Nguyen Learning Analytics Taxonomy Synopsis

Created for the Spring 2018 Blended Learning Fellowship on Evidence-Based Teaching & Learning Analytics

Source

Nguyen, A., Gardner, L., & Sheridan, D. (2017).
A multi-layered taxonomy of learning analytics applications.
In Pacific Asia conference on information systems. Retrieved from http://aisel.aisnet.org/pacis2017

Overview

Nguyen, Gardner, and Sheridan (2017) created a multi-layer taxonomy of learning analytics applications. The taxonomy was created using machine learning techniques which forced classification of an application until mutual exclusivity and collective exhaustion was reached; that is, every application is classed in only a single category and at least one application lives in every category. Nguyen et al. used the proceeding of the Learning Analytics & Knowledge (LAK) Conference as well as *The Journal of Learning Analytics* as their source for LA applications. The proposed taxonomy is offered as a proof-of-concept, and was designed to be explanatory and extensible.

It is important to note that Nguyen et al.'s taxonomy was established to answer the question, "why are people using learning analytics?" This taxonomy only evaluates the purpose of a LA application, it does not consider larger meta- issues such as policy, ethics, etc.

Classification Layers

Nguyen et al.'s taxonomy consists of a seven step iterative method. This means each application must go through seven layers of evaluation before it is classified. Once an application has been evaluated at each of the seven layers (think component or characteristics) determination conditions are assigned to each application. In other words, once an application goes through the evaluation stage, a category is assigned that, in machine learning terms, correctly attributes various characteristics to the application ultimately allowing a "purpose" label to be assigned. The seven layers are summarized below.

Layer 1: Main Purpose

The main purposes layer captures the objective of the LA application. There are three potential outcomes in this category:

1. **Learner-centric:** the application of the analytics is specifically on an individual as a learner.
2. **Event-centric:** the main focus of the analysis on the interactions of a learner.
3. **Content-centric:** the primary emphasis of the application is on the curriculum, course content, or materials.

Layer 2: Feedback Type

The feedback type classifier focuses on the time based element of provided feedback. There are three possible outcomes:

1. **Retrospective/Reflective**: presenting historic data analysis.
2. **Real-time/Adaptive**: presenting real-time data analysis.
3. **Prospective/Predictive**: presenting predicted future state or events.

Layer 3: Data State

The data state layer describes the "physical" state of the data. In today's technologically advanced ecosystems, systems are designed to anticipate the speed at which data may need to be accessed. The associated state then, describes how frequently data is updated in warehouses/data stores. This taxonomy required has three relevant data states:

1. **Static**: data that rarely changes, so frequency of updates is rare. Static data rarely changes over time. For example, GPA because, generally, GPA only changes once a term. Demographic data is another example of static data.
2. **Semi-dynamic**: data that changes at more regular intervals, but are not necessary in real-time. A very common example of this is data in standard educational technology ecosystems. Standard protocol for LMS data refreshes tend to be 24 hour refresh. In the LA realm, the most notable type of semi-dynamic data is log data.
3. **Dynamic**: data that is updated in very short periods of time, typically seconds. Operationally, dynamic data is synonymous with real-time applications. Dynamic data differs from semi-dynamic not necessarily in type of data (typically log/interaction data), but rather the desired frequency of access. For example, just-in-time intervention applications such as scaffolds for quiz items need to access data immediately to be useful, rather than being able to wait for a standard 24-hour refresh update.

Layer 4: Data Source

Quite simply, the data source layer describes where the data for the LA application comes from. While many LA applications have multiple data sources, this taxonomy focuses only on primary data source. Since this classification exercise used LAK proceedings and the *Journal of Learning Analytics* as the source of application, there is notable bias in data sources, but remember that this taxonomy requires exclusivity in determination. The three requisite categories in this layer are:

1. **LMS**: the learning management system (virtual learning environment – VLEs—as they are commonly referred to in Europe and Australia). Typically this is a central system such as BlackBoard, D2L, Moodle, or Canvas.
2. **Social Network**: a large majority of LA work that has theoretical underpinnings of a social nature will leverage social networks such as Twitter, Facebook, or home grown social applications.
3. **Other**: any other data sources. While not as common, LA initiates often have a primary focus of data not related to EdTech data sources. For example, advising applications tend to focus on information from student information system.

Layer 5: Stakeholder

In the context of this taxonomy, Nguyen et al. use stakeholder as a means to express what we commonly refer to as lens/perspective. In other words, the stakeholder is the group of people that either directly use or gains direct benefit from the application. Since LA is grounded in contextual work about the learning environment, the three main stakeholder groups are: teachers, learners, and researchers. For practical purposes, this classification layer collapsed three main stakeholder groups (teacher, learner, and researcher) onto a theoretically driven framework widely used in the LA domain.

1. **Micro:** learner
2. **Meso:** teacher
3. **Macro:** institutional
4. **Mega:** researchers (construed as work "for the greater educational good")

Layer 6: Expertise Requirement

The expertise requirement layer was added to the taxonomy to address the operationalization of the application itself. Simply put, this layer categories the application into broad buckets describing the expertise necessary to run the application. The outcomes are:

1. **Novice**
2. **Intermediate**
3. **Advanced**
4. **Expert**

While these labels are subjective to a degree, they are derived based on the sophistication and complexity of the methods; this classifier is specifically about the level of requisite knowledge and skills needed to successful operationalize the application.

Layer 7: Operating Complexity

Similar to the expertise requirement, the operating complexity layer is a simple classification which gauges the level of effort necessary to use the application. This classifier takes into account level of automation available. The three levels are:

1. **Low**
2. **Medium**
3. **High**

As in the expertise requirement layer, there is a fair level of subjectivity in the classifications of the operating complexity layer.

Final Categorization

Once a LA application has a determination for each of the seven layers in the taxonomy, a final classification can be assigned. The six final categories from Nguyen et al.'s taxonomy that we will focus on for the fellowship are empirically derived from scholarly learning analytics literature. Six application purposes were established (*all adopted from Nguyen, Gardner, and Sheridan, 2017):

1. **Visualize Learning Activities**: Learning analytics traces all learning activities performed by users in a digital ecosystem to produce visual reports on the learning process. The reports can support both students and teachers to boost learning motivation, adjust practices and leverage learning efficiency.

2. **Access Learning Behavior**: Learning analytics can be used to collect user-generated data from learning activities and offer trends of learning engagement. Analyzing those trends can discover learning behavior of the students and identify their learning styles.

3. **Predict Student Performance:** There have been several attempts using learning analytics to predict student's success and identify at-risk students. Based on existing data about learning engagement and performance, learning analytics apply statistical models and machine learning techniques to predict later learning performance. By doing so, likely at-risk students can be spotted out for early intervention.

4. **Individualize Learning:** Adaptive or individualized learning systems apply learning analytics to consume a relatively small user-generated data to adjust its content for each learner. Furthermore, user profiles and other sets of data can be collected and analyzed to offer greater personalized learning experiences.

5. **Evaluate Social Learning:** Not limited to the assessment of formal learning on the LMS, learning analytics can be applied to investigate learner's activities on social networks to evaluate the benefits of social learning.

6. **Improve Learning Materials and Tools**: Learning analytics can track student's usage of learning materials and tools to identify potential issues on those. LA can also offer objective evaluation of learning materials and tools.

Nguyen, Gardner, and Sheridan's Multi-layered Taxonomy of Learning Analytics

Now that the individual layers have been explained, and the final application categories defined, the taxonomy can be explored. The application's assigned value at each of the seven layers *should*, in theory, classify the overall purpose of the LA application. As a reminder, each of these categories have been generated for exclusivity. So the initial training and test data set confirms this taxonomy. However, it is important to note that this taxonomy's extensibility has not been tested. For the purposes of the EBT BLFP we have adopted Nguyen et al.'s taxonomy for its explanatory power. Our discussion will also do a low tech validity check of its extensibility.

We have partially adopted the taxonomy and added sample applications and research. It is this document we will spend the bulk of our time grappling with at next week's session. We focused on only six of the nine categories for ease of discussion; the categories selected are those that align with the lines of inquiry we have been building collaboratively in the fellowship.