Notes on complexity in Learning Analytics utilising Category Theory and Topology

Ben Hicks

June 2020 - present

Notes on research around understanding the theory - practitioner gap, particularly in regards to Learning Analytics.

Defining the problem

Let us start at this beginning; there is a thing out there we call the world, and there are people in it.

The latter often find the aspects of the former somewhat mysterious and as a result, try to make sense of it. Now we have three things; aspects of the **World** that a particular **Person** tries to make sense of by some **process** ⁒ which we can draw together with some boxes and an arrow:

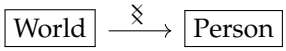

Figure 1: A diagram to help make sense of a person making sense of an aspect of the world through some sense making process, ⁒.

Here we have represented the process of sense-making, moving from the world space to the person space, with an arrow. This is an example of *map* or *mapping*, as it 'maps' things from World to things in Person.[1] The sense-making *map* is denoted using the symbol *samekh*, ⁒ (think of a braid, a weaving together of ideas and evidence).

There are many of ways to make sense of the world, but let us explore a rough dichotomy between *experiential* and *analytical*.

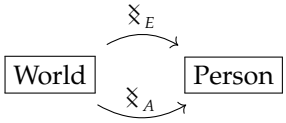

Figure 2: Different ways to make sense of the world, experiential ($⁒_E$) and analytical ($⁒_A$).

The *experiential* way of making sense of the world represents what the mind does to connect ones lived experience to a conceptual model of the world, an emic approach from the point of view of the subject. The *analytical* way can be thought of as a process that can be separate from the lived experiences of the person, an etic approach from the point of view of an observer, that instead attempts to build an understanding of the world through *data* and analytical methods[2].

At this stage it is worth diving into what World and Person are. World represents an (or any, or all) *aspect* of the world we live in

[1] More formally these are examples of *functors* between *categories*, but more on this later.

[2] Not to say that the analytical way is somehow more true than the experiential way - the end result is an interpretation of the world.

that we are trying to understand. In a way this only makes sense if we introduce a particular person, which is another *aspect* (this time of the Person structure[3]) who is trying to make sense of some aspect of the world.

For example, a parent might be trying to understand why the later it gets the more energy their four year old child seems to have, seemingly contradicting the first law of thermodynamics. Here the aspect of World could be described as *energy levels of small children approaching bedtime*, the aspect of Person is a parent. This is puzzling because they have an analytical way to make sense of this phenomena (a mapping from World to Person using the law of conservation of energy in a closed system) and this does not seem to align with their experiential understanding of the aspect (said child is now doing laps of the dining table whilst singing, despite having not eaten dinner).

[3] Later this will be defined as a Category, which has a precise mathematical meaning, but for now the word *structure* lends itself better to developing the correct intuition.

The composition of analytical sense making

We can pull apart the analytical sense making map, $\aleph_A$, into a series of steps.

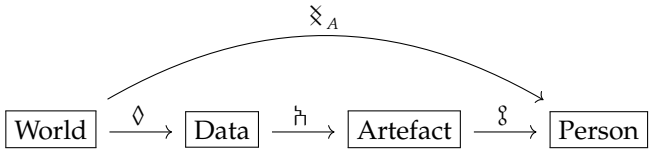

Figure 3: Decomposing the analytic sense-making, $\aleph_A$, into three steps; the *framing* $\lozenge$, the *analysis* $\beth$, and the *interpretation* $\gimel$.

We have already loosely defined the spaces World and Person, so now let us work through the process and talk about the new maps and spaces.

$\lozenge$: Framing

The first step in analytical sense-making we shall call the *framing* and takes us from some aspect of the world we are interested in to a data set. This map is denoted using the symbol *fe*, $\lozenge$ (think of a picture frame[4]). This is a subjective map from the 'real' world to the world of symbols and information. It should be apparent that this is critical to the final understanding of the phenomena of study, for instance if one decides not to include *race* as variable of the data set then it would be hard to find any effects of racial discrimination.

[4] On an angle and slightly squashed. I just did not want to use a box.

Suppose some teacher is interested in how well a group of students have understood the content of their course.[5] Some possible

[5] It would be good to reference the measurement sciences here, or to state that this is an example taken from that field directly.

options for $\lozenge$ include:

- Setting an examination on the topic. The choice of questions here is an epistemic cut in itself; open or closed questions, multiple choice, essay with an unseen or seen question.
- Conducting interviews / discussions with the students, possibly in groups or as a whole.
- Having students reflect on their learning in a journal. This could be marked according to some rubric or may just be read by the teacher, with no mark assigned.
- Observing the students as they learn and discuss the topic.

Note that the framing may involve actions to produce the data; for instance notes would probably be taken from the observation of the students as they learn in order to produce data to analyse.

Data

The Data space (or structure[6]) contains anything that can be recorded as data. To impose further structure on this imagine that the object that is $\lozenge$ maps to is a data frame, a rectangular matrix whose columns hold data of all the same type, including the possibility of a column of data frames. This captures the straight forward cases of records in a spreadsheet to more complicated data such as a collection of audio and video recordings. The chief requirement here is that the data must be *recorded* somewhere, and this requires a decision of what is recorded, and that decision is encapsulated in the framing, $\lozenge$.

[6] We shall use *space* and *structure* interchangeably for these concepts depending on the context, as both words hold true for the things we are trying to describe. Sometimes it is simpler to think of them as an amorphous space holding the ideas we want. Sometimes it is better to think of how the parts of this space relate to each other and how this space is structured and behaves.

ה : Analysis

The second step in analytical sense-making we shall call *analysis* and takes us from the data set to some kind of artefact. This map is denoted using the symbol *alef*, ה (think of exploring the data along different paths to find the best representation).

Imagine we have a set of examination results as out data, which we can view as an $n \times 2$ matrix of $(student, score)$ pairs. Some possible options for ה include:

- Listing the students whose *score* is below a certain threshold (say 0.5).

- Taking the mean of the *score* values.
- Plotting the distribution of the *score* values, and marking the mean and median on the plot.

Artefact

The Artefact space contains anything that can be the result of an analysis, that is used to communicate ideas to some other person. This artefact could be a graph, a report, a number, an email, a linear model - anything that is used to convey the results of the analysis.

$\wr$: Interpretation

The third and final step in analytical sense-making we shall call *interpretation* and this takes us from the artefact to a person's understanding of what the artefact is saying about this aspect of the world. This map is denoted using the symbol *thaw*, $\wr$ (think of a pair of neurons connected by some new pathway). This map is part cognitive and biological and dependent on many external factors; the role of the observer, the type of artefact, the ...

Expanding on the examples previously, the interpretation of a set of "how well did my students understand the topic" is quite different from the point of view of different end users. The teacher might want to inform pedagogy, a director of learning might want to assess performance of subjects, a pastoral leader might want to identify under performing students, and at a higher level we might be interested in ensuring students pass the course. So the person at the end of the process usually has a *purpose* for looking at the artefact, but more often than not they will also be seeking to confirm some existing belief[7] or at least carry some pre-existing notions about the analysis in question. All of this influence what inferences are made about the analysis from the artefact.

[7] This surely must have some backing somewhere in the literature. Where?

The theory - practitioner gap

In Figure 3 it is assumed that the two paths from World to Person are equivalent, which means the diagram *commutes*. This may not always be the case, particularly when order is important.

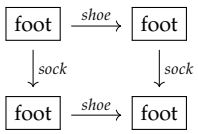

Figure 4: Different paths sometimes lead to different places

One could argue that life is interesting precisely when the paths $\aleph_E$ and $\aleph_A$ do *not* form a commutative diagram between World and Person; that is when the analytical sense making does not quite agree with the experiential sense making of the world. A subset of this problem is known as the theory - practitioner gap[8] which amounts to when the analytical or research view of the world does not consistently agree with the experiential view of the world. We articulate this question diagrammatically by Figure 5.

[8] Need reference

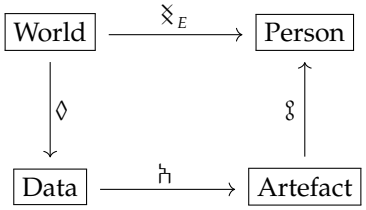

Figure 5: Two different paths for a person understanding an of the world; an analysis process broken into three steps ($\aleph_A = \beth\daleth\lozenge$) or through interpreting experiences ($\aleph_E$).

So we have begun to break up the problem of the theory - practice gap into a fledgling mathematical structure.

Scrapbook

Figures, tables, less cohesive bits to keep.

Thoughts

THE EPISTEMIC CUT[9] still needs to be designed and placed within the system, but seems to permeate the whole process. Might be best viewed in parallel and examining how it influences the process. Exactly how it connects with the structure is an issue (see below in *mathematics* section).

[9] H.H. Pattee. The physics of symbols: Bridging the epistemic cut. *Biosystems*, pages 5–21, 2001

OVERLAP WITH OTHER FIELDS will need to be addressed. Whatever framework is put in place should cover existing work. A notable example is the field of measurement science; any model applied here will need to overlap nicely with applying this to the process of assessing a student, for instance. [10]

[10] Sandra K. Milligan. Methodological foundations for the measurement of learning in learning analytics. 2018

Mathematics

CONTEXT AS A SHEAF above the sense-making process, the epistemic cuts mapping down to it. Possibly as natural transformations, or, where the naturality is not possible this indicates a complex system - a complex system being one that cannot be defined in a single canonical way.[11]

[11] Check with Kirsty the proper definition of this and where it actually came from

COMPLEXITY OF THE SPACES could be framed in the separation schema of topological spaces (see the Wikipedia entry on Separation Axioms) developed initially by Tychonoff.[12] The metrizability of spaces could also be a key indicator of complexity. The other notion (appearing somewhere on Twitter, as well as Jules Hedges talk at ACT2020) is the idea of 'cohomology as complexity'. Coherence properties of categories might be another path to follow - basically looking at how 'well behaved' the categories are.

[12] Andrey. N. Tychonoff. Über die topologische erweiterung von räumen. *Mathematische Annalen*, pages 544—-561, 1930

COMPOSITIONAL GAME THEORY story shows a nice way to develop these ideas. It starts with work on a compositional approach to machine learning[13], the categories of open learners, and builds into a broader approach to economic game theory.[14]

[13] Brendan Fong, David I. Spivak, and Rémy Tuyéras. Backprop as functor: A compositional perspective on supervised learning. 2017

[14] Jules Hedges. From open learners to open games, 2019

Philosophy

Process philosophy might have something useful to say - advocates process over spatial understanding of phenomena. See ep 144 of *Philosophize This!* about Max Weber, but the next episode should introduce more. It might also be worth checking what Category Theory can do for Philosophy because it seems that the intersection of ideas here is pretty evident, and how this is being viewed towards different fields might provide some steps forward.

Gilles Deleuze and Felix Guattari have some interesting views about ontology itself, the rhizome as apposed to the tree of knowledge. Easily digested in the *Philosophise This!* podcast episodes sourced on the works *Capitalism and Schizophrenia* and *What is Philosophy?* They have many viewpoints that might be interesting to apply, but the non-hierarchical view of ontology (or non-transcendent) that they call immanence has some genuine flow through effects as to how one implements LA. For instance, the transcendental view of measuring something like learning assumes that there is some 'true' value out there that you are trying to get close to, however Deleuze and Guattari would argue that these phenomena are purely emergent, from this plane of immanence. There approach to ontology (?) is also interesting, rather than focusin on the individual that conceptualise *machines* and then *flows* between *poles* - all sounding very category theoretic.

Walter Lippman had some interesting ideas about how public opinion is formed[15]. He said that opinions are formed by individuals constructing stereotypes of the world, and the crystalisation of these from the masses forms public opinion. This sounds much like some kind of functor or transformation or embedding. Also introduced in this podcast and a more recent paper is available here.[16]

[15] W. Lippmann. *Public Opinion*. Macmillan, 1922

[16] W. P. Bottom and D. T. Kong. "the casual cruelty of our prejudices": On walter lippmann's theory of stereotype and its "obliteration" in psychology and social science. *Journal of the History of the Behavioral Sciences*, pages 363—-394, 2012

Examples to play with at a later date

Example of conflict between experiential and analytical - inserting a USB drive.

Astrological predictions / star signs.

ENA

ML of some sort

Building a linear model

Bias. Such as selection bias - Tony always thinks that he has to wait for a car turning into Felton road because he already believes that, and remembers each time it happens, and forgets each time it doesn't. How would this pan out if a proper analysis and data was recorded.

p-hacking. Does that work as a way to analyse where this process goes wrong? Maybe use this article as a starting point for other ways statistics is done wrong (or the book of the same name).

Dictionary of Symbols

These are constructed using south Arabian letters, using SARABIAN package. The name of the letter precedes what they are used for.

𐩧 : *Samekh* : SENSE MAKING : Think of weaving together ideas.
𐩦 : *Fe* : FRAMING : Think of a picture frame. On an angle. I just didn't want to use a box.
𐩡 : *Alef* : ANALYSIS : Think of exploring the data to find the best representation.
𐩸 : *Thaw* : INTERPRETATION : Think of some neurons building a pathway.

References

[1] W. P. Bottom and D. T. Kong. "the casual cruelty of our prejudices": On walter lippmann's theory of stereotype and its "obliteration" in psychology and social science. *Journal of the History of the Behavioral Sciences*, pages 363—-394, 2012.

[2] Brendan Fong, David I. Spivak, and Rémy Tuyéras. Backprop as functor: A compositional perspective on supervised learning. 2017.

[3] Jules Hedges. From open learners to open games, 2019.

[4] W. Lippmann. *Public Opinion*. Macmillan, 1922.

[5] Sandra K. Milligan. Methodological foundations for the measurement of learning in learning analytics. 2018.

[6] H.H. Pattee. The physics of symbols: Bridging the epistemic cut. *Biosystems*, pages 5–21, 2001.

[7] Andrey. N. Tychonoff. Über die topologische erweiterung von räumen. *Mathematische Annalen*, pages 544—-561, 1930.