A Comprehensive Methodology for Concept Map Assessment

Carlos T. Calafate, Juan-Carlos Cano, Pietro Manzoni
Universidad Politécnica de Valencia
Camino de Vera, S/N, Valencia, Spain
E-mail: {calafate, jucano, pmanzoni}@disca.upv.es

Abstract

Concept maps have been around for quite some time, and their principles are deeply rooted on well-known learning theories. When used in the evaluation process as a tool to assess learning they have obvious benefits by allowing students to externalize their own mental trees of assimilated concepts seamlessly. However, the assessment of concept maps includes a strong degree of subjectiveness, which should be mitigated. In this paper we propose partitioning the concept map evaluation process according to the steps followed for creating them, along with objective metrics to assign a score to each of these steps. We also propose a formula that combines the partial scores to obtain the final score. Afterward we validate the proposed methodology, showing that the score variability associated with the evaluator is reduced by up to 23%.

1. Introduction

Concept maps were first proposed by Novak [1] a few decades ago. They consist of a very flexible structure that allows creating meaningful relationships between key concepts so as to improve the processes of learning and knowledge acquisition in general. Since their invention, they have received much attention from researchers and experts worldwide, being used in various areas of knowledge [2-4].

Evaluating students using concept maps is quite interesting since it allows them to externalize their own mental trees of assimilated concepts to a simple format, thereby offering the instructor an Ausubelian perspective of the knowledge acquired [5]. Adapting such a student-centered evaluation strategy also helps at detecting misunderstandings and flaws in the students’ learning process.

One of the main drawbacks when evaluating concept maps has to do with the subjectiveness in associating a score with one particular concept map. In fact, different instructors evaluating a same set of concept maps are prone to assign different scores within a relatively wide range of values.

One of the first works in the field of concept map evaluation was that of Cronin et al. [6] in 1982. More recently, works such as [7-8] have proposed an evaluation strategy based on the structural and relational evaluation models. From our perspective, the main drawback of these solutions is that evaluating concept maps of very heterogeneous sizes is not straightforward; additionally, defining a Master-map (see [7] for a definition and more details) is not always appropriate. Finally, notice that these proposals are validated with a set of only two evaluators.

In this paper we introduce a technique that takes into account the different steps taken when creating a concept map, defining a set of criteria and heuristics to evaluate each of these steps. Finally, we proceed to validate our proposal using concept maps developed by our students.

This paper is organized as follows: in the next section we described our systematic technique to evaluate concept maps avoiding high result variability. Validation of our proposal is done in section 3, including the main conclusions drawn from our work.

2. Objective metrics for concept map assessment

An important characteristic of concept maps is that they tend to be unique for every student. However, uniqueness also prevents the instructor from doing a quick evaluation of their work since the object of evaluation is not right or wrong, being more complex, elaborate and precise in direct relation to the
interiorization and understanding of the students about the concepts being addressed. Therefore, the evaluation process is prone to be complex, time-consuming and, in general, quite subjective.

In this section we derive a set of metrics that allow making the evaluation process much more objective, as well as reducing the variability due to the different factors involved. This is achieved by splitting the evaluation process into different steps, and defining a set of criteria and heuristics to assign a score to each of these steps. The proposed fragmentation of the evaluation process takes into consideration the different parts involved in creating concept maps themselves. We then derive a formula that aggregates the individual scores in a comprehensive manner.

### 2.1. Number and significance of the concepts

The first task when creating a concept map that describes, e.g., a course unit, is to discriminate the most meaningful concepts from the rest so as to make them the skeleton of the concept map created. Thus, the first step taken by the instructor should be to determine which are the most essential concepts (those that the student should not obviate). From now on, this number will be denoted as $N$.

When analyzing the concept map developed by an individual student, the instructor should first count the total number of essential ($n_e$) and secondary ($n_s$) concepts involved. Afterward, we propose using the following equation to derive a first score with regard to concepts:

$$ S_i = \frac{n_e}{N} \log_N (n_e + n_s) \quad (1) $$

This formula allows combining the student’s effectiveness in detecting essential concepts with the total size of the concept map developed.

It is also important to define a maximum value ($M_i$) according to the evaluator’s criterion on what should be the maximum ratio between $n_e$ and $n_s$ to achieve the maximum score. For example, if we consider that the ratio between $n_e$ and $n_s$ should by of 4 to 1, the value of $M_i$ would become: $M_i = \log_N (N + 4 \times N)$.

### 2.2. Degree of meshness and relationship accuracy

The second step when creating a concept map is to relate the different concepts involved to construct a meaningful information structure. Usually a linear construction with no loops and minimal concept linking is considered as a poor construction, lacking the richness expected from well-designed concept maps. However, not all concepts have relationship will other concepts, and so the optimum number of relationships is only slightly higher than the minimum indispensable value ($R_{min}$).

Thus, we propose a metric to assess the degree of meshness ($DM$) of the concept map that compares the total number of relationships ($r$) against the minimum number possible ($R_{min}$) in the following manner:

$$ DM = \frac{r}{R_{min}} = \frac{r}{n-1} \quad (3) $$

Mapping DM values to actual scores requires analyzing the characteristics of typical concept maps. Taking the ones developed by our students into consideration, we propose the following meshness score ($MS$) assignments according to the $DM$ value obtained:

$$ MS = \begin{cases} 
0.7 & DM < 1.04 \\
0.85 & 1.4 \leq DM < 1.08 \\
1 & DM \geq 1.08 
\end{cases} \quad (4) $$

The proposed score assignments penalize by 15% maps with low degree of meshness (less than 1.08), and by 30% maps with a very low degree of meshness (less than 1.04).

Besides the degree of meshness ($DM$), which is a strictly objective metric, a qualitative evaluation of the relationships proposed must also be introduced. This new metric, which we denote as Relationship Accuracy ($RA$), is a score between zero and one assigned by the evaluator where he/she subjectively evaluates the overall correctness and accuracy of the relationships proposed. So, while the $DM$ metric attempts to detect whether the amount of relationships between concepts allows for a strong bonding between them, the $RA$ metric reflects the coherence of such relationships.

### 2.3. Other quality factors

Besides the characteristics defined above, there are other quality details that illustrate the student’s dedication and interest when creating the concept map, including segregating the most important concepts from the rest through highlighting (font, color, box shape, etc.), including representative figures or icons, and linking to outside elements such as web pages, applications or even other concept maps.

Depending on context, the evaluator should thus establish some basic criteria for determining the
quality parameter \((Q)\), which takes this extra effort into account by assigning a score between zero and one.

2.4. Proposed formula for objective evaluation

Up to now we have described a set of metrics that are related to each of the three steps involved in creating a concept map: concept definition, relationship definition, and the introduction of auxiliary information (highlighting, figures, links, etc.). Based on those metrics, we now propose a final formula that seeks to integrate these different metrics in order to assign the student a final score. The proposed formula is the following:

\[
Score(\%) = \alpha \cdot \frac{S}{M_1} + \beta \cdot MS \times RA + \chi \cdot Q \quad (5)
\]

where \(\alpha\), \(\beta\) and \(\chi\) are weights between zero and one assigned to each of the three components of the evaluation. Obviously, these three weights should add up to one, and in general we recommend evenly splitting most of the weight among the first two factors, and providing a residual weight \((\chi)\) to the quality parameter depending on the topic and requirements. In particular, for our case study we will use values: \((\alpha, \beta, \chi) = (0.6, 0.35, 0.05)\).

3. Validation and conclusions

To validate the proposed assessment technique we used concept maps developed by students in a Computer Networks course during the two previous academic years (i.e., 2006-07 and 2007-08). These concept maps were then handed-over to five different instructors for evaluation using both the traditional method and the proposed technique.

Table 1 shows a summary of results that highlight the improvements achieved at reducing the differences between different evaluators in terms of score when the proposed objective metrics are used.

In terms of average score the differences are low (about 6%); the trend towards lower values achieved with the objective metrics is expected since more strict criteria tend to decrease the scores. With respect to the median, though, its value remains practically unaltered.

In terms of standard deviation, the objective metrics proposed offer a 45% reduction for this parameter. In terms of min-max differences, we see that the average values have been reduced by 47%, while the top min-max difference was reduced by 2.24 points over 10 down to 2.14. This means that, for all the concept maps evaluated, the differences between two evaluators in terms of score for a particular concept map was never greater than 2.14 points over 10.

Overall, we found that the objective metrics proposed achieve the desired goal of reducing differences between evaluators, while not causing the mean and median scores to vary significantly.

Despite the improvements achieved are evident, we consider that further refinements of this evaluation strategy should be developed to reduce differences between evaluators to a minimum.

References