Learning Analytics in Support of Qualitative Analysis

Bruce Sherin, Northwestern University, bsherin@northwestern.edu
Nicole B. Kersting, University of Arizona, nickik@email.arizona.edu
Matthew Berland, University of Wisconsin–Madison, mberland@wisc.edu

Abstract: The purpose of this manuscript is to describe a style of learning analytics research that makes closer contact with the qualitative methods of the learning sciences. We do so by drawing on two examples of existing research. The first is our prior work attempting to automate part of the Classroom Video Analysis assessment, which seeks to measure knowledge of teachers that is predictive of their performance. The second is from work which looks at how middle school students explain the Earth's seasons. In that work, we attempted to use unsupervised methods to capture elements of what was previously a fully qualitative analysis. Our goal is to provide the reader with a sense for this style of research that brings qualitative analysis and analytic methods into closer contact. To accomplish this, we make use of a new system called Tactic, designed specifically to support this mode of research.

Introduction

The purpose of this manuscript is to argue for a style of learning analytics research that makes closer contact with the qualitative methods of the learning sciences. As a set of innovations in educational research, learning analytics is not only about the application of new analytic algorithms to data. These new algorithms are also applied to new types of data, such as computer log files and biometric data. They are also applied to data that is larger in scope. Furthermore, the types of questions being answered are often somewhat different (Baker & Yacef, 2009; Martin & Sherin, 2013).

However, it is possible to apply some techniques from learning analytics to more traditional qualitative data, and with aims and methods that hew more closely to established methods. Why would we want to do this? There are a number of obvious reasons. First, if we apply traditional and learning analytic methods to similar corpora and with similar aims as existing studies, we have the potential to produce distinct but complementary analyses. A second benefit is the potential for reducing the labor associated with traditional types of analysis, especially in qualitative research. Even when a data corpus is small, from the point of view of learning analytics, manual analysis can be slow and laborious.

But, more exciting than these somewhat obvious benefits is the possibility of bringing established and learning analytic methods together in a manner that augments both. If we work with smaller data corpora about which human analysts have detailed, intimate knowledge, there is the potential to support the design of more finely tuned computational analyses, and to better interpret the results of those analyses. Computational analysis can in turn provide new perspectives on qualitative data analysis, even helping to surface tacit knowledge of human coders. Finally, tightly integrated traditional and computational analyses have the potential to provide a kind of triangulation that increases our understanding of—and confidence in—both.

Our purpose in this paper is to illustrate these points. We do so by drawing on two examples of existing research. The first is our prior work attempting to automate part of the Classroom Video Analysis assessment, an assessment that seeks to measure knowledge of teachers that is predictive of their performance (Kersting, 2008; Kersting, Givvin, Thompson, Santagata, & Stigler, 2012; Kersting, Givvin, Sotelo, & Stigler, 2010). In that analysis, we used data coded by human raters to train classifiers using a modified Naïve Bayes approach (Kersting, Sherin, & Stigler, 2014). The second is from work which looks at how middle school students understand the Earth's seasons (Sherin, Krakowski, & Lee, 2012). In that work, we used unsupervised methods to capture elements of what was previously a fully qualitative analysis (Sherin, 2013). Both of these lines of work followed a similar trajectory; in each case, we revisited prior research, applying learning analytic methods to existing data.

Tactic: Infrastructure for learning analytics research

We begin with a brief discussion of the infrastructure available for this new style of learning analytics research. Learning analytics requires, of course, software tools that can perform the relevant computational analyses. One approach that analysts can take is to write their own code, usually in a way that is specific to the analysis at hand. Even in this case, however, analysts draw on publically available software libraries. In addition, there do exist GUI tools that reduce—but generally don't eliminate—the programming required. These include Weka (Hall et al., 2009), RapidMiner (Hall et al., 2009), and LightSide (Mayfield & Rosé, 2013).

In this paper, we will display analyses as they look in a new web-based environment text mining environment called Tactic, that we are currently developing. Tactic is specifically tuned for working in an interactive mode in which the analyst moves back and forth between the data and analytics. The data is always front and center, and Tactic incorporates visual features that assist the user in seeing relationships between the original data and analytics applied to that data. In addition, Tactic does not impose a particular decomposition of the analytic process on users. So, for example, we do not assume that a user's analysis will be separated into a pre-processing step, followed by a modeling step. We believe instead that the interactive style of research should allow for a more emergent workflow, which might be somewhat specific to each project.

Figure 1 shows an analysis workspace, in process, with two *tiles* visible on the right. (This will be further explained below.) Tiles do most of the computational work, and Tactic includes a base set of tiles which users modify or use unchanged. Users can also create entirely new tiles. All programming is done in Python, within the web environment. There is nothing in Tactic that represents a radical advancement beyond existing tools. What is important is that it is tuned for a particular style of work, which we now explain in more detail.

First example: The CVA assessment

Original CVA research

The second author of this manuscript and her colleagues developed the Classroom Video Analysis (CVA) assessment as a means of assessing the *usable* knowledge of mathematics teachers (Kersting, 2008; Kersting et al., 2012; Kersting et al., 2010). Here, by "usable" knowledge, we mean the knowledge that teachers can access and apply in the classroom. In the CVA, teachers watch a series of videos of classroom instruction. After each video, they are asked to "analyze how the teacher and student(s) interacted around the mathematical content." They type their comments into a web-based form, and these responses are then scored by trained coders. For example, in response to a video clip showing part of a lesson on fractions, a teacher responded: ''The teacher is asking the student questions to narrow down his search to find equivalent fractions. I don't know if the boy understood the meaning of the giant 1. It was almost like the student was guessing, and judging how the teacher responded he would know if he was right or wrong.''

The CVA has been the focus of an extended program of traditional (non-computational) research. That work has found that the CVA can predict instructional quality, as measured by direct observations of teaching. Strikingly, one dimension measured by the CVA, teachers' ability to make suggestions for improving the teaching episode, has been found to predict the learning of students in the classrooms of teachers studied.

Prior computational methods

The CVA is potentially a powerful means of assessing teachers. However, its wider use has been somewhat hindered by the fact that it is laborious for coders to score teachers' written responses to the video clips and requires coders to undergo extensive training. For this reason, automating some part of the CVA analysis could have substantial benefits. In addition, the power and success of the CVA merits additional study. Learning analytic methods have the potential to not only replicate the work of manual coders; they might also help us to better understand why and how coders assign the codes that they do.

In prior computational work on the CVA (Kersting, Sherin, & Stigler, 2014), we made use of CVA assessments targeting three mathematical content areas: (1) fractions, (2) ratios and proportions, and (3) variables, expressions, and equations. For the fractions data, 238 teachers responded to 13 clips. For the ratio and proportion data, 238 teachers responded to 13 clips. Finally, for the variables, expressions, and equations data, 249 teachers responded to 14 clips.

All of this data was manually scored according to four rubrics: mathematical content (MC), student thinking (ST), suggestions for improvement (SI), and depth of interpretation (DI). Each response was given a code of 0, 1, or 2 for each of these four rubrics, with 2 being the best score, and 0 being the worst. For example, for the MC rubric (on which we will focus) a response was given a score of 0 if it didn't mention mathematical content. In contrast, it was given a score of 2 if there was an in-depth analysis of the content, which went beyond the mathematics described in the clip. A more superficial discussion of the mathematical content received a score of 1. For instance, the teacher response cited above was given a score of 1 because it mentions the mathematical content but doesn't go beyond the mathematics described in the clip.

In our computational analysis, we trained a unique Naïve Bayes classifier for each of the 40 clips. In essence, each classifier is a software model, that can take a teacher response as an input, and output the most likely code. The model is initially "trained," by feeding in sample responses that have been coded by human analysts. In the case of the Naïve Bayes classifiers we constructed, this training is essentially done by gathering statistics; the algorithm looks at how frequently specific individual words are associated, in the human-coded data, with a specific code. Furthermore, to train our classifiers, we did not include all of the words that appear in responses. We first discarded all words that appeared on a "stop list" of common words. We then kept only the 100 most frequent words that remained. After training, the Naïve Bayes algorithm combines these statistics to code new responses. More expert readers will recognize that, from the point of view of machine learning, the Naïve Bayes algorithm we have selected is relatively simple, and likely not optimal. The simple nature of our selected method was driven, in part, by our desire to produce classifiers with internals that could be easily inspected and understood, as we illustrate below.

In our computational work, one fundamental question was whether we could ultimately hope to replace human coders with automated scoring, as a means of predicting teacher quality. If this is our goal, it means that we should not be primarily interested in replicating individual codes on specific clips. Rather, we are interested in the ability of the automated coding to predict teacher effectiveness. This leads us to be most interested in (1) the *composite* scores, across all clips, for a given subject-matter area, and (2) the *correlations* between these composite scores and the composite scores given by human coders.

Table 1: Average quadratic weighted kappa and correlations for the CVA analysis

Topic	**Kappa**	**Spearman Correlation**
Fractions	.51	.89
Ratio	.51	.86
VEE	.55	.91

Table 1 shows the average results for this analysis, in which the composite scores are averaged across multiple runs. As shown in the table, overall average correlations were high. The correlations for three of the four sub-scores – MC, ST, and DI – were similarly high, ranging from .77 to .91. For SI, the correlations were somewhat lower, ranging from .49 to .69. (The correlations for sub-scores are not shown in the table.)

We also computed quadratic weighted kappa values to provide a measure of agreement, beyond chance, between human and automated scoring. The results shown in Table 1 report the Kappa values for each of the subject matter areas, averaged across clips. These values represent moderate agreement. As with the correlation results, there was variation across the four rubrics. Results for MC and DI ranged from .56 to .64. In contrast, results for SI and ST ranged from .36 to .43 and .43 to.47 respectively.

Interactive analysis with tactic

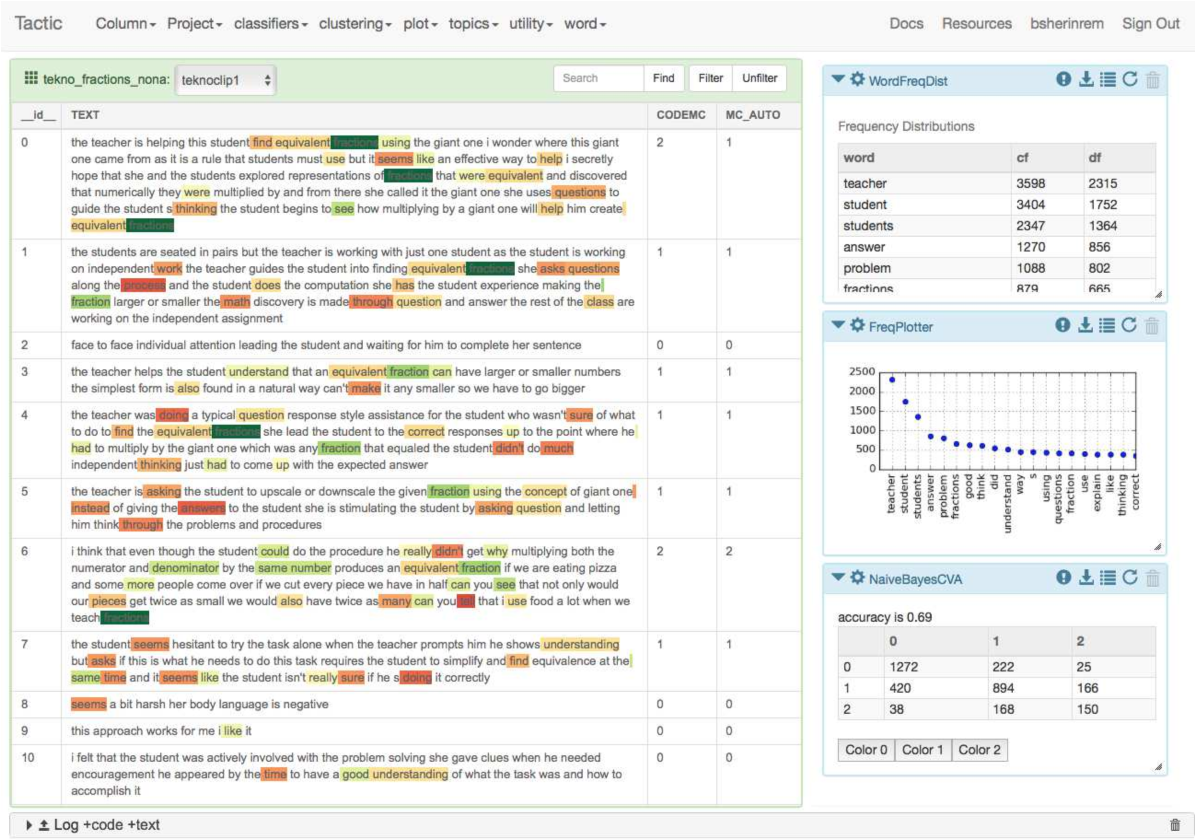
The results summarized above suggest that applications of computational methods to the CVA data can achieve one of the benefits mentioned earlier; namely it seems to have the possibility of reducing the labor required by human analysts. We next want to show what the benefits might be for a more interactive style of analysis.

Figure 1 shows an analysis of the CVA fractions corpus, as it might appear in Tactic, at an intermediate stage. The data table is on the left. It has the text of each response. In addition, the column labeled CODEMC has the codes given by the human raters – 0, 1, or 2 – for the MC rubric. (Here, for simplicity, we will focus just on that rubric.) The popup list above the table allows the user to select other documents in the corpus, which here correspond each of the 13 fractions clips.

Analysis tiles are added using the menus at the top of the display. Once a tile has been added to the environment, the user clicks on the gear icon at the top-left of the tile to configure options on the back of the tile. Here three analysis tiles have been added to the work environment. The first, very basic, example, is the tile labeled WordFreqDist. When this tile is configured and run, it displays a table of the corpus frequency and document frequency for the most frequent words. The tile also provides a simple type of interactivity; namely, clicking on a word highlights matching words in the main data table. In addition, other tiles can access python data structures that are exported by a tile. In this case, there is a tile showing a plot of the document frequencies exported by the WordFreqDist tile.

The third tile allows us to illustrate some more interesting possibilities. This tile is set up to run a very simple Naïve Bayes analysis—a much more rudimentary analysis than was used in (Kersting, Sherin, & Stigler, 2014). Here we are training one classifier to work across all 13 of the fractions clips. In addition, (for the purposes of keeping things simple) we are committing what is usually an egregious error; we are training and testing on the entirety of the data.

When run, this tile trains a classifier using all of the fractions data. It then codes each response, and writes those codes into a new column, here labeled MC_AUTO. In addition, a confusion matrix is displayed on the front of the tile. Again, this provides some simple interactivity: clicking in a cell of the confusion matrix shows a new table of responses that correspond to that cell. (This isn't shown in the figure.) Those responses can, in turn, be clicked to be viewed in context in the main data table.



__id__	TEXT	CODEMC	MC_AUTO
0	the teacher is helping this student find equivalent fractions using the giant one i wonder where this giant one came from as it is a rule that students must use but it seems like an effective way to help i secretly hope that she and the students explored representations of fractions that were equivalent and discovered that numerically they were multiplied by and from there she called it the giant one she uses questions to guide the student s thinking the student begins to see how multiplying by a giant one will help him create equivalent fractions	2	1
1	the students are seated in pairs but the teacher is working with just one student as the student is working on independent work the teacher guides the student into finding equivalent fractions she asks questions along the process and the student does the computation she has the student experience making the fraction larger or smaller the math discovery is made through question and answer the rest of the class are working on the independent assignment	1	1
2	face to face individual attention leading the student and waiting for him to complete her sentence	0	0
3	the teacher helps the student understand that an equivalent fraction can have larger or smaller numbers the simplest form is also found in a natural way can't make it any smaller so we have to go bigger	1	1
4	the teacher was doing a typical question response style assistance for the student who wasn't sure of what to do to find the equivalent fractions she lead the student to the correct responses up to the point where he had to multiply by the giant one which was any fraction that equaled the student didn't do much independent thinking just had to come up with the expected answer	1	1
5	the teacher is asking the student to upscale or downscale the given fraction using the concept of giant one instead of giving the answers to the student she is stimulating the student by asking question and letting him think through the problems and procedures	1	1
6	i think that even though the student could do the procedure he really didn't get why multiplying both the numerator and denominator by the same number produces an equivalent fraction if we are eating pizza and some more people come over if we cut every piece we have in half can you see that not only would our pieces get twice as small we would also have twice as many can you tell that i use food a lot when we teach fractions	2	2
7	the student seems hesitant to try the task alone when the teacher prompts him he shows understanding but asks if this is what he needs to do this task requires the student to simplify and find equivalence at the same time and it seems like the student isn't really sure if he s doing it correctly	1	1
8	seems a bit harsh her body language is negative	0	0
9	this approach works for me i like it	0	0
10	i felt that the student was actively involved with the problem solving she gave clues when he needed encouragement he appeared by the time to have a good understanding of what the task was and how to accomplish it	0	0

Frequency Distributions

word	cf	df
teacher	3598	2315
student	3404	1752
students	2347	1364
answer	1270	856
problem	1088	802
fractions	879	665

accuracy is 0.69

	0	1	2
0	1272	222	25
1	420	894	166
2	38	168	150

Figure 1. Analysis of the CVA fractions corpus in process.

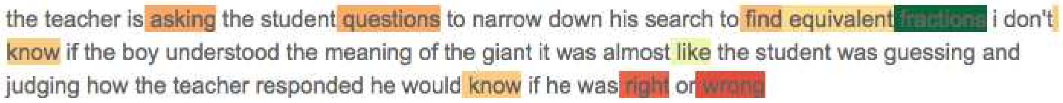

the teacher is asking the student questions to narrow down his search to find equivalent fractions i don't know if the boy understood the meaning of the giant it was almost like the student was guessing and judging how the teacher responded he would know if he was right or wrong

Figure 2. Text coloring for the sample response.

However, a more interesting type of interactivity is possible, if we provide the qualitative data analyst with a window into the automated analysis. The Naïve Bayes tile shown in Figure 1 has three buttons, one corresponding to each of the three codes: 0, 1, and 2. Clicking on one of these buttons causes the text in the main table to be highlighted with varying colors, with the colors selected according to the conditional probability of the highlighted word, given the code corresponding to the clicked button. The color palette used goes from red to yellow to green. So, for example, dark green indicates a high probability of the word for the given code, and dark red a low probability.

Figure 1 shows the coloring obtained when the button corresponding to a score of 2 is pressed. Again, this is the highest code, and a score of 2 represents a deep analysis of mathematical content. Thus, when colored in this way, we are seeing which words the coders associate with a desirable response on this dimension. Figure 2 shows how the response we quoted earlier is colored. The word "fractions" is dark green, indicating a high probability. This is not surprising, given that the coders are looking at whether the teachers are attending to the mathematical content. In contrast, the words "right" and "wrong" are less indicative of a deep analysis. This is also perhaps not surprising – coders might well believe that saying that an answer is right or wrong implies a focus on superficial aspects of the mathematical content.

Second example: Seasons analysis

Original seasons research

In this second example, we apply computational methods as part of what was originally a program of pure qualitative research (Sherin, Krakowski, & Lee, 2012). The work draws on interviews in which middle-school students were asked a range of questions about the Earth, its climate, and space science. More specifically, it draws on a portion of these interviews in which the students were asked to explain the Earth's seasons.

The interviews were conducted in the manner of a clinical interview, which meant that the interviewer had the freedom to ask follow-up questions for clarification, and to further probe the student. The seasons portion of the interview began with the interviewer asking, "Why is it warmer in the summer and colder in the winter." The interviewer then asked various follow-up questions. This included asking the student to draw a picture. Interviewers were also prepared with challenges to be introduced based on the initial explanations provided by students.

The explanations given by students were quite varied. However, in describing the spread of explanations, we have found it helpful to have in mind three prototype explanations. The first prototype we call *close-farther* explanations. In a closer-farther explanation, the whole Earth moves so that it is sometimes closer to the sun and sometimes farther away. When it is closer to the sun, the whole Earth experiences summer. When it is farther, the whole Earth experiences winter.

The second type we call *side-based* explanations. In a side-based explanation, the Earth moves in such a manner that one side, at times, faces toward the sun, while the other side faces away. (Usually—though not always—this is caused by the rotation of the Earth on its axis.) The side facing toward the sun experiences summer, and the side facing away experiences winter.

The final type of explanation is *tilt-based.* In these explanations, the Earth somehow moves so that one hemisphere is tilted toward the sun and the other away, with the side tilted toward the sun experiencing summer. This category includes, but is not limited to, the scientifically-accepted answer. In the accepted answer, the seasons are intimately linked to the Earth's tilt. The Earth's axis always points in the same direction. But, because the it orbits around the sun, first one hemisphere, then the other, will be inclined toward the sun. Furthermore, the hemisphere of the Earth tilted toward the sun receives more direct sunlight, and this more direct sunlight causes warmer conditions.

In the original work on this corpus, our argument was not that students each give one from amongst a small set of explanations. Rather we argued that students generally construct explanations out of a number of small components, which we called nodes, leading to a larger number of explanation structures that mix and match these components. We called these explanation structures *dynamic mental constructs,* or DMCs. Furthermore, we presented evidence, in the form of examples, that these DMCs could shift rapidly, over the course of an interview.

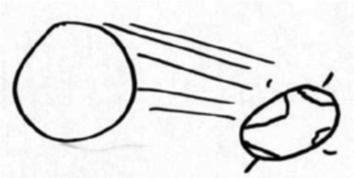


Figure 3. Edgar's seasons diagram.

For illustration, we present a portion of the text of the interview with one student, who we refer to as Edgar. In his initial explanation, Edgar drew the diagram shown in Figure 3, as he said:

> E: Here's the Earth slanted. Here's the axis. Here's the North Pole, here's the south pole, and here's our country. And the sun's right here [draws circle on the left], and the rays are hitting directly right here, so things are getting hotter in the summer and when this thing turns, the country will be here and the sun can't reach as much. It's not as hot as the winter.

In our prior work, we argued that Edgar's first explanation, given in the passage above, is a variant side-based explanation; the Earth spins, and the side facing the sun is warmer because it receives more direct sunlight. As the interview proceeded, Edgar seemed to recall that the Earth orbits the sun, in addition to rotating on its axes. This led him to transition to a fairly traditional closer-farther explanation.

E: Actually, I don’t think this moves [indicates Earth on drawing] it turns and it moves like that [gestures with a pencil to show an orbiting and spinning Earth] and it turns and that thing like is um further away once it orbit around the s- Earth- I mean the sun.

I: It’s further away?

E: Yeah, and somehow like that going further off and I think sun rays wouldn’t reach.

These brief excerpts illustrate, first, how a student explanation could shift, from one moment to the next, over the course of an interview. Edgar began with a side-based explanation, and shifted to a fundamentally different explanation, a closer-farther explanation. We can also get some sense for how explanations can be seen as built out of components, that are mixed and matched. For example, as part of his initial side-based explanation, Edgar makes use of the notion that *more direct light causes parts of the Earth to be warmer*, a component that we would more typically associate with a tilt-based explanation. However, here he has incorporated this idea as part of a side-based explanation.

Prior computational methods

The issues here are quite different than those encountered in the CVA research described above. In that case, there was a well-defined coding scheme, but that scheme was laborious to execute. Here, in contrast, the initial publications did not use a coding scheme at all; instead, they used examples from the corpus to illustrate and support theoretical claims. This meant that, in important ways, readers had to trust our interpretive ability – our ability to “see” the components and DMCs in the idea.

We believed that, if an automated analysis could replicate at least some aspects of our analysis, that would place the whole endeavor on more solid ground. Furthermore, we felt it was necessary to set the bar high. We did not want to specify, in advance, the components or DMCs that our automated analysis would use. The discovery of these elements is where the important, and controversial work, lies. Thus, we wanted to have the automated analysis discover, on its own, both the components and DMCs.

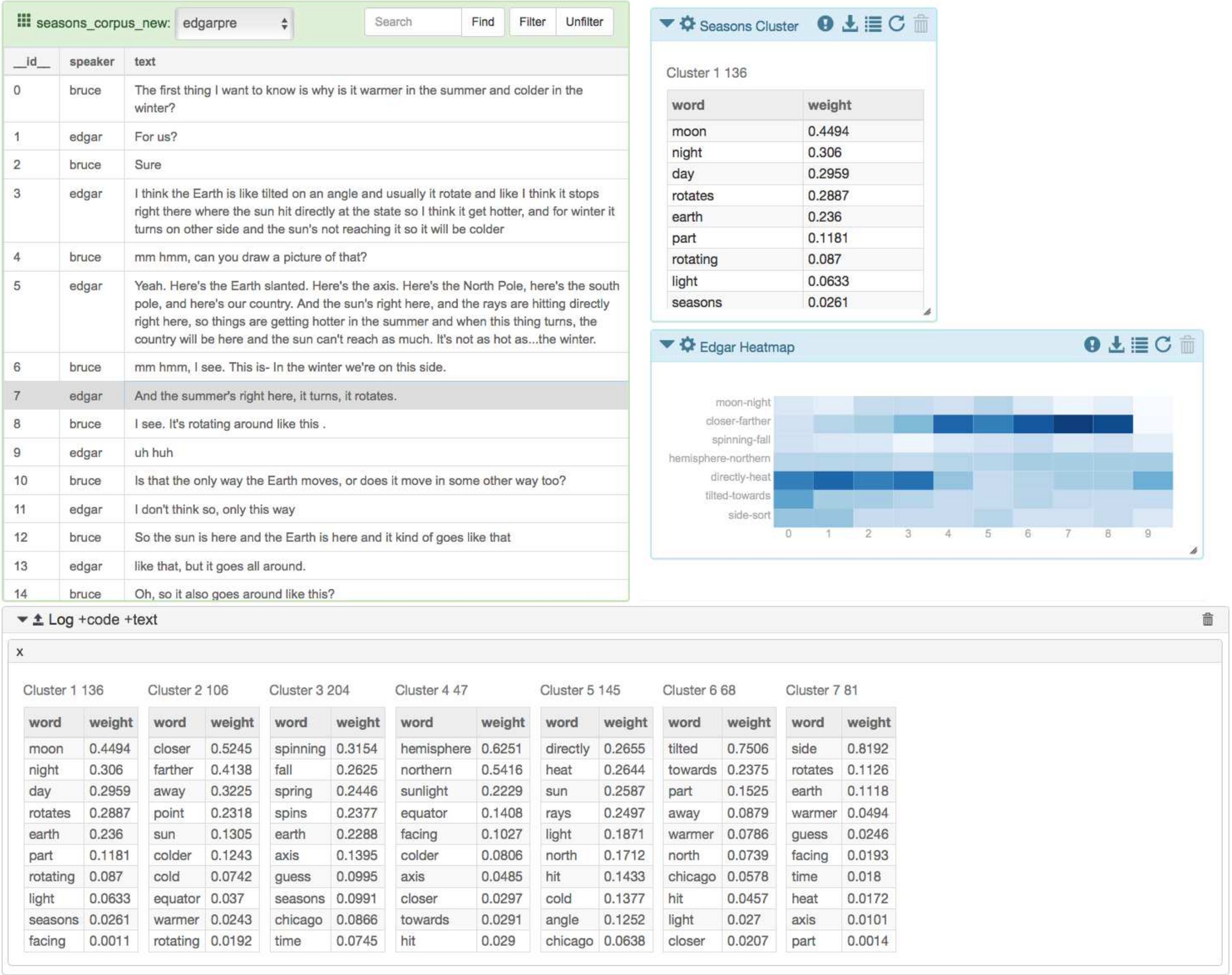
Going in, there were many reasons to think that the data might not be amenable to much in the way of a computational analysis. As we will see, the size of the data corpus is quite small. In addition, the interviews can be rambling and unclear. Furthermore, as in the interview with Edgar, there was a lot of gesturing, and references to drawings, important communicative elements that weren’t included in our computational analysis. These challenges are severe. But they are the sort of challenges that we will frequently need to overcome if we are going to use learning analytics in tandem with traditional qualitative data analysis.

Our computational analysis is presented in its most extended form in (Sherin, 2013). The data we used were the transcripts of interviews with 54 middle school students. As a first step in the analysis, all comments by the interviewer were removed, and utterances by the student in each interview were concatenated. The interviews were then broken into overlapping 100-word segments, using a moving window that moved forward in 25-word increments. So, the first segment had words 1-100, the next segment words 25-125, etc. This segmenting process resulted in a total of 794 segments, across all 54 of the interviews.

Words in a stop list of 782 words were removed from each of the segments. When this was done, the segments contained words from a vocabulary 647 unique words. Each of the 794 segments were then converted to word vectors, using a weight function of 1 + log(tf), where tf was the number of times the word appeared in the segment. What this means is that each segment of text was converted to a list of 647 numbers, where each number corresponded to one word in the vocabulary of 647 words. Finally, we performed one non-standard form of preprocessing. Namely, we computed what we have called *deviation* vectors. All of the 794 vectors were added, and the result normalized, to construct a sort of super-average. Then this average vector was subtracted from each of the 794 segment vectors. As explained in more detail in Sherin (2013), this was necessary for the next stage of the analysis to produce meaningful results. The result of the above process was that each segment of text was mapped to a point in a 647-dimensional space. As a final step, these points were clustered into 7 groups, using hierarchical agglomerative clustering. These clusters were interpreted as aligning with the knowledge components in our qualitative analysis. Ultimately, we argued that the analysis could in fact reproduce important aspects of the qualitative analysis, thus providing a new type of support for the original theoretical claims. We attempt to briefly illustrate this below.

Analysis in tactic

Figure 4 shows a Tactic workspace in which we have replicated some elements of the analysis reported in Sherin (2013). The top right tile in the workspace replicates the entire clustering analysis, described above. When this analysis is complete, the face of the tile displays a set of 7 tables, one for each of the clusters. These can be scrolled through (but only part of the first table is visible in the figure). To make the results more visible here, these were sent to the Log area at the bottom of the workspace. These tables show the highest weighted words in each cluster.



seasons_corpus_new: edgarpre

Search | Find | Filter | Unfilter

__id__	speaker	text
0	bruce	The first thing I want to know is why is it warmer in the summer and colder in the winter?
1	edgar	For us?
2	bruce	Sure
3	edgar	I think the Earth is like tilted on an angle and usually it rotate and like I think it stops right there where the sun hit directly at the state so I think it get hotter, and for winter it turns on other side and the sun's not reaching it so it will be colder
4	bruce	mm hmm, can you draw a picture of that?
5	edgar	Yeah. Here's the Earth slanted. Here's the axis. Here's the North Pole, here's the south pole, and here's our country. And the sun's right here, and the rays are hitting directly right here, so things are getting hotter in the summer and when this thing turns, the country will be here and the sun can't reach as much. It's not as hot as...the winter.
6	bruce	mm hmm, I see. This is- In the winter we're on this side.
7	edgar	And the summer's right here, it turns, it rotates.
8	bruce	I see. It's rotating around like this .
9	edgar	uh huh
10	bruce	Is that the only way the Earth moves, or does it move in some other way too?
11	edgar	I don't think so, only this way
12	bruce	So the sun is here and the Earth is here and it kind of goes like that
13	edgar	like that, but it goes all around.
14	bruce	Oh, so it also goes around like this?

Seasons Cluster

Cluster 1 136

word	weight
moon	0.4494
night	0.306
day	0.2959
rotates	0.2887
earth	0.236
part	0.1181
rotating	0.087
light	0.0633
seasons	0.0261

Edgar Heatmap

Log +code +text

x

Cluster 1 136		Cluster 2 106		Cluster 3 204		Cluster 4 47		Cluster 5 145		Cluster 6 68		Cluster 7 81	
word	weight	word	weight	word	weight	word	weight	word	weight	word	weight	word	weight
moon	0.4494	closer	0.5245	spinning	0.3154	hemisphere	0.6251	directly	0.2655	tilted	0.7506	side	0.8192
night	0.306	farther	0.4138	fall	0.2625	northern	0.5416	heat	0.2644	towards	0.2375	rotates	0.1126
day	0.2959	away	0.3225	spring	0.2446	sunlight	0.2229	sun	0.2587	part	0.1525	earth	0.1118
rotates	0.2887	point	0.2318	spins	0.2377	equator	0.1408	rays	0.2497	away	0.0879	warmer	0.0494
earth	0.236	sun	0.1305	earth	0.2288	facing	0.1027	light	0.1871	warmer	0.0786	guess	0.0246
part	0.1181	colder	0.1243	axis	0.1395	colder	0.0806	north	0.1712	north	0.0739	facing	0.0193
rotating	0.087	cold	0.0742	guess	0.0995	axis	0.0485	hit	0.1433	chicago	0.0578	time	0.018
light	0.0633	equator	0.037	seasons	0.0991	closer	0.0297	cold	0.1377	hit	0.0457	heat	0.0172
seasons	0.0261	warmer	0.0243	chicago	0.0866	towards	0.0291	angle	0.1252	light	0.027	axis	0.0101
facing	0.0011	rotating	0.0192	time	0.0745	hit	0.029	chicago	0.0638	closer	0.0207	part	0.0014

Figure 4. Sample Tactic work environment for the seasons analysis.

The clusters are, we feel, relatively easy to align with the terms of our theoretical analysis. First, Cluster 2 has *closer, farther*, and *away*, as its top words. This suggests that the corresponding knowledge component could play a role in any explanation that focuses on the proximity of all or part of the Earth to the sun (i.e., a closer-farther explanation). There are also clusters we would most expect to see in side-based explanations; Cluster 7 has *side* as its highest-weighted word and *rotates* second. Cluster 3 talks about the spinning of the Earth. In addition, Cluster 1 is about the *moon*, *day*, and *night*. (Discussion of day and night often showed up with side-based explanation.) Finally, Clusters 4, 5, and 6 correspond to components we would typically associate with tilt-based explanations. Cluster 6 has *tilted, toward,* and *tilt* as its highest weighted terms, and Cluster 4 seems to be about the Earth's hemispheres. Lastly, Cluster 5 appears to focus on the directness of light striking the Earth.

One of the core elements of our original work was the analysis of the dynamics of interviews. To capture this with our computational analysis, we performed a secondary analysis in which the clusters were applied back to each of the original transcript. The heatmap in Figure 4 shows the result when this analysis is performed for the full text of the interview with Edgar. To create this heatmap, the text of the interview was prepared and segmented just as for the clustering analysis, and a vector computed for each segment. The result, for the interview with Edgar, was 10 vectors. We then found the dot product of each of these vectors, with the centroids of each of the 7 clusters. In the heatmap, the time of the interview proceeds from left to right, and each row of the heatmap corresponds to one of the 7 clusters. Darker shades represent a higher dot product between the section and cluster centroid. Thus, for example, the plot tells us that, the vector for segment 1 has its highest dot product with the *directly-heat* cluster, and segment 7 has the highest dot product with the *closer-farther* cluster. More generally, the plot does seem to broadly align with the qualitative analysis of the interview with Edgar. Recall that Edgar initially gave a side-based explanation, in which the side of the Earth facing the sun is warmer because it receives more direct sunlight. We see that, in fact, the segments in the first part of the interview do seem to align strongly with the *directly-heat* and *side-rotate* clusters. The first segment also seems to align with the *tilted-towards* cluster. This is because Edgar initially mentions the axis and poles. We also saw that, in the latter part of the interview, Edgar shifted to giving a closer-farther explanation. This is also captured in the heatmap. Thus, we can see that the automated analysis has captured at least some features of the dynamic account produced by the fully qualitative analysis, in our original work with this corpus.

Conclusion

Grounded theory, discovery-focused qualitative methods, and arduous qualitative coding have shown remarkable value for educational research over the last 50-100 years; but, so far, few environments or methods have leveraged learning analytics and machine learn to focus primarily on these analyses.

Our argument is that by using learning analytics as a tool for existing qualitative analyses, bringing the two fields closer together, could result in many new exciting *qualitative* findings, much as learning analytics has done in more traditionally quantitative fields. Furthermore—perhaps paradoxically—there are reasons to believe that qualitative methods and learning analytics can work particularly well together. A significant chunk of qualitative analysis has focused on *listening to the data* – building theory and hypotheses from data rather than prefiguring hypotheses before collecting those data. A similarly large portion of learning analytics concerns mining for patterns in data that are not immediately human-perceptible. By enabling these strands of learning analytics and qualitative learning sciences to mesh, and building tools that afford and support that mesh, we open up the potential for radical new understandings for and between the fields.

Finally, we suggested that this new meshed practice is more likely to be successful if it is supported by tools that are specifically tuned for this style of work. In that spirit, we presented the Tactic text mining environment as a tool that was designed with these aims in mind.

References

Baker, R., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, *1*(1), 3–17.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter*, *11*(1), 10–18.

Kersting, N. (2008). Using video clips of mathematics classroom instruction as item prompts to measure teachers' knowledge of teaching mathematics. *Educational and Psychological Measurement*.

Kersting, N. B., Givvin, K. B., Thompson, B. J., Santagata, R., & Stigler, J. W. (2012). Measuring Usable Knowledge: Teachers' Analyses of Mathematics Classroom Videos Predict Teaching Quality and Student Learning. *American Educational Research Journal*, *49*(3), 568–589.

Kersting, N., Givvin, K. B., Sotelo, F. L., & Stigler, J. W. (2010). Teachers' analyses of classroom video predict student learning of mathematics: Further explorations of a novel measure of teacher knowledge. *Journal of Teacher Education*, *61*(1–2), 172–181.

Kersting, N., Sherin, B. L., & Stigler, J. W. (2014). Automated Scoring of Teachers' Open-Ended Responses to Video Prompts: Bringing the Classroom-Video-Analysis Assessment to Scale. *Educational and Psychological Measurement*.

Martin, T., & Sherin, B. L. (2013). Learning Analytics and Computational Techniques for Detecting and Evaluating Patterns in Learning: An Introduction to the Special Issue. *Journal Of The Learning Sciences*, *22*(4), 511–520.

Mayfield, E., & Rosé, C. P. (2013). Open Source Machine Learning for Text. *Handbook of Automated Essay Evaluation: Current Applications and New Directions*.

Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., & Euler, T. (2006). Yale: Rapid prototyping for complex data mining tasks. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 935–940). ACM.

Sherin, B. L. (2013). A computational study of commonsense science: An exploration in the automated analysis of clinical interview data. *Journal of The Learning Sciences*, *22*(4), 600–635.

Sherin, B. L., Krakowski, M., & Lee, V. R. (2012). Some assembly required: How scientific explanations are constructed during clinical interviews. *Journal of Research in Science Teaching*, *49*(2), 166–198.