A KNOWLEDGE METRIC
WITH APPLICATIONS TO LEARNING ASSESSMENT

Rafik Braham
PRINCE Research Group, College of ICT, University of Sousse, Sousse, Tunisia

Keywords: Information Theory, Similarity Measure, Knowledge, Ontology, Learning Assessment.

Abstract: We present a framework within which Knowledge is decomposed into basic elements called knowlets so that it can be quantified. Knowledge becomes then a measurable quantity in very much the same way data and information are known to be measurable quantities. An appropriate metric is thus defined and used in the specific domain of learning assessment. The proposed framework may be utilized for Knowledge acquisition in the context of ontology learning and population.

1 INTRODUCTION

Students learning assessment in the context of e-learning has been the focus of attention of several research studies for the last few years. A number of assessment environments have been developed (Gardner 2002, He 2006). Some related standards have also emerged, for example the Question-and-Test-Interoperability from IMS, better known as IMS-QTI (IMS 2006). Several web sites offer now several tools for the generation of assessment material. Hot Potatoes (from the University of Victoria) is a well known software tool employed in the generation of tests especially those of the MCQ category. In pursuing research in this field (Cheniti-Belcadhi et al. 2004, 2008), we have been intrigued by a fundamental question concerning students' assessment: "how much they know?" Other questions relative to assessment and testing have been raised, for example "how we know they know?" (Palloff and Pratt 2006), but to the best of our knowledge, the question we ask has not been dealt with. That is precisely our objective in this paper.

We define the problem (Knowledge acquisition) by the following algorithm.

1. Consider a system (a certain KB for example) which contains at some time t>0 an amount of stored Knowledge denoted by $K_S(t)$ (according to some positive metric to be defined later). We assume $K_S(t=0) = 0$.

2. At time t'>t, some "Knowledge" denoted by K_{in} is presented to the system.

3. The system will compare K_{in} to K_S. Only the part of K_{in} that is novel (with respect to K_S) shall be stored. Since the Knowledge increment is greater than or equal to zero, then $K_S(t') \geq K_S(t)$.

Three basic questions can be raised at this stage:

- What Knowledge metric to use?
- How can K_{in} and K_S be compared?
- What use can be made of this metric?

These are the questions we intend to answer in the following sections.

2 DEFINING KNOWLEDGE LEVELS

The ideas we develop and consider in this work should be regarded within the framework of ontology. It is well known that ontology is relative to a domain of study. As explained in Section 5 below, we consider science and engineering domains. This is important for concept definition when we consider specific documents. For example, certain terms such as verbs and nouns may not be important to us and will not be counted as "concepts", whereas they may be capital for someone studying the English language.

Any framework of knowledge has to make some assumptions about the levels of granularity because knowledge is necessarily hierarchical. This question may be debated on psychological and cognitive grounds. Our arguments are however purely technical. We define four Knowledge levels,

Braham R..
A KNOWLEDGE METRIC WITH APPLICATIONS TO LEARNING ASSESSMENT .
DOI: 10.5220/0003058800050009
In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development* (KEOD-2010), pages 5-9
ISBN: 978-989-8425-29-4

although it may be argued that more levels may exist. The concept of "Knowledge Level" (KL) in our work should not be confused with the one described by Newell (1981), nor with the KLs in the sense of philosopher J. Locke (also called degrees). Our KLs are defined from a logics point of view. They allow us to present corresponding metrics as we shall explain below.

a. **Knowledge of Level 1:** this is basic Knowledge. It describes concepts, items or objects, for example animal, tree, person …
b. **Knowledge of Level 2:** Here we have properties and relations defined on concepts. Elements of Knowledge at this level require two K-elements of Level 1. Examples: a parrot is-a bird; Coca-Cola is-a soft-drink; Mozzarella cheese is-made-in Italy, lions are-faster-than humans. It includes simple relations of the type 5=2+3 and 5>4 as well.
c. **Knowledge of Level 3:** this level incorporates three cases:

- Rules and inferences, for example: hasUncle ← hasParent^hasBrother
- Logical structures of the type IF-THEN
- Equations.

d. **Knowledge of Level 4:** this is the highest level. It includes logical structures of the form IF-THEN-ELSE such as those encountered in theorems. To simplify the terminology, we will call elements (grains or items) of Knowledge of any level "Knowlets," a word inspired from applets and servlets in computer science. Note that this definition is not quite the same as Mons' (2008) and knowlets are not just the smallest "piece" of Knowledge. They are hierarchical elements of Knowledge.

3 KNOWLEDGE ENTROPY

Two Knowledge kinds are of interest to us: "stored Knowledge" (K_S) and "learned Knowledge" (K_L). The latter one is new Knowledge actually, i.e. Knowledge to be learned and added to K_S. When a person (a learner in our case) is presented with some Knowledge K_{in}, the amount of gained Knowledge, denoted by $H(K_L)$ must be computed having the following properties:

1. $H(K_L)$ is positive.
2. $H(K_L) = 0$ if $K_{in} \subset K_S$.

Shannon in his seminal work on information theory (Shannon 1948) was inspired by Hartley and used the well-known logarithmic measure for information. Since then, information theoretic approaches have flourished (Smyth and Goodman 1992, Lin 1998, etc.). We employ a logarithmic measure as well.

Using the KLs defined earlier, we have for K_{in} in the general case:

$$K_{in} = K_{in,1} \cup K_{in,2} \cup K_{in,3} \cup K_{in,4}$$

where $K_{in,n}$ is the knowlet of level n contained in K_{in}. In other words, K_{in} must be decomposed according to KLs before proceeding further. Let us assume without any loss of generality, that:

$$K_{in} = K_{in,n} \text{ (just one level), } n = 1, 2, 3 \text{ or } 4.$$

Under these assumptions, we define our Knowledge metric with the following fundamental equation:

$$H(K_L) = \alpha_n \log_2\left(1 + \frac{K_{in}^2 - K_{in} * K_S}{K_{in}^2}\right) \qquad (1)$$

where α_n is the Knowledge unit of Level n, $K_{in} * K_s$ is a measure of correlation defined as follows:

$$K_{in} * K_S = \mathrm{Sim}(K_{in}, K_{in} \cap K_S) \qquad (2)$$

where "Sim" represents a similarity function that we will discuss in more detail in Section 4 next.

Furthermore, we use the notation:

$$K_{in} * K_{in} = \mathrm{Sim}(K_{in}, K_{in}) = K_{in}^2$$

(this is consistent with notation from the field of signal processing). We give $H(K_L)$ as defined by (1) the name of "Knowledge entropy" and we choose the "bit" as a unit of measure in line with information theory since logarithm base 2 is used in (1) and throughout.

When we compare the Knowledge levels defined in Section 2 in terms of number of concepts (or ideas) involved, we find appropriate to take $\alpha_n = n\alpha_1$. Furthermore we set $\alpha_1 = \log_2 2^{1/2} = ½$ (actually $\alpha_n = \log_2 2^{n/2} = n/2$).

Let us go back to Equation (1) and examine its main properties. Two cases are to be considered:

$K_{in} \subset K_S$ and $K_{in} \not\subset K_S$.

a. $K_{in} \subset K_S$: in this case $K_{in} \cap K_S = K_{in}$ so that $K_{in} * K_S = K_{in}^2$ and thus $H(K_L) = 0$ as precisely desired.
b. $K_{in} \not\subset K_S$: if $K_{in} \cap K_S = \emptyset$ then mathematically speaking K_{in} and K_S are orthogonal ($K_{in} \perp K_S$). In this case $H(K_L) = \alpha_n \log_2(2) = \alpha_n$ (its maximum value).

If $K_{in} \cap K_S \neq \emptyset$ then $H(K_L)$ lies anywhere between zero and this maximum value.

4 KNOWLEDGE SIMILARITY

At this stage of the discussion, we need to define the similarity employed in Equation (1). Several similarity measures have been proposed in the literature, for example Lin's (1998) and Resnik's (1999). More recent similarity metrics have been proposed in d'Amato (2006) and Slimani et al. (2008). Mihalcea et al. (2005) and Warin et al. (2006) give comparative studies of these measures among others. Some of the measures are defined based on information theoretic approaches while others use a logics and/or ontology point-of-view.

The choice of a particular measure depends on the form of the objects to be compared: texts, semantic maps, rules, etc. In our case, we need to compare knowlets (concepts, properties, rules/equations, theorems). We define a similarity measure adapted from Lin's in the following way:

$$\mathrm{Sim}(K_1\ , K_2) = \frac{K_1 \cap K_2}{K_1 \cup K_2} \quad (3)$$

We could have used cardinals but we prefer to keep the notation simple. Let us illustrate the use of this definition with an example (More on this in Section 6).

Let K_1 = {father ≡ man ^ parent} and K_2 = {mother ≡ woman ^ parent}. Then:

$$\mathrm{Sim}(K_1, K_2) = \frac{\wedge + parent}{man + \wedge + parent + woman}$$

$$= \frac{2}{4} = 0.5.$$

The obvious cases of $K_1 = K_2$ and $K_1 \perp K_2$ can be easily checked (maximum and minimum similarity values).

In practice and for correct knowledge acquisition, a threshold value μ should be chosen to decide for new versus learned knowledge, for example μ=.25.

5 KNOWLETS IN PRACTICE

We are interested in this paper in documents (course materials, papers, exams) from scientific and engineering fields. These documents comprise four types of knowlets:

(1) Concepts: in the form of one or more words.

(2) Theorems: generally in the form of IF-THEN or IF-THEN-ELSE.

(3) Equations: usually definitions or a series of derivations.

(4) Examples: applications of theorems and equations for specific values and conditions.

Transforms such as Fourier, Laplace, Z, may be considered as special cases of equations and transforms come in pairs (analysis and synthesis equations). We may extend this logic to laws of physics and other entities like those suggested by Gruber (1993).

The case of examples is less straightforward and requires a more elaborate analysis. Examples may be applied to equations, theorems and so forth. They actually help us understand them. But a fundamental question is the following: how many examples are necessary to fully understand a theorem say? The answer is, in theory, a large number, approaching infinity. Of course all depends on the theorem and the examples themselves. It is however safe to assume that examples may be ranked in a decreasing order of usefulness. We make use of the fact that:

$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots \left(\frac{1}{2}\right)^n + \cdots = 1$ and assume that:

H(n examples) = β $\log_2[1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots \left(\frac{1}{2}\right)^n]$

where β = H(theorem) in the case of a theorem for example. Note that $\lim_{n\to\infty}$ H(n examples) = β.

6 APPLICATIONS AND CASE STUDY

The metrics that we have proposed may find numerous applications such as benchmarking ontologies, concept maps, T-Boxes and A-Boxes. Our own interest lies in the field of e-learning and student learning assessment more specifically. We believe that these metrics can be employed as effective tools to evaluate exams with respect to course contents. Furthermore, they may be quite useful in the automatic generation of assessment items from course material.

We illustrate these ideas with a practical example using a course on information theory that we have been teaching for a few years now. First of all we need a text reference. We have chosen Shannon's paper (1948) as it is known to a wide audience. Furthermore, it may be easily employed as a reference (at least in part) for any course on information theory.

Let us first clarify the use of the two notions of similarity measure defined in Section 4 and correlation metric defined in Section 3 with examples of concepts taken from Shannon's paper which are considered different.

Suppose K_{in} = {information source} and K_S = {discrete information source}. This corresponds to case **a** of Section 3. We have then:

$\mathrm{Sim}(K_{in}, K_S) = 3/5 = 0.6$,

$K_{in} * K_S = \mathrm{Sim}(K_{in}, K_{in} \cap K_S) = 1$

and $H(K_L) = 0$.

Now let K_{in} = {second-order approximation of English} and K_S = {first-order approximation of English}. This corresponds to case **b** of Section 3. Then:

$\mathrm{Sim}(K_{in}, K_S) = 3/7 = 0.43$,
$K_{in} * K_S = \mathrm{Sim}(K_{in}, K_{in} \cap K_S) = 3/5 = 0.6$, and
$H(K_L) = .24$ bit.

The analysis of Shannon's paper (without the appendices) reveals at least 16 concepts, 36 relations/properties (these two numbers can only be more or less subjective), 9 equations, 12 theorems and 17 examples. Out of the 12 theorems, 7 are of the form IF-THEN (equivalent to equations) and the rest 5 are of the form IF-THEN-ELSE. This analysis would have been carried out ideally with automatic techniques. But it was done manually due to the lack of appropriate tools at the present time.

According to our metrics and using the results of the above analysis, we have:

$H(\mathrm{Sh1948}) = (16+36\cdot 2+9\cdot 3+7\cdot 3+5\cdot 4)\cdot\alpha_1 + H(\mathrm{examples})$.

The examples case is somewhat complex. Seven examples are for concepts (distributed as 1, 1, 1, 1, 1, 2), five for equations (distributed as 1, 1, 3) and five for theorems (one If-Then and 2, 1, 1 If-Then-Else). Therefore:

$H(\mathrm{examples}) = [5\log_2(1+\frac{1}{2}) + \log_2(1+\frac{1}{2}+\frac{1}{4})]\alpha_1 + [2\log_2(1+\frac{1}{2}) + \log_2(1+\frac{1}{2}+\frac{1}{4}+\frac{1}{8})]\cdot 3\alpha_1 + \log_2(1+\frac{1}{2})\cdot 3\alpha_1 + [2\log_2(1+\frac{1}{2}) + \log_2(1+\frac{1}{2}+\frac{1}{4})]\cdot 4\alpha_1$,

i.e.: $H(\mathrm{examples}) = 19.6\alpha_1 = 9.8$ bits.

We have finally: $H(\mathrm{Sh1948}) = 175.6\alpha_1 = 87.8$ bits of Knowledge entropy.

It may be useful to compute the average Knowledge entropy. In this case it is equal to $\frac{87.8}{90} = .975$ b/knowlet.

This figure is relatively low (less than 1), but if we do not take into account the examples, it becomes 1.07 b/knowlet.

We looked at exams for the last three years. We have found an average of 10 concepts and 4 equations per exam. We may conclude that H(exam) $= 22\cdot\alpha_1 = 11$ bits, i.e. 12% of Shannon's paper. Although there are no standards in the literature to tell us what value would be acceptable, a coverage value of 10% minimum should be in our opinion teachers' target.

7 RELATED WORK AND CONCLUSIONS

In this paper we gave a foundation for Knowledge metrics. Ideally, an automatic generation of Knowledge from text or documents should be made. Then documents are compared automatically as well. This undertaking is for that matter impractical at the present time as necessary tools are still under development. Significant progress has been made in the area of ontology learning and population during the last few years. Valuable tools have been proposed in this regard (Zouaq and Nkambou 2008, Buitelaar et al. 2003), but some time is still needed before they become fully operational.

To the best of the author's knowledge, the only works that grade exams and course contents using quantitative metrics are those based on Bloom's six-level knowledge taxonomy, for example Oliver et al. (2004) and Zheng (2008). We therefore believe that we have presented original ideas that would allow us to assess quantitatively our exams with respect to course contents we present to students.

The metrics we defined may be used for comparative purposes but with due precaution. The scientific importance of any specific theorem for example is measured with its impact on the course of science and technology and is by no means an absolute value.

We should note that the knowledge metrics we have defined open a large scope of applications especially those related to ontology development and comparison and not just learning assessment.

Another possible further exploration can be done to assess the validity of theses metrics from a cognitive standpoint. Indeed we have refrained from speculation on how this work would compare against human perception of knowledge. We leave it for a future exploration.

ACKNOWLEDGEMENTS

The author would like to thank ISITCom and Prince Research Group for their financial support. We thank also the anonymous reviewers for their suggestions to improve the final manuscript.

REFERENCES

d'Amato, C. et al., 2005. Semantic Similarity Measure for Expressive Description Logics. in: Proc. CILC, Rome, Italy, June.

Buitelaar, P., Cimiano, P. and Magnini, B., 2003. Ontology Learning from Text: An Overview. *IOS Press*, pp. 1-9.

Cheniti-Belcadhi, L., Braham, R., Henze, N. and Nejdl, W., 2004. A Generic Framework for Assessment in Adaptive Educational Hypermedia. In: Proc. *IADIS WWW / Internet, Madrid, Spain, pp. 397-404.*

Cheniti-Belcadhi, L. and Braham, R., 2008. Assessment Personalization on the Semantic Web. Journal of Computational Methods in Sciences and Engineering (IO Press), *Special issue: Intelligent Systems and Knowledge Management, 8, 1-20.*

Gardner, L., Sheridan, D. and White, D., 2002. A web-based learning and assessment system to support flexible education. *Journal of Computer Assisted Learning* 18 (2), 125-136.

Gruber, T. R., , 1993. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. Technical Report KSL 93-04, Knowledge Systems Lab.

He, L. 2006. A Novel web-based educational assessment system with Bloom's taxonomy. *Current Developments in Technology-Assisted Education*, 1861-1865.

Hot Potatoes, 2010. University of Victoria. Available at http://hotpot.uvic.ca/.

IMS Global Learning Consortium, 2006. IMS Question and Test Interoperability Overview, Version 2.1, 8 June. URL: http://www.imsglobal.org/question/qtiv2p1pd2/imsqti_oviewv2p1pd2.html. Accessed 5 march 2010.

Lin, D., 1998. An Information Theoretic Definition of Similarity. I*n Proc. of the 15th Int. Conference on Machine Learning.*

Mihalcea, R. et al., 2006. Corpus-based and Knowledge-based Measures of Text Semantic Similarity, in: Proc. of *21st National Conference on AI, AAAI,* pp. 775-780.

Mons, B., 2008. Calling on a million minds for community annotation. In WikiProteins. Available at: http://genomebiology.com/2008/9/5/R89.

Newell, A., 1981. The Knowledge Level, AI Magazine, Summer, 1-20.

Oliver, D. , Dobele, T., Greber, M. and Roberts, T., 2004. This Course Has A Bloom Rating Of 3.9, in: Proc. Sixth Australasian Computing Education Conference *(ACE2004), Conferences in Research and Practice in Information Technology,* Vol. 30. Available at: http://crpit.com/confpapers/CRPITV30Oliver.pdf.

Palloff, M. and Pratt, K., 2006. How do we know they Know? Student Assessment Online. *In Proc. 22nd Annual Conf. on Distance Teaching and Learning, Wisconsin,* pp. 1-5.

Resnik, P., 1999. Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language. *Journal of Artificial Intelligence Research,* 11, 95-130.

Shannon, C., 1948. A Mathematical Theory of Communication, *The Bell System Technical Journal,* 27, July/October, 379–423/623–656,.

Slimani, T., Ben Yaghlane, B. and Mellouli, K., 2006. A New Semantic Similarity Measure based on Edge Counting, *World Academy of Science, Engineering and Technology,* 23, 34-38.

Smyth P. and Goodman, R. M., 1992. An Information theoretic Approach to Rule Induction from Databases, *IEEE Trans. on Knowledge and Data Engineering*, 4 (4), 301-316.

Warin, M., Oxhammar, H. and Volk, M., 2005. Enriching an Ontology with WordNet based on Similarity Measures, in: *Proc. of the MEANING Workshop.*

Zheng, A. Y., 2008. Application of Bloom's Taxonomy Debunks the MCAT Myth, Science, 319, 25 Jan., 414-415.

Zouaq, A. and Nkambou, R., 2008. Building Domain Ontologies from Text for Educational Purposes, *IEEE Trans. on Learning Technologies*, 1 (1), 49-62.