SEMANTIC REPRESENATION WITH KNOWLEDGE MEASURING FOR DISTANCE LEARNING

Ivashenko V.P.

Belarusian State University of Informatics and Radioelectronics, Minsk, Belarus, ivashenko@bsuir.by

Abstract. Several approaches to measuring individual knowledge are considered. The article concerns on ontological models of educational material, tests and its features which can be used to measure correspondence between expert and got answer for scoring tested person.

Attestation is one of the parts of the educational process.

Consider some approaches to estimate knowledges of tested person. These approaches depend on type of tests, structure of educational material and person capabilities to give answers. Usually there is distinguishing between closed and open (free) test. But in practice the real difference is between number of answers, its forms and sizes and number and properties of answer patterns, rules or schemas accessible for tested person.

The process of attestation can be considered as a measuring of the correspondence between knowledge of tested person and teacher (expert). This process depends on kinds of scales used for measurement.

There are several approaches:

– expert ranking and classification;

– heuristics and similarity measures;

– probabilistic models;

– precision measures and semantic metrics based on ontology models [5, 6, 7, 8, 9].

The last approach is seemed be most precise. Many test systems have the test material *P*, auxiliary textbook material T and some of them have the ontological model B. The existence of the ontological model means that there are mappings from *P* or *T* to *B* and reverse. In the case of the absence of the ontological model we need to use expert knowledge or heuristic approach. In such case, the language dependent lexical measures can be used such as Levenshtein distance or others. Finally the result of attestation has to be mapped on the scale *S*.

The scale *S* can be one of the Stevens scales: nominal, ordinal, interval or ratio [1, 2]. This scale related to *T* or *B* is usually used with the purpose to qualify and select tested person by ranking and classification for the specific activity.

Each part of the knowledge in the educational material and each answer can be interpreted as phenomena which can be modelled with a semantic model as the part of the ontology [5, 6, 9].

As concerns probabilistic models we have some assumptions about distribution of cases for random variable and events. For example, consider set of *n* independent closed tests with probability p_i to get right answer with expectation μ.

$$\mu = \frac{\sum_{i=1}^{n} p_i}{n}. \quad (1)$$

The mapping to the scale S may to have the next form:

$$\max\left(\left\{s_{\min}, \left\lfloor \frac{s_{\max} - s_0}{1-\mu} * \left(\frac{q}{n} - \mu\right)\right\rfloor\right\}\right), \quad (2)$$

where *q* is a number of right answers, s_0, $s_{\max}$ and $s_{\min}$ are correspondingly a zero, maximum and minimum value of ratio scale *S*. For *n* equals 20, $\mu = 4$, $s_0 = s_{\min} = 0$ and $s_{\max} = 10$, we get the next correspondence (see table 1).

Table 1. Attestation scoring with probabilistic models

number of right answers	attestation score
0-5	0
6-7	1
8	2
9-10	3
11	4
12-13	5
14-15	6
16	7
17-18	8
19	9
20	10

It is rather simple model. Further progress in the process of the scoring unification can be achieved by the transition to z-normalized scores. This and more sophisticated models are investigated and printed in the next publications [2, 3, 4, 6].

Let enumerate some kinds of test based on several features. Let take in account the set of next features: world assumption (WA) (closed (CWA), open (OWA)), number of alternative answers (binary, several (alternative), many (multiple choice), finite (“open” type test), infinite), distance from prompts, answer prototypes or patterns in the question text (short, middle, long), regular («deterministic») or irregular («stochastic») question appearance. It is to be noted that the matching type (building correspondence or sequence) test despite of its form is close to alternative or multiple choice test (see tables 2 and 3).

Table 2. Test model complexity

	world assumption	number of answers				
		finite				infinite
		binary	several	many		
	CWA					
	OWA					

Table 3. Test interpretation complexity

	question appearance	interpretation distance		
		short	middle	long
	regular			
	irregular			

It is important to admit that the each test is accounted to contain only one question being independent of any others. Consideration of a series of questions and strategies of querying tested person is beyond the scope of this article. By the similar reason there is no consideration of such forms of test as written test or oral tests despite this one to have psychological issues.

In the case of the CWA the hard computing model can be used to classify answer and get attestation score. While with the OWA case some more flexible computing models including soft computing with ordinal or metric scales can be used to compute answer completeness or distance to the complete answer.

The existence of ontological model allows using knowledge specification model to compute measures based on properties of the correspondences between formal models of ontologies.

Depending on the structure of the ontology, its formal representation on the knowledge representation language, there is an acceptable transformation distance of the structure of the one of ontologies to the structure of another. For symmetric graph languages [4], this distance can be given by a quadruple containing number of added and removed vertices and edges (v^+_{ij}, v^-_{ij}, e^+_{ij}, e^-_{ij}). Therefore, the distance can be defined as follows.

$$d\left(\left\langle i,j\right\rangle\right)=\sqrt{\left(v^+_{ij}\right)^2+\left(v^-_{ij}\right)^2+\left(e^+_{ij}\right)^2+\left(e^-_{ij}\right)^2}\ . \quad (3)$$

The defined measure matches all properties of the distance.

$$d\left(\left\langle i,i\right\rangle\right)=0\ . \quad (4)$$

The added vertices are replaced with deleted ones with a change in the direction of transformation.

$$v^+_{ij}=v^-_{ji}\ ;\ e^+_{ij}=e^-_{ji}\ . \quad (5)$$

So, the last implies symmetric properties.

$$d\left(\left\langle i,j\right\rangle\right)=d\left(\left\langle j,i\right\rangle\right). \quad (6)$$

The triangle inequality can be derived from the follows.

$$\begin{aligned}
V^+_{ik}&=\left(\left(V^+_{ij}/V^-_{jk}\right)\cup\left(V^+_{jk}/V^-_{ij}\right)\right)\subseteq\left(V^+_{ij}\cup V^+_{jk}\right)\\
V^-_{ik}&=\left(\left(V^-_{ij}/V^+_{jk}\right)\cup\left(V^-_{jk}/V^+_{ij}\right)\right)\subseteq\left(V^-_{ij}\cup V^-_{jk}\right)\\
E^+_{ik}&=\left(\left(E^+_{ij}/E^-_{jk}\right)\cup\left(E^+_{jk}/E^-_{ij}\right)\right)\subseteq\left(E^+_{ij}\cup E^+_{jk}\right)\\
E^-_{ik}&=\left(\left(E^-_{ij}/E^+_{jk}\right)\cup\left(E^-_{jk}/E^+_{ij}\right)\right)\subseteq\left(E^-_{ij}\cup E^-_{jk}\right),
\end{aligned} \quad (7)$$

where the powers of the sets of vertexes and edges are:

$$v^+_{ij}=\left|V^+_{ij}\right|;\ v^-_{ij}=\left|V^-_{ij}\right|;\ e^+_{ij}=\left|E^+_{ij}\right|;\ e^-_{ij}=\left|E^-_{ij}\right|, \quad (8)$$

so that:

$$v^+_{ik}=\left|V^+_{ik}\right|\le\left|\left(V^+_{ij}\cup V^+_{jk}\right)\right|\le\left|V^+_{ij}\right|+\left|V^+_{jk}\right|=v^+_{ij}+v^+_{jk}\ . \quad (9)$$

This all leads us to the result.

$$d\left(\left\langle i,k\right\rangle\right)\le d\left(\left\langle i,j\right\rangle\right)+d\left(\left\langle j,k\right\rangle\right). \quad (10)$$

Other measures taking in account operational semantics can be defined [9].

The measuring of the correspondences between ontologies can be applied with the combination of answer representation in the logical knowledge representation languages and its transformation to the normal and canonical forms. Some software components implementing elements of the described approaches were probed in the educational process of the Department of the Intellectual Information Technologies of the BSUIR [10].

Literature

1. Stevens, S. S. On the Theory of Scales of Measurement. Science, 1946. – 103 (2684). – pp. 677–680.

2. Educational measurement and evaluation. USI Publications 2/31, Nehru Enclave, Kalkaji Ext., New Delhi-110019 for Lovely Professional University, Phagwara, 2012. – 400 p.

3. Yen, W.M. The choice of scale for educational measurement: an IRT perspective. [Electronic resource]. – Mode of access: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1745-3984.1986.tb00252.x Date of access: 10.11.2018.

4. Borgatti, S.P. Carboni, I. On Measuring Individual Knowledge in Organizations. Organization Research Methods. – Vol. 10 (3). – 2007. – pp. 449–462.

5. Golenkov V.V., Guliakina N.A., Tarasov V.B., Eliseeva O.E. et al. Intellektual'nye obuchayushchie sistemy i virtual'nye uchebnye organizatsii. Minsk: BSUIR, 2001. – 488 p. (in Russian).

6. Ivashenko, V.P. Ontologicheskaya model' prostranstvenno-vremennykh otnoshenii sobytii i yavlenii v protsessakh obrabotki znanii [Ontological model of space-time relations for events and phenomena in the processing of knowledge]. Vestnik BrGTU, 2017. – №5(107). – pp. 13 –17. (in Russian).

7. Rybina, G.V. Obuchayushchie integrirovannye ekspertnye sistemy: nekotorye itogi i perspektivy. Iskusstvennyi intellekt i prinyatie reshenii 1, 2008. – pp. 22–46. (in Russian).

8. Gavrilova T.A., Gorovoy V.A., Bolotnikova E.S., Gorelov V.V. Sub"ektivnye metriki otsenki ontologii. [Electronic resource] – Mode of access: http://www.math.nsc.ru/conference/zont09/reports/39Gavrilova-Gorovoi-Bolotnikov-Gorelov.pdf. Date of access: 10.11.2019.

9. Ivashenko, V.P. Attributes, scales and measures for knowledge representation and processing models. Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nykh system [Open semantic technologies for intelligent systems], 2019. – pp. 247–250.

10. Variety Virtual Machine. [Electronic resource]. – Mode of access: http://scnedu.sourceforge.net/variety/_/index.html?variety=examinator.LFAI.E.1.json&gradient=tester.json. Date of access: 11.11.2019