Identifying Students' Mechanistic Explanations in Textual Responses to Science Questions with Association Rule Mining

Yu Guo
Northwestern University
Evanston, IL, USA
bguo@u.northwestern.edu

Wanli Xing
iSchool
University of Missori-Columbia
wxdg5@mail.missouri.edu

Hee-Sun Lee
The Concord Consortium
West Emeryville, California, USA
hlee@concord.org

***Abstract*—Reasoning about causal mechanisms is central to scientific inquiry. In science education, it is important for teachers and researchers to detect students' mechanistic explanations as evidence of their learning, especially related to causal mechanisms. In this paper, we introduce a semi-automated method that combines association rule mining with human rater's insight to characterize students' mechanistic explanations from their written responses to science questions. We show an example of applying this method to students' written responses to a question about climate change and compare mechanistic reasoning between high- and low-scoring student groups. Such analysis provides important insight into students' current knowledge structure and informs teachers and researchers about future design of instructional interventions.**

Keywords—association rule mining; learning analytics; mechanistic reasoning; science education; assessment

I. Introduction

Constructing explanations of real-world phenomena is an essential practice for scientists as well as an important ability for students to learn in science classrooms [1]. Several types of explanations are identified. Teleological explanations attribute natural phenomena to the design by a creator or the need to fulfill some purpose and thus are not acceptable as scientific explanations. Causal explanations, on the other hand, involve a sequence of reasoning where earlier events lead to later events. However, not all causal explanations are sufficiently precise. For example, some causal explanations may only answer the "what" question, without answering the "how" question, or the causal mechanism [2, 3]. Therefore, constructing explanations with multi-step causal mechanisms to account for natural phenomena is a central goal of science education.

Reasoning about causal mechanisms is central to scientific inquiry [4]. Mechanisms are "entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions" [5]. Mechanistic ideas are more "scientific" than occult or teleological ones, whether or not they are correct [3]. In this paper, we focus on students' mechanistic explanations—using entities, activities, and their relationships to account for a real-world phenomenon in their written responses to a scientific question.

Students' written responses are underutilized as data in teaching and learning. Even though large amounts of texts are generated from tests and day-to-day teaching and learning processes, they are often superficially evaluated or left unexamined. Students in science classes are typically asked to construct explanations to answer different types of questions. However, science teachers are usually too busy to provide thorough evaluations on students' written responses in part because they are collected at much faster speed than can be manually examined. Even if teachers perform manual analyses, they are prone to human errors due to fatigues and biases. In addition, although it is possible for teachers to focus on individual students' responses and conduct in-depth evaluations of the students' knowledge and reasoning, due to human cognitive limits, it is very difficult, if not impossible, to detect overall patterns from large amounts of students' responses.

In this paper, we introduce a semi-automated method in order to identify students' mechanistic explanations from their written responses to scientific questions by combining a simple data minding technique—association rule mining—with human raters' insight. This method can leverage the use of written texts, a data source easily accessible and widely available, to provide timely and frequent feedback related to students' understanding of the content knowledge. Such feedback can be useful to diagnose students' learning and formatively evaluate the effects of existing instructional practices. It can also inform teachers and educational researchers of what instructional remedies are necessary and appropriate.

Note that automated scoring is not the goal of this project. There exists a large amount of literature addressing mature methods on automated scoring and classification [6, 7, 8]. The purpose of this project is to gain insight into students' understanding of a complex natural system [9, 10, 11], especially their mechanistic explanations, as opposed to merely classifying students' responses. Although we used pre-existing categories that human coders created to assign high- and low-scoring student groups, human efforts are not always required in creating these categories. This task can be automated using other existing data mining methods.

II. Background

A. Mechanistic Explanations

Mechanistic explanations of real-world phenomena consist of entities, activities, and the relationships between them within the phenomena. Activities are "the things that entities do... and they constitute stages of mechanisms" [12]. To illustrate this definition, compare the two responses below to a question about how a decrease in sea ice levels can affect global temperatures.

Student A wrote: "The ice will have melted, letting more heat be absorbed into the ocean." In this response, there are two activities, or sub-mechanisms: ice melting and ocean absorbing heat. Although this student did not fully explain a causal chain between sea ice change and temperature change, he or she provided two important mechanistic accounts that contribute to the big picture. Student B explained that "less ice caps means more water, the water is darker." This explanation is less mechanistic because activities are missing from the explanation. It is not clear what happened to the entities "ice caps" and "water".

It is not hard to see from the above two examples that good markers for entities and activities are their part of speech. Entities are nouns, while activities are usually verbs. Therefore, the task of detecting mechanistic reasoning becomes detecting nouns, verbs, and the relationships between them.

B. *Association Rule Mining*

Association rule mining is a simple, popular, and effective data mining technique used to extract interesting correlations, frequent patterns, associations, or causal structures among sets of items in large datasets [13]. The resulting format is usually association rules expressed as X => Y. The meaning of such rule is intuitive: it shows how X and Y concurrently occur in the whole database. The probability of such rule is shown as confidence and support. Therefore, the importance of a particular rule can be reflected in the values on confidence and support. The confidence control and support is the main difference between association rule mining and standard natural language processing co-occurrence count. Association rule mining finds all frequent itemsets and then generates strong association rules from the frequent itemsets.

For example, Table I lists four students' responses to the climate change question. To use association rule mining for natural language processing, the first step is to tokenize natural sentences—to break sentences into items. Words with too high frequencies, such as "to" and "the", or those without substantial meanings are removed. Table I is, therefore, converted to Table II, which lists items and their frequencies.

When the algorithm is running, it detects, for example, "absorb" and "ocean" co-occur in 50% of all transactions, as they appeared in two out of four transactions. Therefore, {absorb, ocean} is a frequent itemset. The algorithm also finds that these two words co-occur with the word {ice}, thus, an association rule is generated:

$$\{\text{absorb, ocean}\} \Rightarrow \{\text{ice}\} \qquad (1)$$

This rule has a *support* of 0.5, as it appears in 2 out of 4 transactions (Responses 1 and 2). It means that when "absorb" and "ocean" are present in a transaction, there is a 50% chance that "ice" is also present.

$$\text{supp}(\{\text{absorb, ocean, ice}\}) = 2 / 4 = 0.5 \qquad (2)$$

TABLE I. EXAMPLE RESPONSES

Response	**Content**
1	More sun light is being absorbed by the ocean since there is less ice to cover it. Ice is lighter than the ocean water.
2	The ice will have melted, letting more heat be absorbed into the ocean.
3	The ice melt will cause the atmosphere to be warmer.
4	Less ice caps mean more water, the water is darker.

TABLE II. TOKENIZED TRANSACTIONS

Transactions	**Items**																	
	more	sun	light	absorb	ocean	less	ice	lighter	cover	water	melt	heat	let	atmosphere	warmer	cause	dark	cap
1	x	x	x	x	x	x	x	x	x	x								
2				x	x		x				x	x	x					
3							x				x			x	x	x		
4	x					x	x			x							x	x

III. ANALYZING STUDENTS' MECHANISTIC EXPLANATIONS ABOUT CLIMATE CHANGE

To apply our method to real data, we performed association rule mining on a dataset with 2,434 students' textual responses to a climate change question. The question was asked as part of pre/post-tests administered to students who engaged in an online climate change module [14] developed by the Concord Consortium (http://authoring.concord.org/sequences/47). See Fig. 1.

This climate change module has been implemented since 2011. The data corpus used in this study were collected between 2012 and 2014. The climate change module was designed to improve students' scientific argumentation practice with an emphasis on scientific reasoning and uncertainty articulation [15].

Students first read the question prompt "More sunlight can be absorbed by an object with a darker surface than one with a lighter surface. In the 1970s, sea ice covered about 10.8 million square kilometers of the Arctic Ocean. In 2010, sea ice covered 8.7 million square kilometers of the Arctic Ocean" and then made a prediction about "How might the decrease in the sea ice affect Earth's atmospheric temperature in the future?"

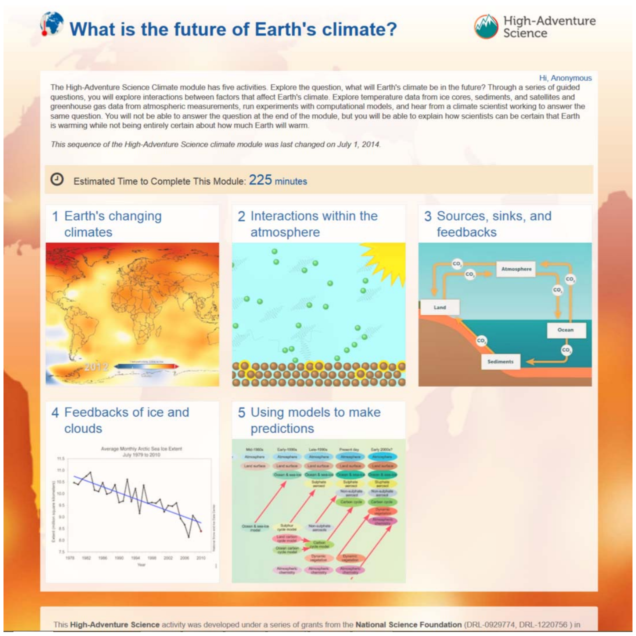

Fig. 1. A screenshot of the curriculum module called “What is the future of Earth’s climate?”, where the data used in this study were collected.

Students picked one of three choices listed below and explained the rationale for their choices in an open-ended question format.

a. It will increase the atmospheric temperature.
b. It will decrease the atmospheric temperature.
c. There will be no effect on the atmospheric temperature.

The responses were given a score between 0 – 4 by human coders, where 0 denotes irrelevant answers, 1 means non-normative ideas, while 2 – 4 means normative ideas with an increasing degree of links between these ideas. Based on these scores, the 2,434 responses were divided into three groups: the off-task group, where students did not answer the question and scored 0 (195), the no-link group, where students scored 1 because they provided a non-normative explanation without any link with other normative ideas (1,526), and the linked group, where students scored 2-4 for providing multiple normative ideas with increasing complexity in links (713). The off-task group was dropped from our analysis.

We performed association rule mining on both the high and low-scoring groups and discovered three predominant patterns in the data:

A. High-Scoring Group Generated More Well-Connected Rules

Although there are twice as many low-scoring students (1,526) as high-scoring students (713), association rule mining on high-scoring students’ responses generated significantly more rules (2,889 with 1,149 quality rules) than on low scoring-students’ responses (1,230 with 527 quality rules). This result shows that in high-scoring students’ responses the diversity of words and links are much higher.

Fig. 2 shows a balanced and well-connected network of rules for high-scoring students, while Fig. 3 shows a less connected network for low- scoring students. These results also imply that high-scoring students possess a much more complex knowledge structure pertaining to the climate change question.

B. High-Scoring Group Used More Sophisticated Mechanistic Explanations

We found 42 most frequent itemsets (see Fig. 2) with 35 unique items (see Table III) in high-scoring students’ responses and 31 most frequent itemsets (see Fig. 3) with 29 unique items (see Table III) in low-scoring students’. These two graphics visualize the rules, which contain these frequent itemsets. They are selected from the 2,289 rules. The size of an itemset corresponds to the frequency it appeared in the rules.

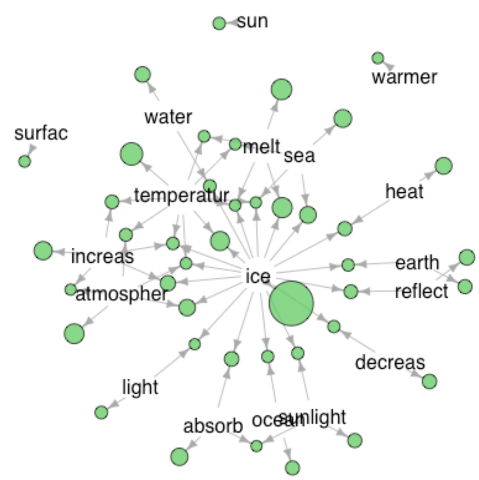

Fig. 2. Itemsets of the high-scoring groups. (Some words were stemmed so their last ltters were missing.)

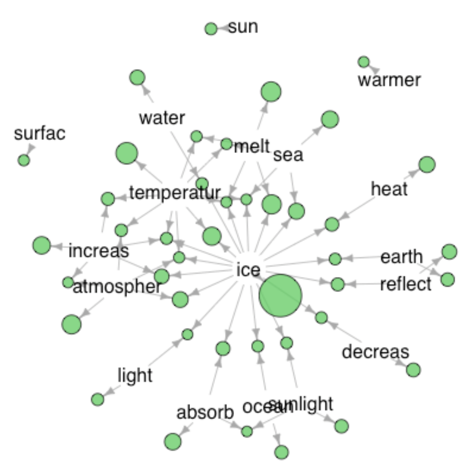

Fig. 3. Itemsets of the low-scoring groups.

In Table III, we list the unique items in the high-scoring group in a descending order by their support and matched them with the same items in the low-scoring group. Nine items that are present in the high-scoring group (Order 22, 24-26, 28-31, and 35) do not exist in the low group, and two items that are in the low group (Order 28 and 29) did not appear in the high group.

From these items, we identified six key verbs as activities in these responses: melt, absorb, reflect, rise, cover, and evaporate (bold items in Table III). These six verbs, together with other nouns or entities in the itemsets, comprise multiple activities, or sub-mechanistic explanations to the question prompt:

- ice **melts**,
- earth or dark color **absorbs** sunlight or heat,
- ice or water or light color **reflects** sunlight or heat,
- sea level or temperature **rises**,
- ice or water **covers** the earth surface,
- water **evaporates** into the atmosphere.

Although a few other verbs are present in this list, such as “increase”, “decrease”, and “affect”, they are not included as key verbs because they describe general changes instead of specific mechanisms. This is where human experts’ insights can be used to determine salient activities as compared to generic ones.

This itemset comparison shows that while most low-scoring students were only able to use the “melt” mechanism, very few of them used the “absorb” and “reflect” mechanisms in their explanations, even when the word “absorb” is in the prompt of this question. The support on these two items in the high group is about three times higher than the low group (0.18 vs. 0.06, and 0.15 vs. 0.04), and the order of these two words are 7 vs. 15 and 10 vs. 20 in the high and low groups respectively

C. *Color-Related Entities are Missing from the Low-Scoring Group*

Although the first 21 items in the high-scoring group are also present in the low-scoring group, starting from the 22nd item, some words are missing from the low-scoring group, which indicates low-scoring students did not take these factors into account when giving explanations (see Table III). Items 22, 24, and 26—“color”, “dark”, and “white”, (the italicized items in Table III) —are all color related.

TABLE III. COMPARING ITEMSETS OF HIGH- AND LOW-SCORING GROUPS

Item	High Group		Low Group		Item	High Group		Low Group	
	Order	Support	Order	Support		Order	Support	Order	Support
{ice}	1	0.64	1	0.64	{air}	20	0.06	14	0.08
{temperature}	2	0.27	2	0.33	{warm}	21	0.06	16	0.07
{atmosphere}	3	0.23	4	0.23	*{color}*	*22*	*0.05*		
{melt}	**4**	**0.24**	**3**	**0.27**	{cold}	23	0.07	10	0.12
{sea}	5	0.19	6	0.17	*{dark}*	*24*	*0.05*		
{increase}	6	0.2	5	0.2	*{ray}*	*25*	*0.04*		
{absorb}	**7**	**0.18**	**15**	**0.06**	*{white}*	*26*	*0.04*		
{heat}	8	0.17	12	0.1	**{cover}**	**27**	**0.03**	**22**	**0.03**
{water}	9	0.15	8	0.13	**{evaporate}**	**28**	**0.03**		
{reflect}	**10**	**0.15**	**20**	**0.04**	{level}	29	0.03		
{ocean}	11	0.13	11	0.09	{amount}	30	0.02		
{sunlight}	12	0.13	19	0.05	{area}	31	0.02		
{earth}	13	0.13	9	0.12	{affect}	32	0.03	24	0.03
{decrease}	14	0.14	7	0.16	{effect}	33	0.02	25	0.03
{light}	15	0.11	21	0.04	{global}	34	0.02	26	0.03
{sun}	16	0.1	18	0.06	{answer}	35	0.03		
{surface}	17	0.09	23	0.03	{predict}			27	0.02
{warmer}	18	0.08	13	0.1	{hot}			28	0.02
{rise}	**19**	**0.06**	**17**	**0.06**					

Color is a very important property of entities in the mechanism of climate change related to albedo effect. As mentioned in the prompt of the question, a darker surface absorbs more sunlight and heat than a lighter surface. However, the consideration of color difference in absorbing heat is missing from the low-scoring group's explanation. Further instructional intervention is needed for this group to take these factors into account.

After identifying the six key verbs shown in Table III, we generated words that were associated with each key verb to show the mechanistic explanations centering on these key verbs. Table IV shows the comparison between the high- and low-scoring groups on the words associated with one of the key verbs "absorb". Students in the high-scoring group made connections between "absorb" and many other concepts, such as energy and radiation, which are absent from low-scoring students' explanations.

IV. CONCLUSION

We introduced a method of detecting students' mechanistic explanations of scientific phenomena through a semi-automated technique that involves association rule mining and human insight. As the example of applying this method on the climate change data shows, this method leverages the use of widely available data—students' written responses to questions generated in science classes and tests in online learning environments— and provides teachers and researchers important insight into students' knowledge structures. Such insights can be used to improve teaching and learning, and to inform the design of future instructional interventions.

TABLE IV. UNIQUE ITEMS ASSOCIATED WITH "ABSORB" IN HIGH AND LOW-SCORING GROUPS

High Group	Low Group	High Group	Low Group
absorb	absorb	melt	melt
air		object	
amount		ocean	ocean
area		radiation	
atmosphere	atmosphere	ray	
bounce		reflect	reflect
color	color	rise	
cover		sea	sea
dark		sun	sun
decrease	decrease	sunlight	sunlight
earth	earth	surface	surface
energy		temperature	temperature
heat	heat	warmer	
ice	ice	water	water
increase	increase	white	
light	light		

ACKNOWLEDGMENT

This material is in part based upon work supported by the National Science Foundation under grants No. DRL-0929774 and No. DRL-1220756. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] National Research Council, A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas. Committee on a Conceptual Framework for New K-12 Science Education Standards., Washington, DC: National Academy Press, 2012.

[2] T. S. Kuhn, The essential tension. Chicago: University of Chicago Press, 1977.

[3] R. S. Russ, R. E. Scherr, D. Hammer, and J. Mikeska, "Recognizing mechanistic reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science," Sci. Educ., vol. 92 no. 3, pp. 499–525, 2008.

[4] B. Koslowski, Theory and evidence: The development of scientific reasoning. Cambridge, MA: MIT Press, 1996.

[5] P. Machamer, D. Darden, and C. F. Craver, "Thinking about mechanisms," Philos. Sci, vol. 67, pp.1 – 25, 2000.

[6] W. Xing, G. Rui, E. Petakovic, and S. Goggins. "Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory." *Computers in Human Behavior* vol .47, pp. 168-181, 2015.

[7] S. Goggins, W. Xing, X. Chen, B. Chen, and B. Wadholm. "Learning Analytics at" Small" Scale: Exploring a Complexity-Grounded Model for Assessment Automation." *Journal of Universal Computer Science* 21, no. 1, pp. 66-92. 2015.

[8] W. Xing, and S. Goggins. "Learning analytics in outer space: a Hidden Naïve Bayes model for automatic student off-task behavior detection." In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, pp. 176-183. ACM, 2015.

[9] B. Sherin, "A computational study of commonsense science: An exploration in the automated analysis of clinical interview data," *J. Learn. Sci.,* vol. 22 no. 4, pp. 600-638, 2013.

[10] W. Xing, R. Wadholm, E. Petakovic, and S. Goggins. "Group Learning Assessment: Developing a Theory-Informed Analytics." *Journal of Educational Technology & Society* 18, no. 2, pp. 110-128, 2015.

[11] W. Xing,, S. Kim, and S. Goggins. "Modeling Performance in Asynchronous CSCL: An Exploration of Social Ability, Collective Efficacy and Social Interaction." Vol.1, pp. 276-283, CSCL, 2015.

[12] C. F. Craver, "Interlevel experiments, multilevel mechanisms in the neuroscience of memory," Philos. Sci, (Suppl), vol. 69, pp.S83 – S97, 2002.

[13] R. Agrawal and M. Batra, "A detailed study on text mining techniques," IJSCE, vol. 2, no. 6, pp.118–121, 2013

[14] Pallant, A., & Lee, H. -S. (2015). "Constructing scientific arguments using evidence from computational models". Journal of Science Education and Technology, 24(2/3), 378-395.

[15] H. Lee, S., Liu, O. Pallant, K. Crotts, and S. Pryputniewicz (2014). Assessment of uncertainty-infused scientific argumentation. Journal of Research in Science Teaching, 51(5), 581-605.