|  |
| --- |
| Literature Review |
| By Mark Brackenrig  Due: 18th August  Time: 5:00 pm |

Executive Summary 3

Article Notes 4

Upgrading the K-12 Journey through Curriculum Mapping 4

Mapping or tracing? Rethinking curriculum mapping in higher education 4

Indexing by Latent Semantic Analysis 4

Method 5

Implementation Example 5

A cognitive processing framework for Learning Analytics 5

Comparative study of LSA vs Word2Vec embedding in small corpora: a case study in dreams database 5

LSA 6

Word2Vec 6

Differences 6

Chapter 1: Theory and Learning Analytics (Handbook of Learning Analytics) 6

Chapter 2: Computational Methods for the Analysis of Learning and Knowledge Building Communities 6

Chapter 7: Content Analytics: The Definition, Scope, and an Overview of Published Research 7

Chapter 8: Natural Language Processing and Learning Analytics 7

# Body

Text analysis techniques are widely used in the field of Learning Analytics (LA). (McNamara, et al., n.d.) suggests that language is of ‘central importance’ to education because it facilitates interactions between learners and educators. There is a large body of literature around analysing student-teacher interactions, student to student interactions and assessment. Much of this literature aims to predict student outcomes, such as failing a class or marking an assessment without human involvement. However, a much smaller body of research has been conducted into

# Article Notes

## Upgrading the K-12 Journey through Curriculum Mapping

(Jacobs, 2000)

* **Curriculum mapping:** Is a tool for gathering data on what teachers are actually working on with their students through the course of the school year. When a school undertakes curriculum mapping, all teachers at the school enter information about their classroom curricula into a database that is geared to the school calendar. Teachers enter three types of data:
  + **Content –** key concepts, essential question, etc.
  + **Specific skills –** often based on standards
  + **Assessments –** tests, products or performances.
* The article identified these key uses of curriculum maps:
  + **Gain information** – Allows teachers to understand what their students are learning in other classes and what they have learnt before.
  + **Identify Gaps** – There is an assumption that all the committee outlines are accurate and that all teachers are addressing the stated curriculum. This would provide educators the ability to target learning at gaps in knowledge.
  + **Identify Repetitions** – Often there are too many repeated units throughout a K-12 journey. Educators may not be aware that a concept, process, or book has already been introduced to students in another course. Differentiating between spiralling and repetition is a critical task that curriculum mapping addresses.
  + **Identify potential areas for integration** – Identifying overlap between curriculum of two or more disciplines can present a powerful opportunity for transdisciplinary unit design or concurrent teaching.
  + **Align assessments with standards** – By matching desired learning outcomes with evidence of student learning, you can measure educational effectiveness.

## Indexing by Latent Semantic Analysis

(Dumais, et al., 1990)

This paper describes a ‘new’ method of automatic indexing and retrieval of text documents. This methodology (LSA), takes advantage of the higher-order structure in the association of terms with documents (“semantic structure”) in order to improve the detection of relevant documents on the basis of terms found in queries.

Essentially this method tackles two main problems in text classification and retrieval:

* Synonymy: A document may be relevant to a query but does not contain any words used in the query. This is because humans may use many different words to describe the same concept.
* Polysemy: A document may NOT be relevant to a query but does contain words used in the query. This is because the same word may have multiple meanings in different contexts (i.e. “Chip”).

### Method

*I need help in the methodology section – further reading required*

LSA uses Singular Value Decomposition over a vector space of both terms and documents in order to identify an underlying semantic structure. From my understanding, this is similar to dimension reduction techniques (such as factor analysis) to identify underlying structures in the data – similar to how you would interpret factors in factor analysis as ‘latent variables’. However, in Latent Semantic Indexing, you are less concerned with the ability to ‘understand’ these underlying structures, but rather wish to attribute terms and documents into these underlying structures for their retrieval. In the indexing context, this means that you would return all documents and terms similar to your query based on proximity of the created vector space.

### Implementation Example

*Find a python implementation of LSA*

<https://github.com/llazzaro/lsa_python/blob/master/lsa/lsa.py> - Still need to review

<https://github.com/cran/lsa> - in R if I start crying by looking at python

## A cognitive processing framework for Learning Analytics

(Gibson, et al., 2014)

This paper describes the application of a learning analytics framework called “COPA”, in which the application of Bloom’s Revised Taxonomy to learning outcomes and assessment criteria is useful in understanding the levels of cognitive operation. Essentially, this paper describes a process of curriculum mapping that extends beyond understanding whether the learner has achieved learning outcomes, but describes the level of cognition that the learner has to operate at to achieve these outcomes.

This method has been shown to be a useful tool in identifying a mismatch between assessment criteria and stated learning outcomes (CLOs), after applying COPA to an undergraduate science unit. This proof of concept showed that only a small proportion of the CLOs required high levels of cognition (creating): 4.2%, yet accounted for 20.3% of the assessment.

* **Learning analytics (LA)** is understood to be “measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs"
* **Bloom’s Revised Taxonomy:** Bloom’s Revised Taxonomy describes a hierarchy of learning outcomes from basic recall to the ability to apply a particular learning outcome. The taxonomy has six levels:
  + Remembering
  + Understanding
  + Applying
  + Analysing
  + Evaluating
  + Creating

## Comparative study of LSA vs Word2Vec embedding in small corpora: a case study in dreams database

(Altszyler, et al., 2017)

This paper is a comparative study between two types of text analysis techniques: Latent Semantic Analysis and Word2Vec. The results of the study showed that while much literature found that Word2Vec performed better in semantic categorisation, this was done using large corpora (i.e. a large number of documents), and LSA performed better at these tasks with small corpora. The author suggested that this was due to Word2Vec being a prediction based model, where as LSA is a ‘counter-based’ model. The discussion of the distinct advantages and disadvantages of each technique are as follows:

### LSA

* Counter-based model that considers the entire corpus.
* Uses a bag of words model.
* Typically uses normalization to reduce weight of uninformative high-frequency words
* Applies SVD (dimension reduction) which projects every word into a pre-defined number of dimensions.

### Word2Vec

* Prediction based method that considers parts of the corpus at a time.
* Variations: Continuous bag of words and skip-gram.
* A CBOW model will move through the corpus and attempt to predict the word in the centre of the window of the corpus based on the surrounding words
* A skip-gram model is trained to predict the contexts based on the central word.
* Uses neural networks

### Differences

* Word2Vec uses less memory during training – especially in large corpora since it doesn’t consider everything in the training corpus at once.
* Word2Vec does not perform well in small corpora since it is trying to fit an extremely large number of parameters.

## Chapter 1: Theory and Learning Analytics (Handbook of Learning Analytics)

(Knight & Buckingham Shum, n.d.)

This paper acts as a high level overview of the purpose and theory of learning analytics, and how LA differs from related education fields. The paper raises important points about the relationship between pedagogy, epistemology, and assessment, and that the use of learning analytics is effectively grounded in assessment. The paper outlines a series of questions designed to assist in the development and design of learning analytics systems:

* What are we trying to measure?
* How are we measuring?
* Who is the assessment/Analytic for?
* Why is the knowledge important to us?
* Where does the assessment happen?
* When does the assessment and feedback occur?

These questions structure the paper (and are broadly categorised into pedagogy, epistemology, and assessment) and ask important questions when designing analytic systems. The paper argue that, broadly, learning analytics tools are grounded in assessment regimes, which in turn, relies on epistemological assumptions and pedagogical practices.

## Chapter 2: Computational Methods for the Analysis of Learning and Knowledge Building Communities

(Hoppe, n.d.)

## Chapter 8: Natural Language Processing and Learning Analytics

(McNamara, et al., n.d.)

This paper act as an overview to NLP techniques in Learning Analytics. The purpose for this paper is to discuss specific tools and techniques specific to LA. The main reason for the study of language in Learning Analytics is because it is the conduit for communicating and understanding information. The paper outlines multiple tools and techniques and categorises these techniques into a few main categories:

* ‘The Words’: Techniques that analyse the words used in the language directly. This encompasses ‘bag-of-words’ approaches.
* ‘Word features’: techniques that consider the features of words and sentences in the text. One such example is LSA, as LSA looks to analyse the ‘meaning’ of the word.
* ‘Sentence features’: Techniques that analyse sentence groupings, for example, looking at the complexity of a sentence.
* ‘Text features’: Techniques that analyse text features, such as cohesion, sophistication, etc.

The paper looked at many applications for the use of text mining in Learning Analytics:

* Writing assessment – Applications such as automated essay scoring
* Tutoring systems – NLP systems allow students to interact with a computer and receive feedback.
* Computer Supported Collaborative Learning
* Massive Open Online Courses

## Curriculum Mapping to Embed Graduate Capabilities

(Spencer, et al., 2012)

This paper discusses a ‘manual’ methodology to map generic graduate attributes across different first year subjects in Law and Management degrees. The generic graduate attributes are attributes that all graduates at La Trobe university should have. The methodology was as follows:

* Initial consultations with discipline academics
* Defining graduate attributes
* Document-based data collection
* ‘Coffee meetings’
* Data analysis (quantitative survey)
* Heat maps (visualisation)

This process was very manual, requiring expert mappings between curriculum documents to agreed upon graduate capabilities, and confirmation from subject coordinators regarding what was actually taught.

## Using Contexts for Object Similarity: Exploratory Investigations

(Niemann, et al., 2011)

This was a short technical paper outlining the use of text mining techniques for educational content that are not text based (such as images, audio and videos). The premise that context is significant for the meaning of a word: “You shall know a word by the company it keeps”, may also apply to objects other than words. In the LA context, this means that objects used together in the same session for searching for resources are most likely related. The authors suggest that this can be used to generate a recommender tool that assists students in searching for relevant material.