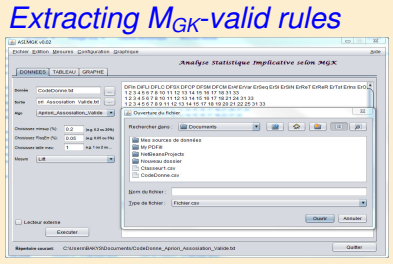
	Q1	Q2	Q3	Q4	⋯	Q16
Stu1	F	Old	Multimedia	17	⋯	Gifted
Stu2	M	Old	Games	16	⋯	Gifted
Stu3	M	New	Games	7	⋯	Weak
Stu4	F	Old	Office	10	⋯	Weak
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Stu75	M	Old	Internet	9	⋯	Weak

ASI-MGK processing

Launching the software

1) Importing the .CSV file;
2) Settings:
   - MinSup
   - Validity threshold
3) Algorithm execution

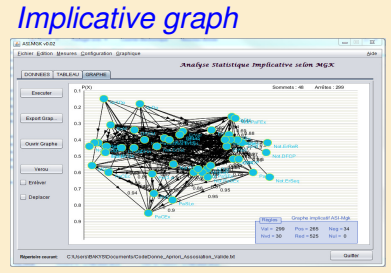
Extracting M_{GK}-valid rules



Boolean matrix

	F	M	BVR	PTB	ACT	PRT	NVE	OBJ	RES	PRS	RTR	PDL
A01	0	1	1	0	0	0	0	0	0	1	1	1
A02	0	1	0	0	1	1	0	1	1	0	0	0
A03	1	0	0	0	0	0	0	1	1	1	1	1
A04	1	0	0	0	0	0	0	1	1	1	1	1
A05	0	1	1	0	0	0	0	0	0	0	1	1
A06	0	1	1	0	1	1	0	1	1	0	0	0
A07	1	0	0	0	1	1	0	1	1	0	0	0
A08	1	0	1	0	0	0	0	1	1	1	1	1
A09	1	0	0	0	0	0	0	1	1	0	0	0
A10	1	0	0	0	1	1	0	1	1	0	0	0
A11	1	0	0	0	1	1	0	1	1	0	0	0
A12	0	1	1	0	1	1	0	1	1	0	0	1
A13	0	1	1	1	1	1	0	0	1	1	0	1
A14	0	1	1	0	1	1	0	1	1	0	0	0
A15	0	1	1	0	1	1	0	0	1	0	0	0
A16	0	1	0	0	1	1	0	1	0	0	0	0
A17	0	1	1	0	1	1	0	1	1	0	0	0
A18	0	1	1	0	1	1	0	1	0	0	0	0

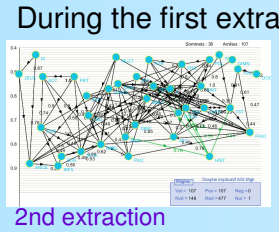
Implicative graph



Results

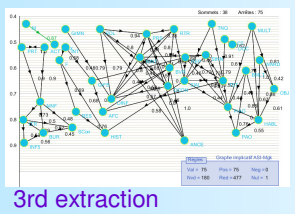
Implicative graph followed by improvement of sight by displacement.

During the first extraction: with minsup = 0.2 and minEur =0.05, we have 280 valid rules.

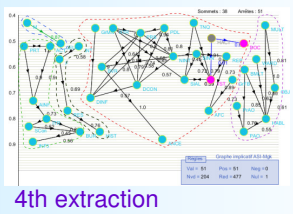


2nd extraction

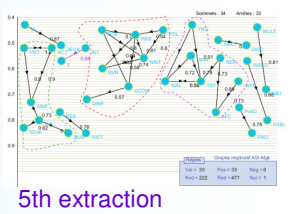
During the second extraction (*choice of minsup*): By setting minsup = 40%, we have 107 valid rules.



3rd extraction



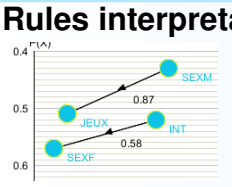
4th extraction



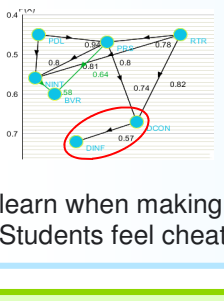
5th extraction

Third to fifth extractions (*choice of minEur):* With minEur = 5.0E-3 we have 79 valid rules, with minEur = 5.0E-4 we have 51 valid rules and with minEur = 5.0E-5, we have 33 valid rules.

Rules interpretation



Rule{SEXM} ⟹{JEUX} : 0.87.
with S*upp*({SEXM}) = 0.43 and S*upp*({JEUX}) = 0.51
Rule{INT} ⟹{SEXF} : 0.58
with S*upp*({INT}) = 0.52 and S*upp*({SEXF}) = 0.57
Here we find that most boys are generally interested in video games. On the other hand, almost all the girls in the class tend to get used to the Internet. We obtained a strong dependence on choice of IT technology or service with gender.



Rule{DCON} ⟹{DINF} : 0.57
with S*upp*({DCON}) = 0.67 and S*upp*({DINF}) = 0.72
Distracting implies hating computers.
Obviously with a percentage of 70%, the students hate computer education without a computer (unplugged) because they are distracted, this is not what they wanted to learn when making contact.
Students feel cheated and hopeless by the switch to unplugged mode.

Conclusions

During these processes, we observed a reduction in the number of valid rules. The less relevant rules disappear as the risk of error is reduced. In addition to the results found for this study, it is very interesting to note that ASI-MGK is an implicative graphing software belonging to the Applied Sciences and that allows the different researchers who use it to verify their hypotheses in terms of cause and effect with a precise risk of error.
It is very interesting to note that this ASI-MGK, as described above has already been used in several scientific works in different disciplines such as :

- Sociology: "*agronomic training, ecosystem and socio-economic progress*"[2].
- Agronomy: *"influence of a university education in agronomy on the feeling of traditions by undergraduate university students: case of the district of Mandritsara in Madagascar"*[3].

This confirms the interest of this technique of statistical analysis for the researchers.

Bibliographic references

THE 3RD INTERNATIONAL CONFERENCE ON APPLIED SCIENCES
ICAS 2022
Ton Duc Thang University, Ho Chi Minh City | 4-5 May, 2022

[1] AGRAWAL, R., IMIELINSKI, T. and SWAMI, A. (1993). Mining association rules between sets of items in large databases. *In* BUNEMAN, P. et JAJODIA, S., éditeurs : *Proc. of the ACMSIGMOD International Conference onManagement of Data*, volume 22, pages 207–216,Washington,U.S.A.

[2] *C.* RAZANATSOAVINA, B B RALAHADY, J. C. Laberche. and H. BEARINIAINA, (2021). *Formation agronomique, ecosysteme et progres socioconomique.* Journal of Sciences of Technologies and the Environment *RSTE4, Mahajanga, Madagascar.*

[3] *C.* RAZANATSOAVINA, B B RALAHADY and J. C. Laberche, (2019). Influence of a university education in agronomy on the feeling of traditions by undergraduate university students: case of the district of Mandritsara in Madagascar 10th International Colloquium on Implicative Statistical Analysis (ASI10) October 2-5 2019, Belford, France.

[4] GRAS, R. (2014). Genese et developpement de l'analyse statistique implicative : retrospective historique.*EducMatem Pesq São Paulo*, 16(3):645–661.

[5] GUILLAUME, S. (2000). *Traitement des données volumineuses. Mesures et algorithmes d'extraction des règles d'association et règles ordinales*. Thèse de doctorat, Université de Nantes, France.

[6] RALAHADY, B. B. and TOTOHASINA, A. (2019a). Asi-mgk : Implicative statistical analysis tool based on *m*GK. *IJCST*, 3(1).

[7] RALAHADY, B. B. and TOTOHASINA, A. (2019b). Experimental study of the valid rules according to the measure*m*GK. *IJCST*, 3(1).

[8] TOTOHASINA, A. (2008). *Contribution à l'étude des mesures de qualité des règles d'association : normalisation sous cinq contraintes et cas de MGK : propriété, base composite des règles d'association et extension en vue d'applications en statistique et en sciences physiques.* Thèse de doctorat, Université d'Antsiranana, Madagascar.