USING LOGIC TO DETERMINE KEY ITEMS IN MATH EDUCATION

Manuel Ojeda-Hernández, Francisco Pérez-Gámez, Ángel Mora Bonilla
and Domingo López-Rodríguez
Departamento de Matemática Aplicada - E.T.S.I.Informática – Universidad de Málaga, Spain

ABSTRACT

Due to the COVID-19 restrictions, high-schools in Spain are having both online and in-class lectures. As a result, the students can use not only the information provided by the teachers in class, but they can also use several other methods such as videos and online examples that allow the students to have materials from different places. In this paper, we analyse the Mathematics results of the first two terms in a secondary school from Andalusia. This analysis can help find the central units of the subject found, that is, giving enough background knowledge to keep up with the module. When these are found, the teachers can improve the learning and the results in the following years.

1. INTRODUCTION

e-Learning is an exciting approach to learning that we can use to improve the teaching and learning process quality. It has advantages such as allowing self-paced learning for each of the students, availability of the content at any time, and community-based support, given in the form of forums where the students can ask and solve their questions and their partners'.

A mixture of e-Learning and classical learning is Blended Learning (Dziuban et al., 2018). This term refers to every teaching style that combines traditional face-to-face lectures with an online platform where some content is uploaded, namely, theoretical material, quizzes, tasks, and other stuff related to the module.

Usually, nowadays, we could say we approach our teaching using Blended Learning. That is, we provide on-site courses with live instruction but with solid support from online systems.

In the Spanish Education System framework, the use of Learning Management Systems such as Moodle is widely extended (Almansa-Martinez, 2019) (Cabero-Almenara, 2019). Moodle (Modular Object-Oriented Dynamic Learning Environment) is, at the moment, the most popular system for e-Learning purposes.

The wide use of e-Learning platforms has brought about a massive amount of data that, with proper processing, may bring out meaningful information beneficial for teaching purposes. Anyone with minimal technical knowledge can use classical techniques like Machine Learning or Data Mining to extract knowledge from data.

In recent years, there have been many different approaches in the literature showing the advantages of using methods of Machine Learning or Data Mining to e-Learning (Mohamad, 2013) (Romero, 2008), (Viloria, 2019). For example, it is possible to develop systems that adapt or recommend modifications of particular contents according to different students' aspects defined by their behaviour (Ashraf, 2020) (De Maio, 2012) (Asil Oztekin, 2013) (Hooshyar, 2020). Other applications of this kind of techniques to e-Learning are the development of tools to detect and prevent academic dropouts (Burgos, 2018), (Chung, 2019), (Chui, 2020), (Gray, 2019) (Blundo, 2021), the detection of problematic aspects in evaluation tasks (García 2011), the representation of feelings or preferences of students (Carmona, 2007), (Zengin, 2011) or the measurement of the student experience during an e-course (Shukor, 2015), (Hew, 2020) among others.

Since one of the main goals is to provide an understandable analysis of datasets, we chose to use Formal Concept Analysis (FCA) tools. FCA is a solid mathematical framework to manage information based on logic, lattice theory and Galois connections. It defines two explicit representations of the tacit knowledge present in a dataset (called formal context). One of these representations is in the form of concepts, which are closed sets under a closure operator, entities characterised by the non-formalized relationships among the attributes or features in a dataset.

The other representation of the information is in the form of implications, which are exact association rules. The use of implications extracted from a dataset allows us to employ logic to reason with them. The implications obtained by this method capture all the non-trivial knowledge included in the dataset.

FCA provides the methods to find those representations and extract the concepts and implications that can be deduced from the dataset, introducing logic tools to reason and infer new knowledge. In this sense, FCA allows discovering knowledge in datasets analogously to what other techniques (e.g. in Machine Learning) do. Still, this logic-based approach is more suitable to provide explainable answers when dealing with real-world datasets. Comparing our system with other techniques used in machine learning and data mining, such as association rules, FCA extracts more knowledge. It gives the implications (exact association rules) and the concepts organised in a hierarchical structure, called the concept lattice. Moreover, our team is exploring the use of logic, specifically the Simplification Logic (see Mora 2012), to manipulate implications and build automated methods to reason.

In this work, we use FCA to retrieve the implications from a dataset with the exam marks obtained in the different Maths subject lessons, inspect them, and acquire valuable knowledge for the teacher.

Mainly, our research deals with searching for hidden patterns and relationships between different courses or lectures in pre-university degrees. Specifically, in this paper, we present some preliminary results obtained using FCA techniques to the data collected from the Maths subject in High School, taught following the Blended Learning scheme, due to the COVID-19 restrictions. Data collected in this course include mainly exam marks. Other researchers (Buldu, 2010) (Yahya, 2019) have addressed this general problem of finding links between different lessons. Still, the difference concerning our approach is that we use Formal Concept Analysis (FCA) (Ganter, 1999) (Ganter, 2019) as the tool to extract the knowledge from the data. Although FCA has been applied in the context of e-Learning, e.g., in (De Maio, 2012), to the best of our knowledge, it has never been used for this specific goal, which is studying the weight of each unit within the subject, thus helping the teacher improve the marking scheme for the following year, rather than giving result statistics.

In this way, Section 2 describes the subject details and how we have approached FCA's use to extract knowledge. The results obtained are explored in Section 3, and some conclusions and future works appear in the last section.

2. MATERIALS AND METHODS

In this paper, we analyse the marks obtained in mathematics in a class of the third and fourth year of compulsory education in Spain. The data used has been taken from a secondary school in Andalusia (Spain). The third year class has forty-seven students and the fourth year class has fifty-five students. We cannot give more information about the dataset due to the data protection regulations in Spain.

This year, due to COVID restrictions, the students have both in-class and online teaching, i.e., one day they have in-class lectures, and the following one, they have online classes. This way, there are weeks when they have three math classes and others where they have only two. In the days with online lectures, the students must do some different exercises to check that everything they were taught in class was understood. Some other online activities include watching videos that explain some theory in order to solve some exercises later on.

Although we do not have many students in these groups, we have analysed the qualifications obtained from different exams and courseworks made in the first and second term (see Table 1) to check if any implications would lead us to pass the term.

Table 1. Table of the exams analyzed

Third course	Fourth course
First term:	
Integers and fractions	Real numbers
Decimal numbers and scientific notation	Polynomials and algebraic functions
Polynomials and numerical sequences	Equations, inequations and systems I
Mark of the first term	Coursework I (solving eq. systems)
	Mark of the first term
Second term:	
Equations and systems of equations	Equations, inequations and systems II
Functions and graphs	Elemental functions and characteristics
Coursework I (Geogebra)	Similarity, applications and trigonometry
Mark of the second term	Coursework II (Geogebra)
	Mark of the second term

This way, the teachers can focus on those exams or exercises in the following years to get better results and a more balanced marking scheme. The students also have different activities where they get the skills and knowledge necessary to get a high level in maths this year and further years due to mathematics being a pyramidal study. The students need a good background if they want to improve their knowledge and add new information about the subject.

Even though the subject's teacher thinks that the work made so far has been satisfactory enough to get that background needed for further courses, analysing the data could potentially provide new (hidden) information. For instance, the teacher could determine the most critical unit that gives the students the mindset needed to understand the latest concepts and pass the terms.

In summary, the goal of the paper is to extract interesting knowledge from this data. Some questions that we want to answer are the following:

● Is there any exam or set of exams important enough to imply passing the two terms?
● Is there any connection between the exams, i.e., is there any implication telling that if a student passes an exam, they will pass any of the following ones?

The method chosen to analyse this dataset and answer these questions is Formal Concept Analysis (FCA). Since it is based on Logic and Lattice Theory, this technique can provide more interpretable answers to the questions above.

FCA is useful to capture and infer knowledge stored in binary data sets, usually referred to as formal context. Usually, this formal context is described as tables where the rows are called objects and the columns are called attributes. Also, the formal context can be given as $K = \langle G, M, I \rangle$ where for all $g \in X$ and $m \in M$ $\langle g, m \rangle \in I$ if and only if the object g has the attribute m .

To capture the information we use the following mappings:

$X^{\uparrow} = \{m \in M \mid \langle g, m \rangle \in I, for\ all\ g \in X \}$ i.e. given a subset $X \subseteq G$ the subset $X^{\uparrow}$ is the subset of all attributes shared by all the objects in X.

$Y^{\downarrow} = \{g \in G \mid \langle g, m \rangle \in I, for\ all\ m \in Y \}$ i.e. given a subset $Y \subseteq M$, $Y^{\downarrow}$ is the subset of all objects that have all the attributes in Y.

With these mappings, we can form the so-called formal concept: a pair of subsets (A,B) where A is the set of objects that share all the attributes in B whilst they do not share any other attribute.

We can define an order among formal concepts as $(A_1 , B_1) \leq (A_2 , B_2)$ if and only if $A_1 \subseteq A_2$ or (equivalently) if $B_2 \subseteq B_1$. This way, the set of all the formal concepts will have a structure (the concept lattice) to give important information about the formal context. This concept lattice has already been used previously to infer hidden knowledge in Maths education (Pérez-Gámez, 2020).

However, in this paper, we are going to use the implications, we call implication to any expression A →B where A and B are subsets of M. We say that an implication A →B is valid in a formal context if and only if $A^{\downarrow} \subseteq B^{\downarrow}$, that is if every object satisfying the attribute set A also satisfies the attribute set B..

Due to the length restrictions, we do not introduce an extensive summary of Formal Concept Analysis. See (Ganter, 1999) and (Ganter 2011) for more details about FCA.

We have used the R language to develop the analysis of the data using FCA and, specifically, the R package named **fcaR**[1]. This package allows extracting the knowledge inside a binary dataset using FCA.

Using **fcaR**, the user can easily extract the formal concepts, the whole concept lattice, relationships between attributes. In the following section, we show how we have used **fcaR** to extract the valid implications in the provided datasets.

3. RESULTS OF THE ANALYSIS

To answer the questions posed above, we have introduced in R the information provided by the High School. This data could be simply an excel file or a CSV file. Then, the data has been binarized; that is, all the passed items were put to 1, and the rest were put to 0. Therefore, we get a formal context having the students as objects, the exams as attributes, and a binary relation representing ``a student is related to an exam if and only if the student has passed the exam''.

The package **fcaR** allows us to use the latter defined formal context as such and, hence, compute its basis of implications, also called Duquenne-Guigues basis (Duquenne-Guigues, 1986). All the implications were computed with **fcaR** using the **find_implications** command. The following are all the implications obtained from the third course database.

Implication set with 13 implications.

- Rule 1: {Functions, Geogebra} → {2nd term}
- Rule 2: {Equations} → {2nd term}
- Rule 3: {Polynomials, Equations, 2nd term} → {1st term}
- Rule 4: {Decimals, 2nd term} → {1st term}
- Rule 5: {Decimals, Geogebra} → {1st term}
- Rule 6: {Decimals, Functions} → {1st term}
- Rule 7: {Decimals, 1st term, Equations, 2nd term} →{Integers, Functions}
- Rule 8: {Decimals, Polynomials} → {1st term}
- Rule 9: {Decimals, Polynomials, 1st term, Geogebra} → {2nd term}
- Rule 10: {Integers, 2nd term} → {1st term}
- Rule 11: {Integers, 1st term, Equations, 2nd term} → {Functions}
- Rule 12: {Integers, Equations} → {1st term}
- Rule 13: {Integers, Decimals} → {1st term}

Even though the package gives the implications on its own, some may be a little confusing. Moreover, the rules extracted have a high degree of redundancy. Thus, these implications have been simplified using logical simplification rules, which give a simpler read but contain the same amount of information.

3.1 Analysis of the Third Year Class

Among the total number of implications, we will only consider the ones with the first or the second term mark as a consequent. This is due to these implications being the most important to both teachers and students.

[1](https://CRAN.R-project.org/package=fcaR). See https://github.com/Malaga-FCA-group/fcaR for some vignettes explaining the use of this package.

1st term: Implication set with 8 implications.

- Rule 1: {Polynomials, Equations, 2nd term} → {1st term}
- Rule 2: {Decimals, 2nd term} → {1st term}
- Rule 3: {Decimals, Coursework} → {1st term}
- Rule 4: {Decimals, Functions} → {1st term}
- Rule 5: {Decimals, Polynomials} → {1st term}
- Rule 6: {Integers, 2nd term} → {1st term}
- Rule 7: {Integers, Polynomials} → {1st term}
- Rule 8: {Integers, Decimals} → {1st term}

This set of implications shows a balanced weight of each unit during the first term. This is by rules 5, 7 and 8. These rules show that a student passing any two exams during the term will pass the module. This is desirable from a teaching point of view because students pass two out of three exams. In addition, if they fail one, it is with a high mark; that is, students fail the exam but are not completely unaware of the unit contents. This good balance does not appear in the second term results.

2nd term: Implication set with 3 implications.

- Rule 1: {Functions, Coursework} → {2nd term}
- Rule 2: {Equations} → {2nd term}
- Rule 3: {Decimals, Polynomials, 1st term, Coursework} → {2nd term}

On the one hand, notice that rule 2 gives a privileged weight to the Equations unit. Every student who passes the Equations exam gets to pass the second term. This makes sense from a theoretical point of view because the first Equations unit is a milestone to every math student and understanding it makes every other unit easier. Still, the term's weight should be more balanced since this unit is not the most important in the whole course. For instance, every student that chooses the Mathematics module in higher years must have Functions as background knowledge, which is not ensured with this marking scheme.

On the other hand, rule 1 ensures that every student who has passed the Functions exam and the Coursework passes the second term. This is interesting since the coursework's contents are function graph representations in Geogebra. Therefore, a good knowledge of Functions ensures a satisfactory result in the whole term. Again, Functions is one of the most important units in the whole course, but notice that it is not strong enough to be independent of the other exams, particularly the Geogebra coursework.

This set of implications suggests that the first term's marking scheme is well balanced, but the one of the second term should change to give less importance to the Equations unit and try to balance it with the Functions one.

3.2 Analysis of the Fourth Year Class

As in the previous section, we will consider the simplified set of implications. From this set, the implications taken into account are again only the ones such that the consequent is either the first or the second term mark.

1st term: Implication set with 6 implications.

- Rule 1: {Functions} → {1st term}
- Rule 2 : {Equations II, 2nd term} → {1st term}
- Rule 3: {Equations II, Geogebra} → {1st term}
- Rule 4: {Eq. Project, Equations II} → {1st term}
- Rule 5: {Equations I} → {1st term}
- Rule 6: {Polynomials} → {1st term}

From this set of implications, we observe the following conclusions. From rule 1, we infer that if a student does not have enough knowledge to pass the first term, then they are going to fail the Functions unit because it uses information learnt in the first term such as Equations and Polynomials. Also, in rules 2 and 3, we see the same idea: if a student has not passed the first term, they will not pass the second term as they do not have enough knowledge to pass the second term. In rules 5 and 6, we find that Equations I and Polynomials are important enough on their own to assure passing the first term. It could be because if a student passes one of these exams, they are probably going to pass another of the exams from the first term as they need a good knowledge of real numbers to work with polynomials and equations. Also, deep knowledge of polynomials is needed to understand equations.

2nd term: Implication set with 10 implications.

- Rule 1: {Trigonometry, Geogebra}→ {2nd term}
- Rule 2: {Functions} →{2nd term}
- Rule 3: {Equations II, Geogebra} →{2nd term}
- Rule 4: {1st term, Equations II} → {2nd term}
- Rule 5: {Eq. Project, Equations II} → {2nd term}
- Rule 6: {Equations I, 1st term, Geogebra} → {2nd term}
- Rule 7: {Reals, 1st term, Trigonometry} →{2nd term}
- Rule 8: {Reals, Eq. Project} → {2nd term}
- Rule 9: {Reals, Equations I, 1st term}→ {2nd term}
- Rule 10: {Reals, Polynomials, 1st term}→ {2nd term}

In rule 2, again, we find that Functions is the central unit of the course; if a student passes the Functions exam, this means that they have enough knowledge to pass the two terms. In rule 4, we have that if a student understands the first term and Equations II, then they are going to pass the second term. In rule 6, we have again that if a student passes the first term, Equations I and the Geogebra coursework, they will pass the second term. In rules 4, 7, 9 and 10, we have that passing the first term along with other exams implies passing the second term. This could happen because if a student has passed the first term, they should have enough knowledge to afford the second term. In rule 1, we have that if a student passes Trigonometry and Geogebra, they are going to pass the second term. Hence, Trigonometry and Geogebra are important information within the second term. Lastly, in rule 3 again we find Geogebra along with Equations II showing the importance these units have in order to pass the second term.

4. CONCLUSIONS

We have presented our results of the analysis of the evaluations in two courses in a High School in Andalusia. We have analysed some exams and projects made during the first and second term.

In this work, we have used FCA tools to extract logical implications between the exams, coursework and the marks obtained in the first and second terms, finding some of them very interesting from the teachers' point of view. For instance, some exams are relevant enough to make a student pass both terms. The teachers can use this information to improve their lecturing in the following years so that students can get a bigger background knowledge and be better prepared for higher levels. This is not an analysis to be used midterm but from one course to another.

One of the main restrictions during the development of this study is having the data of half a course and not a full course. Therefore, we look forward to extending this study to one taking into account all the exams of the course. Besides, we plan to do a longitudinal study where we consider not just one year but also the following courses, even university studies. Of course, all the teachers and professors will have all this information available from that study to improve their lectures and increase the math level in high schools and universities.

Another direction to improve our study is taking into account not only the fact of passing or failing the exams but also if they get a letter mark such as A, B, C, D, E, F, FF, G or H. This way, a deeper analysis could be made because it would give much more information than a binary set of pass/fail data.

ACKNOWLEDGMENT

REFERENCES

Almansa-Martínez,et al. 2019. *Functionalities of moodle and edmodo in the middle and higher educations. In Revista de Comunicación de la SEECI, (Vol 50, 87-105).*

Asil Oztekin, et al, 2013. *.A machine learning-based usability evaluation method for e-Learning systems. In Decision Support Systems (Vol 56, pp 63-73)*

Ashraf, M. et al, 2020. An Intelligent Prediction System for Educational Data Mining Based on Ensemble and Filtering approaches. *In Procedia Computer Science (Vol. 167, pp. 1471–1483). Elsevier B.V.*

Blundo,C., et al, 2021. A time-driven FCA-based approach for identifying students' dropout in MOOCs, *International Journal of Intelligent systems,*

Buldu, A. and Üçgün, K, 2010. Data mining application on students' data. *In Procedia - Social and Behavioral Sciences (Vol. 2, pp. 5251–5259). Elsevier.*

Burgos, C. et al, 2018. Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout. *Computers and Electrical Engineering, 66, 541–556.*

Cabero-Almenara, et al, 2019. *Technical and Didactic Knowledge of the Moodle LMS in Higher Education. Beyond Functional Use. Journal of New Approaches in Educational Research,* (Vol 8, 25-33)

Carmona, C., et al, 2007. *Discovering Student Preferences in e-Learning.*

Chui, K. T. et al, 2020. Predicting at-risk university students in a virtual learning environment via a machine learning algorithm. *Computers in Human Behavior, 107, 105584.*

Chung, J. Y. and Lee, S., 2019. Dropout early warning systems for high school students using machine learning. *Children and Youth Services Review, 96, 346–353.*

De Maio, C. et al, 2012. RSS-based e-Learning recommendations exploiting fuzzy FCA for Knowledge Modeling. *Applied Soft Computing Journal, 12(1), 113–124.*

Dziuban, C. et al, 2018. Blended learning: the new normal and emerging technologies. *International Journal of Technology in Higher Education. 15(3)..*

Ganter, B. and Wille, R. (1999). Formal Concept Analysis: Mathematical Foundations. Springer.

Ganter, B. and Obiedkov, S. (2016). Conceptual Exploration. Conceptual Exploration, 1–315.

Ganter, B. and Obiedkov, S. 2016 *Conceptual Exploration Springer.*

García, E. et al, 2011. A collaborative educational association rule mining tool. *Internet and Higher Education, 14(2), 77– 88.*

Gray, C. C. and Perkins, D., 2019. Utilizing early engagement and machine learning to predict student outcomes. *Computers and Education, 131, 22–32.*

Hew, K. F. et al, 2020. What predicts student satisfaction with MOOCs: A gradient boosting trees supervised machine learning and sentiment analysis approach. *Computers and Education, 145, 103724.*

Hooshyar, D. et al. 2020. Open learner models in supporting self-regulated learning in higher education: A systematic literature review. *Computers and Education, 154, 103878.*

Mohamad, S. K. and Tasir, Z., 2013. Educational Data Mining: *A Review. Procedia - Social and Behavioral Sciences, 97, 320–324.*

Mora, A. et al, 2012. Closure via functional dependence simplification, *International. Journal of Computational Mathematics, 89,4, 510-526*

Pérez-Gámez, F et al, 2020. Using Formal Concept Analysis to Explore Hidden Knowledge in the Assessment of a Math Course. *International Conference e-Learning*.

Romero, C., et al, 2008. Data mining in course management systems: Moodle case study and tutorial. *Computers and Education, 51(1), 368–384.*

Shukor, N. A., et al., 2015. An Examination of Online Learning Effectiveness Using Data Mining. *Procedia - Social and Behavioral Sciences, 172, 555–562.*

Viloria, A. et al, 2019. Determinating student interactions in a virtual learning environment using data mining. *In Procedia Computer Science (Vol. 155, pp. 587–592). Elsevier B.V.*

Waheed, H. et al, 2020. Predicting academic performance of students from VLE big data using deep learning models. *Computers in Human Behavior, 104.*

Yahya, A. A. and Osman, A., 2019. Using Data Mining Techniques to Guide Academic Programs Design and Assessment. *In Procedia Computer Science (Vol. 163, pp. 472–481). Elsevier B.V.*

Zengin, K., Esgi et al, 2011. A sample study on applying data mining research techniques in educational science: Developing a more meaning of data. *In Procedia - Social and Behavioral Sciences (Vol. 15, pp. 4028–4032). Elsevier.*