Instance-Based Ontology Matching For Open and Distance Learning Materials

Sergio Cerón-Figueroa[1], Itzamá López-Yáñez[2], Yenny Villuendas-Rey[2], Oscar Camacho-Nieto[2], Mario Aldape-Pérez[2], and Cornelio Yáñez-Márquez[1]
[1] *Centro de Investigación en Computación, Instituto Politécnico Nacional, México,* [2] *Centro de Innovación y Desarrollo Tecnológico en Cómputo, Instituto Politécnico Nacional, México*

Abstract

The present work describes an original associative model of pattern classification and its application to align different ontologies containing Learning Objects (LOs), which are in turn related to Open and Distance Learning (ODL) educative content. The problem of aligning ontologies is known as Ontology Matching Problem (OMP), whose solution is modeled in this paper as a binary pattern classification problem. The latter problem is then solved through the application of our new proposed associative model. The solution proposed here allows the alignment of two different ontologies —both in the Learning Objects Metadata (LOM) format— into a single ontology of LOs for ODL in LOM format, without redundant objects and with all inherent advantages for handling ODL LOs. The proposed model of pattern classification was validated through experiments, which were done on data taken from the Ontology Alignment Evaluation Initiative (OAEI) 2014 campaign, as well as on data taken from two known educative content repositories: ADRIADNE and MERLOT. The obtained results show a high performance when compared against some of the classifier algorithms present in the state of the art.

Keywords: open and distance learning, ontology matching problem, e-learning, pattern recognition, associative classifier

Introduction

In recent years, the rising of the World Wide Web has drastically changed the way people learn. In particular, Open and Distance Learning (ODL) makes use of computers and data networks to store,

exchange, and manage learning assets, such that these resources are always available, anywhere, and anytime (Gooley & Lockwood, 2001). The ODL approach also benefits from using Big Data in learning environments (Prinsloo, Archer, Barnes, Chetty, & van Zyl, 2015).

Thanks to the great advances in information technologies, the ODL industry has taken advantage of novel internet tools (Dabbagh et al., 2016). Thus, emerging technologies related to Semantic Web (SW) have allowed ODL systems to change their perspective, from a task oriented approach towards a knowledge discovery based approach (Beydoun, 2009).

Given the importance of SW in environments related to ODL, it is noteworthy that SW is considered to be an evolution of the traditional World Wide Web (Karger, 2014; Thangaraj & Sujatha, 2014). This evolution began as an initiative by the World Wide Web Consortium (W3C, 2016) which followed the original concept introduced by Tim Berners-Lee three decades ago (Berners-Lee, Hendler, & Lassila, 2001).

In the context of the present work, it bears mention that SW is based on ontologies, which describe in a general manner concepts and objects, as well as the relationships among them, pertaining to a specific body of knowledge (Caliusco & Stegmayer, 2010). In this paper, the ontologies of interest are those containing Learning Objects (LOs) related to ODL.

The Ontology Matching Problem (OMP) has recently become a relevant research topic (Shvaiko, & Euzenat, 2013). Several proposals from different research groups around the world have been made, to tackle the problem of ontology alignment. This issue can be seen as a natural result of the increasing popularity of Semantic Web. OMP intends to solve the problem of heterogeneity in SW, looking for certain relationships between similar elements among different ontologies (Otero-Cedeira, Rodríguez-Martínez, & Gómez-Rodríguez, 2015; Mao, Peng, & Spring, 2011).

In order to appropriately describe the OMP in ODL environments, let us consider a typical situation in handling ODL materials, which implies managing LOs applied to ODL. Thus, there is access to LOs belonging to two different ontologies —say A and B— in the LOM format, which is a very popular standard for LO metadata management. Then the OMP solution for ODL environments consists on aligning, by means of an Ontology Alignment Model (OAM), both ontologies A and B into a single ontology C containing LOs for ODL. This resulting ontology will be represented in the same LOM format, with no redundant objects, and with all inherent advantages for LO management in ODL. An illustration of these ideas can be seen in Figure 1.
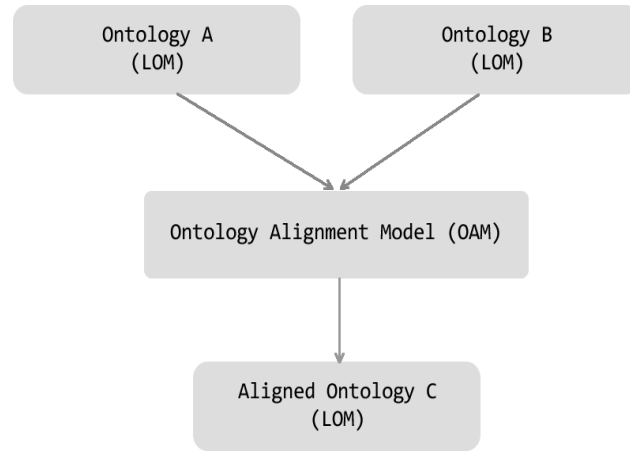
Figure 1. Illustration of an instance based solution for OMP in ODL environments.

LOM is an IEEE 1481.12.3 standard (IEEE Standard for Learning Object Metadata, 2002) which is defined as an extension of XML for the description of LO characteristics. For the LOM standardization, the models are proposed to be represented and used as ontologies. In this standard, the metadata associated to an instance describe relevant characteristics of the LO.

One of the main tools in the OMP is to find an appropriate similarity function, in order to build pairs of objects which are actually close in meaning (Houshmand, Naghibzadeh, & Araban, 2010). Through the adequate use of a given similarity function, it is possible to transform the OMP into a binary pattern classification problem (Mao, Peng, & Spring, 2008) as follows $m(o_{Ai}, o_{Bj}, r_{ij}) \rightarrow \{0,1\}$ Where m is the OAM, o_{Ai} is the i-th element from ontology A, o_{Bj} is the j-th element from ontology B, and r_{ij} is the value of the similarity relationship between these two elements, according to a similarity function chosen previously. The result is class 1 when o_{Ai} and o_{Bj} represent the same object, and class 0 otherwise. As is shown in Figure 1, el OAM plays a central role in this process.

Some of the advantages of tagging documents and resources with metadata consist of enabling the search, acquisition, and use of LOs in a simple manner, both for teachers and students. Also, this allows information exchange between different e-learning systems (Pietranik & Nguyen, 2014).

The main advantages of having no redundant objects in ontologies are reflected in lower storage needs, as well as simpler and faster LO searches in the resulting ontology. Several techniques exist to solve this problem; however, a relevant approach consists on using data mining to find certain patterns in each ontology and thus build the matching model (Rubiolo, Caliusco, Stegmayer, Coronel, & Gareli Fabrizi, 2012). In the current work, a new associative model is introduced to solve the OMP in ODL environments.

The rest of the paper is organized as follows. The Literature Review section presents related work in order to contextualize our proposal and showcase how the OMP has been tackled before. Next, the theory behind the proposed associative model in both phases, learning and classification, is explained; while the next section describes and exemplifies the Implementation of both phases of the proposed model. The Experimental Results and Discussion section follows, which details the experimental design followed and

presents the results obtained by our proposal. The paper concludes with Conclusions, proposals for related Future Work, and References.

Literature Review

This section is dedicated to discussing some recent scientific works, related to several topics relevant to the development of the present paper, such as: metadata extraction, stating the OMP as a binary classification problem, the use of data mining models —e.g., Support Vector Machines (SVMs), decision trees, proximity-based recognition methods, directed graphs, or evolutionary algorithms— in the OMP, SW, and e-learning, massive open online courses (MOOC), pattern recognition, and pattern classification.

The work of Atkinson, Gonzalez, Munoz, and Astudillo (2014) explores several strategies for educational metadata extraction, whose one of the most relevant open problems is the identification of LOs and the metadata that can be extracted from them.

Furthermore, both Mao, Peng, and Spring (2008) and Liu, Yang, Zhang, Wu, and Hu (2012) show how Ontology Matching can be stated as a binary classification problem, making use of pattern recognition algorithms. In the former work, a methodology for finding relationships between two ontologies using Support Vector Machines (SVM) is presented. The experimental results reported therein are outstanding when compared against other mapping methods.

In the same sense, Liu, Qin, and Wang (2013) present a new ontology matching model which uses SVMs, showing a precision of the order of 95% in their experimental results. Other works, such as (Yang & Steele, 2009), show ontology mapping methods based on decision trees. Also, the latter work presents a similarity measure between two elements belonging to different ontologies. However, the work does not present precision results, while arguing that the model presented there is faster at execution due to the fewer comparisons needed.

SW has been widely used along pattern recognition algorithms in order to solve several medical problems. Such is the case of Mohammed and Benlamri (2014), where a recommendation system for clinical diagnostics of some medical conditions is presented. This system uses ontologies and proximity-based pattern recognition methods.

Taking a different approach to the problem at hand, the authors of Vidal, Lama, Otero-García, and Bugarín (2014) introduce a method to represent e-learning resources semantic annotations by means of directed graphs. The resulting directed graphs are then used, along with linked data, in order to obtain document search with high recall and precision.

Mentaheuristics have also had a relevant role in the area of e-learning. In this sense, Luna, Romero, Romero, & Ventura (2014) introduce an association model for discovering learning rules using evolutionary algorithms.

Regarding the union of SW and pattern recognition, Peñalver-Martinez, Garcia-Sanchez, Valencia-Garcia, et al. (2014) apply natural language techniques to resources generated by SW for opinion mining. On the other hand, Wang, Peng, and Liu (2015) presents an advanced model for the classification of less popular web pages. In this work, the authors make use of Latent Semantic Analysis and Density-based Rough Sets for the automatic tagging of internet web pages.

One research area which has branched out of pattern recognition —becoming widely active in recent years— is that of recommendation systems, which have found great acceptance and application in e-commerce platforms. However, Aher and Lobo (2013) present an online courses recommendation system, which combines several clustering algorithms found in the WEKA data mining platform (Hall et al., 2009), such as k-Means. The results of said work show that pattern recognition methods can improve the evaluation process of courses immersed in e-learning environments.

Another work which combines pattern classification and e-learning is by Gladun, Rogushina, García-Sanchez, Martínez-Béjar, and Fernández-Breis (2009), which presents a multi-agent recommendation system for automatic feedback regarding knowledge acquired by students in e-learning platforms, taking advantage of the SW.

Due to the existing relationship between ontologies and SW, there are several works that make use of the normal web to gain knowledge. One such case is that of Li, Xu, Zhang, and Lau (2014), where this tool is applied to a new prediction model for unemployment rates.

Tabaa and Medouri, (2013) show some methods by which data mining can be applied to MOOCs, making a particular emphasis on the tasks of pattern generation and data set building. Other works on distance learning are focused on proposing a novel way of micro-learning through mobile terminals (Wen & Zhang, 2015), while others focused on expanding educational horizons (Walters, Walters, Green, & Lin, 2016).

In this sense, there are few works in the current scientific literature which make use of pattern recognition algorithms for Ontology Mapping. One such case is by Hariri, Abolhassani, and Sayyadi (2006), where decision trees and neural networks are employed for ontology alignment. The results obtained there indicate that neural networks offer a better precision. Similarly, Pietranik and Nguyen (2014) introduces a new model for ontology alignment based on a set of algorithms which work at different levels of granularity.

The Proposed Associative Model

With respect to Figure 1, the original contribution of this paper to the state of the art is an associative model for pattern classification that works as an OAM. Given that the proposed solution is instance based, two LOs will make up the input of the original model (one from ontology A and one from ontology B) and the proposed model will output the class corresponding to that pair of objects: class 1 if the pair represent the same object, and class 0 otherwise.

The proposed associative model belongs to the supervised learning paradigm (Marques de Sá, 2001), and is divided into two stages: the learning phase and the classification phase. Thus, a set of previously classified patterns will be used in the training phase of the classifier, with the intention of having said algorithm correctly classify the patterns in the test set during the classification phase.

Hypothesis 1. Given that the Associative Model (AM) proposed herein belongs to the supervised learning paradigm, it is assumed that there is a pattern bank available. This pattern data set has finite dimension n and finite cardinality q, and its patterns have been previously classified in c classes. It is also assumed that this pattern bank is partitioned into two sets: a training set of cardinality p, and a test set of cardinality q-p, where $n, p, q, c \in N$ such that $n, p, q, c > 1$ and $q > p$.

Let k be a class of the training pattern set with $k \in \{1, 2, \dots, c\}$ and let $N(k)$ be the cardinality of class k; this means that $\sum_{k=1}^{c} N(k) = p$. The form of the i-th pattern belonging to class k in the training pattern set, where $i \in \{1, 2, \dots, N(k)\}$ is as follows: $x^{ki} = \begin{pmatrix} x_1^{ki} \\ x_2^{ki} \\ \vdots \\ x_n^{ki} \end{pmatrix}$

Hypothesis 2. The AM requires working with linear equation systems which may have at least one solution; thus, it is assumed that $n < N(k), \forall k$. This means that the dimension of the patterns is less than the cardinality for each of the classes in the data set.

Assuming there is access to a pattern bank which fulfills hypotheses 1 and 2, the definitions of the operators α, β, u_β, and the modified Gamma similarity operator (López-Yáñez, Argüelles-Cruz, Camacho-Nieto, & Yáñez-Márquez, 2011) are required.

Definition 1: α and β operators. Let sets A and B be defined as $A = \{0, 1\}$ and $B = \{0, 1, 2\}$; then the α and β operators are defined as shown in Tables 1 and 2.

Table 1

α Operator

α: AxA →B		
x	y	α(x,y)
0	0	1
0	1	0
1	0	2
1	1	1

Table 2

B Operator

β: BxA →A		
x	y	β(x,y)
0	0	0
0	1	0
1	0	0
1	1	1
2	0	1
2	1	1

Definition 2: u_β operator. The u_β operator receives as input a binary vector x and outputs a non-negative integer number, which is computed as $\mu_\beta = \sum_{i=1}^{n} \beta(x_i, x_i)$.

Definition 3: Modified Gamma similarity operator. Let there be the set $A = \{0,1\}$, a number $n \in Z^+$, two n-dimensional binary vectors $x \in A^n$ and $y \in A^n$, whose i-th component is represented by x_i, and y_i, respectively. Then, the modified Gamma similarity operator $\gamma(x, y)$ has two n-dimensional binary vectors as input, and its output is an integer number calculated as $\gamma(x,y) = u_\beta[\alpha(x,y) mod 2]$

AM learning phase. One pattern among the $N(1)$ patterns in class 1 is chosen randomly; say x^{1i}, with $i \in \{1,2,...,N(1)\}$. By hypotheses 2, it holds that $n < N(1)$; thus it is possible to compute the modified Gamma similarity operator between x^{1i} and the remaining $N(1) - 1$ vectors in class 1. Yet, since the arguments of the modified Gamma similarity operator must be binary patterns, a conversion of the patterns components is necessary. First each component is converted into a positive integer and then into a binary value, considering for each value the number of bits necessary for converting the larger value in all patterns belonging to class 1, say m. Then, after the conversion there are $N(1)$ vector of mn bits in class 1, which are denoted as $\{x_b^{11}, x_b^{12}, ..., x_b^{1N(1)}\}$.

The $N(1) - 1$ values $g(j) = \gamma(x_b^{1i}, x_b^{1j})$ are computed, with $j \in \{1,2,...,N(1)\}, j \neq i$. The $N(1) - 1$ values $g(j)$ are ordered, selecting the $n - 1$ greatest values.

The set formed by the union of vector x^{1i} and the $n - 1$ vectors x^{1j} whose $g(j)$ values are the greatest in step 4, is the Base Vector Set for class 1 —also known as BVA(1)— whose cardinality is n. Also, the notation of the vectors in $BVA(1)$ is changed for ease of use to $BVA(1) = \{x^{11}, x^{12}, ..., x^{1n}\}$.

This procedure is repeated for each of the remaining $c - 1$ classes $k \in \{2, ..., c\}$, such that at the end there are c Base Vector Sets $BVA(1), ..., BVA(c)$.

AM classification phase. We have as input a pattern from the test set, whose class is unknown for AM; say vector x with dimension n. With the input vector x and the set $BVA(1)$, the following linear combination of vectors, of dimension n, is formed, where $a_{1i} \in R, \forall i: x = \sum_{i=1}^{n} a_{1i} * x^{1i}$. This expression is a

linear system of n equations with n variables, which are the n real numbers a_{1i}. The set of n real numbers a_{ki} for class $k \in \{2, \dots, c\}$ is denoted as $A(k) = \{a_{k1}, a_{k2}, \dots, a_{kn}\}$.

According to linear algebra theory (Grossman & Godoy, 2012), the previous system has the following three exclusive possibilities: it has no solution, it has a unique solution, or it has an infinite number of solutions.

If the previous system has no solution or has an infinite number of solutions, it is conjectured that pattern x does not belong to class 1. On the other hand, if the system has a unique solution, it is conjectured that pattern x may belong to class 1.

The latter process of solving the linear systems is done for each of the remaining $c - 1$ classes $k \in \{2, \dots, c\}$.

If only the equation system corresponding to class $r \in \{1, 2, \dots, c\}$ has a unique solution, while for the rest of the classes the systems have no solution or an infinite number of solutions, then pattern x belongs to class r.

Otherwise, if there is a unique solution for more than one equation system, say for classes $r_1, r_2, r_3, \dots, r_t, t \in \{1, 2, \dots, c\}$; then it is conjectured that pattern x may belong to any of these t classes. This means that a disambiguation criteria is needed to determine to which of these classes does pattern x belong. To do so, the variance of each set $A(1), A(2), \dots, A(t)$ is computed (Sachs, 1982), and pattern x is assigned to the class corresponding to the lesser variance.

Implementation of the Proposed Model: Learning and Classification Phases

When implementing the proposed associative model as an OAM in the OMP described in the Introduction section, we may see that c=2, since only two classes are considered when applying the OAM to each pair of objects, knowing the r_{ij} value of the similarity relationship between these two elements: $m(o_{Ai}, o_{Bj}, r_{ij}) \rightarrow \{0,1\}$. The result is class 1 when o_{Ai} and o_{Bj} represent the same object, and class 0 otherwise.

As a previous step to extracting the features that make up the pattern from the LOs belonging to the ontologies used on the experiments, some basic text filters are used for the pre-processing of the data: punctuation symbols and parenthesis were eliminated, all words were converted to lowercase, and the texts were translated into English, independently of the original language (Microsoft Bing API, 2016).

Next, one of the major problems for the use of pattern recognition algorithms on ontologies is tackled: feature extraction and LO representation (Pancerz & Lewicki, 2014). Given that the ontologies used on the experiments are already in the LOM format, every LO of such ontologies has an identifier and at least the four attributes mentioned in Table 3.

Table 3

LO Attributes Present in the LOM Format

Attribute	Representation
Identifier	String
Title	S_{title}
Description	$S_{description}$
Keywords	$S_{keywords}$
Learning Resource Type	Integer

Below are two examples of these five attributes (4 strings and one integer) as seen in Table 3, on LO taken from specific ontologies:

Table 4

Example of the Five Attributes of Table 3, on LO Taken from Specific Ontologies

Attribute	Value
Identifier	A4189051B
Title	Polynesian Bay
Description	A photograph of a bay near Raietia, Polynesia.
Keywords	Arts: Travel and Tourism: Polynesia
Learning Resource Type	0 (Book)

Attribute	Value
Identifier	C10750GL3
Title	World War II
Description	Events related to military weapons of World War 2.
Keywords	war: military: german: tank: guns
Learning Resource Type	1 (Quiz)

Given that the identifier does not contribute any useful information for determining whether two objects are related, this attribute is discarded for the extraction of features to make up the patterns from the LOs. Thus, the remaining four attributes were selected for feature extraction, through the use of a similarity function. In this work, the similarity function sim presented in (Tang, Li, Liang, Huang, Li, & Wang, 2006) is used, in order to determine the four features of each of the patterns. Let o_{Ai} be the i-th element from ontology A, and let o_{Bj} be the j-th element from ontology B; then the four features of the pattern that represents one pair of LOs is computed as indicated in table 4.

Table 5

Pattern Features

Feature	Operation
1	$sim\left(o_{Ai \text{ (title)}}, o_{Bj \text{ (title)}}\right)$
2	$sim\left(o_{Ai \text{ (description)}}, o_{Bj \text{ (description)}}\right)$
3	$sim\left(o_{Ai \text{ (keywords)}}, o_{Bj \text{ (keywords)}}\right)$
4	$\delta(o_{Ai \text{ (type)}}, o_{Bj \text{ (type)}})$ is the Kronecker delta, defined as: $\delta(x,y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases}$
Class	1 if o_{Ai} match with o_{Bj}, 0 otherwise

The value for class is 1 when o_{Ai} and o_{Bj} represent the same object, and 0 otherwise.

Notice that in this stage of pattern building, the fact that two objects match or not is already known, from the moment that the two objects o_{Ai} and o_{Bj} are selected. Notice also that the similarity value r_{ij} between objects o_{Ai} and o_{Bj} in equation (1) is implicit in the four features that make up the training patterns for the AM, since these features were computed by means of determining the individual similarity values between attributes, using the similarity function sim.

In this manner, the patterns are built as a 4 featured numerical vector, while the class is represented by an extra feature which may be 1 or 0, depending on whether both entities represent the same object.

For building the data bank, random pairs of LOs were selected, combining them with the pairs of LOs which were already known to be related. For this, half of the data bank corresponds to related LOs (class 1), and the other half corresponds to unrelated pairs of LOs (class 0). Considering this approach class balance was guaranteed.

The obtained data bank is divided into two disjoint sets: on one hand the Training set is formed by LOs including the class, so that they can be used during the *Learning Phase* of the AM. On the other hand, the Test set is formed with LO that do not include the class, in order to use them during the *Classification Phase* of the AM. Figure 2 illustrates the former process.
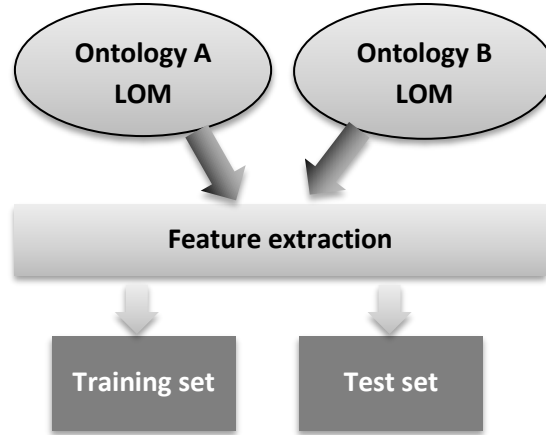
Figure 2. Illustration of the feature extraction process, and the obtaining of the Training and Testing sets.

By applying the trained AM on the Testing set during the Classification Phase, through the use of some model validation technique such as Leave-one-out or *k*-fold cross-validation (Kohavi, 1995), it is possible to evaluate the AM performance. Thus the performance measures presented in the following section are a valuable indicator of the expected behavior of AM, when it is applied to two ODL ontologies as an OAM, in the context of Figure 1.

Experimental Results and Discussion

The proposed model of pattern classification was validated through two experiments, which were done on data taken from the Ontology Alignment Evaluation Initiative (OAEI) 2014 campaign (Dragisic et al., 2014), as well as on data taken from two known educative content repositories: ADRIADNE (ADRIADNE, 2016) and MERLOT (MERLOT, 2016). Thus, two data sets were built, containing patterns representing the relationships between pairs of LOs taken from two different ontologies immersed in the ODL context.

In order to evaluate the performance of the proposed method, the usual evaluation measures were used: precision, recall, and f-measure (Hariri, Abolhassani, & Sayyadi, 2006).

$$precision = \frac{correct\_mappings \cap found\_mappings}{|found\_mappings|}$$

$$recall = \frac{correct\_mappings \cap found\_mappings}{|correct\_mappings|}$$

$$f - measure = \frac{2(precision) * (recall)}{precision + recall}$$

Below, each experiment is presented and the obtained results discussed.

1st Experiment – OAEI Instance Matching

For the first experimental test, the OAEI 2014 data bank was used, for tackling the problem of Instance Matching Track; more specifically for the Identity Recognition Task (Dragisic et al., 2014).

The validation method used for this experiment was k-fold cross-validation, with k=10. This method was selected since the amount of available instances (and thus patterns) is reasonably large, which in turn offers statistically sound performance measurements (Kohavi, 1995). Table 6 shows the experimental results obtained when the proposed method (including the novel classification algorithm) is used.

Table 6

Experiment 1 Results

Algorithm	Precision	Recall	F-measure
Our model	0.8802	0.7118	0.7871
kNN, k=3	0.8634	0.4236	0.5684
ANN	0.8254	0.5375	0.6510
BayesNet	0.7650	0.8259	0.7943
RTree	0.6055	0.2933	0.3952

As can be seen, the proposed model exhibits a better precision, as well as the second best f-measure, compared against other state of the art algorithms, such as those present in the WEKA platform (Hall et al., 2009). Notice that, even when lending some of the key aspects of the proposal to other classification algorithms (namely the statement of the OMP as a binary classification task and the pattern building scheme), the AM introduced here outperforms other classifier present in the current state of the art.

Even though the proposed model requires previous knowledge about the task to be performed (unlike the usual methods), this apparent drawback may be worth it. Figure 3 presents a comparison between the results obtained using the model of this paper and the results presented at the OAEI 2014 (Dragisic et al., 2014).
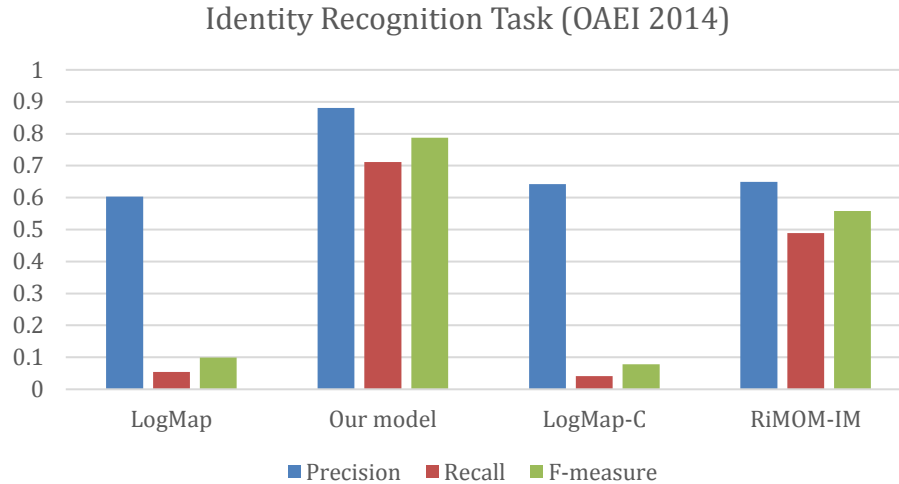
Identity Recognition Task (OAEI 2014)

Figure 3. OAEI 2014 comparison graphic.

The results shown in Figure 3 include the models with the best precision results, independently of other performance indicators such as recall and f-measure. The y axis represents either precision, recall, or the f-measure in a 0 to 1 scale, accordingly. In this case, the proposed method solves the OMP by means of the AM introduced here with higher precision, higher recall, and higher f-measure than the 3 best solutions found at the OAEI 2014 (Dragisic et al., 2014), according to precision.

2nd Experiment – LOM Repositories

The second experiment consists on doing a match between two different educative content repositories (ADRIADNE and MERLOT) in LOM format, based on a sample of 100 from each repository, related to the Computer Sciences topic.

The ADRIADNE Foundation (ADRIADNE, 2016) is a non–profit organization, which offers services intended to improve the creation and exchange of knowledge. One of the services offered by ADRIADNE is a LO repository; this service allows content consulting, content publishing, as well as digital content harvesting. One notable advantage of this service is the ability to convert the metadata of the objects into known standards, such as LOM and Doublin Core.

MERLOT is one of the largest open access repositories for educative contents, and is designed for use by researchers and teachers (MERLOT, 2016). MERLOT contains a collection of tens of thousands of teaching and learning resources, such as: animations, case studies, collections, questionnaires, simulators, among others.

In this experiment, a total of 100 1:1 matching instances were built from both ontologies. The features which were taken into account for the pattern building stage were: title, description, keywords, and type of resource. The results of this experiment were compared against other pattern classification algorithms, using the WEKA platform and the 10-fold cross-validation method (Kohavi, 1995).

Table 7 shows the 5 best results, according to precision. Again, the proposed model exhibits the highest precision among the methods tested. In addition, the results on recall and the f-measure are also the best.

Table 7

Experiment 2 Results

Algorithm	Precision	Recall	F-measure
Our model	0.9724	0.8576	0.9114
kNN, k=3	0.9664	0.3451	0.5086
ANN	0.9246	0.7237	0.8119
BayesNet	0.9152	0.5675	0.7006
RTree	0.8873	0.2452	0.3842

The results of the numerical experiments carried out show the good performance of the proposed algorithm, which clearly outperforms other proposals in terms of precision, recall, and f-measure. The accurate ontology matching allows a better use of the learning materials in open and distance learning repositories.

Thus, it may be said that the proposed AM is used as an OAM to join LOs taken from two different ODL ontologies expressed in the LOM format, outputting a new (as homogenous as possible) ontology with no redundant objects. This task is done by the proposed model in a competitive way, against either other known ontology alignment techniques, or other classification algorithms used in combination with the proposed method. This competitiveness was tested in two different scenarios: with a dataset built from an ontology matching contest, and on a dataset built by considering related OLs from two different repositories.

Conclusion

In this work, a new model for Ontology Matching over two educative content repositories is introduced, with the goal of helping to automatically improve the homogeneity of resources for open and distance learning environments. This task is fulfilled thanks to a novel pattern classification algorithm, which is also presented here. This approach to ontology entity alignment requires precious knowledge, yet the performance is higher than other methods since the classifier takes advantage of the similarity function to find additional information.

Given that the proposed classification algorithm solve only *c* equation systems (where c is the number of classes in the problem, 2 in this case) in order to determine the class corresponding to the test pattern, classification is done in one single step and does not require an iterative process. This, in turn, means that the convergence of the AM is guaranteed, since there is no possibility for the proposed model to not find an answer (either correct or incorrect) in the prescribed single step. This kind of one-shot classification has clear advantages with respect to other known iterative pattern classifiers, such as artificial neural network and SVMs, whose models have shown great precision, at the cost of high computational costs. In the case

of our model, the whole dataset is abstracted for training into c matrices of nxn dimensions, reducing the training space and facilitating the operations during the classification phase.

Future Work

There are several extension which can be clearly done to this work, namely to search for entity relationships with different cardinalities: 1:n, m:1, and m:n. Also, similar experiments can be run on other educative content repositories, which may even use different metadata standards. Other possible extensions include taking into account other metadata attributes, perhaps using other similarity computing strategies. Finally, it remains to do a cost-performance analysis of the proposed method on larger ontologies, particularly over ontologies which belong to the Big Data category.

Acknowledgements

References

ADRIADNE. (2016, March 2). Retrieved from http://www.adriadne-eu.org

Aher, S. B., & Lobo, L. M. R. J. (2013). Combination of machine learning algorithms for recommendation of courses in E-Learning System based on historical data. *Knowledge-Based Systems*, *51*, 1–14.

Atkinson, J., Gonzalez, A., Munoz, M., & Astudillo, H. (2014). Web metadata extraction and semantic indexing for learning objects extraction. *Applied Intelligence, 41*(2), 649–664.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American, 284*(5), 34–43.

Beydoun, G. (2009). Formal concept analysis for an e-learning semantic web. *Expert Systems with Applications*, *36*(8), 10952–10961.

Caliusco, M. L., & Stegmayer, G. (2010). Semantic web technologies and artificial neural networks for intelligent web knowledge source discovery. In Y. Badr, R. Chbeir, A. Abraham, & A.-E. Hassanien (Eds.), *Emergent web intelligence: Advanced semantic technologies* (pp. 17–36). London: Springer.

Dabbagh, N., Benson, A. D., Denham, A., Joseph, R., Al-Freih, M., Zgheib, G., …. Guo, Z. (2016). Massive open online courses. In N. Dabbagh, A. D. Benson, A. Denham, R. Joseph, M. Al-Freih, G. Zgheib,

.... Z. Guo (Eds.), *Learning technologies and globalization* (pp. 9-13). Heidelberg: Springer International Publishing.

Dragisic, Z., Eckert, K., Euzenat, J., Faria, D., Ferrara, A., Granada, R., ... & Montanelli, S. (2014, October). Results of the ontology alignment evaluation initiative 2014. In P. Shvaiko, M. Mao, J. Li, & A.-C. Ngonga Ngomo (Eds.), *Proceedings of the 9th International Conference on Ontology Matching-Volume 1317* (pp. 61-104).

Gladun, A., Rogushina, J., García-Sanchez, F., Martínez-Béjar, R., & Fernández-Breis, J. T. (2009). An application of intelligent techniques and semantic web technologies in e-learning environments. *Expert Systems with Applications*, *36*(2), 1922–1931.

Gooley, A., & Lockwood, F. (Eds.). (2001). *Innovation in open and distance learning: Successful development of online and web-based learning*. London: Routledge, Taylor & Francis.

Grossman, S. I., & Godoy, J. J. F. (2012). *Álgebra lineal*. México: McGraw-Hill.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *ACM SIGKDD Explorations Newsletter, 11*(1), 10-18.

Hariri, B. B., Abolhassani, H., & Sayyadi, H. (2006). A neural-networks-based approach for ontology alignment. *Proceedings of the Joint 3rd International Conference on Soft Computing and Intelligent Systems and 7th International Symposium on advanced Intelligent Systems,* Japan.

Houshmand, M., Naghibzadeh, M., & Araban, S. (2010). Reliability-based similarity aggregation in ontology matching. *Intelligent Computing and Intelligent Systems (ICIS), 2010 IEEE International Conference on* (Vol. 3), (pp. 744–749).

Karger, D. R. (2014). The semantic web and end users: What's wrong and how to fix it. *Internet Computing, IEEE, 18*(6), 64–70.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence* (Vol. 2), 1137–1145.

Learning Technology Standards Committee. (2002). IEEE Standard for learning object metadata. *IEEE Standard*, 1484.

Li, Z., Xu, W., Zhang, L., & Lau, R. Y. K. (2014). An ontology-based Web mining method for unemployment rate prediction. *Decision Support Systems*, *66*, 114–122.

Liu, J., Qin, L., & Wang, H. (2013). An ontology mapping method based on support vector machine. In *Proceedings of the 8th International Conference on Ontology Matching-Volume 1111* (pp. 225-226).

Liu, L., Yang, F., Zhang, P., Wu, J.-Y., & Hu, L. (2012). SVM-based ontology matching approach. *International Journal of Automation and Computing*, *9*(3), 306–314.

López-Yáñez, I., Argüelles-Cruz, A. J., Camacho-Nieto, O., & Yáñez-Márquez, C. (2011). Pollutants time-series prediction using the gamma classifier. *International Journal of Computational Intelligence Systems, 4*(4*)*. 680–711.

Luna, J. M., Romero, C., Romero, J. R., & Ventura, S. (2014). An evolutionary algorithm for the discovery of rare class association rules in learning management systems. *Applied Intelligence 42*(3), 501-513.

Mao, M., Peng, Y., & Spring, M. (2011). Ontology mapping: As a binary classification problem. *Concurrency and Computation: Practice and Experience*, *23*(9), 1010-1025.

Marques de Sá, J. P. (2001). *Pattern recognition: Concepts, methods, and applications*. Berlin Heidelberg: Springer-Verlag

MERLOT. (2016, March 2). Retrieved from http://www.merlot.org

Microsoft Bing API. (2016, March 2). Retrieved from http://www.bing.com/dev/en-us/translator

Mohammed, O., & Benlamri, R. (2014). Developing a semantic web model for medical differential diagnosis recommendation. *Journal of Medical Systems*, *38*(10).

Otero-Cerdeira, L., Rodríguez-Martínez, F. J., & Gómez-Rodríguez, A. (2015). Ontology matching: A literature review. *Expert Systems with Applications*, *42*(2), 949–971.

Pancerz, K., & Lewicki, A. (2014). Encoding symbolic features in simple decision systems over ontological graphs for PSO and neural network based classifiers. *Neurocomputing*, *144*, 338–345.

Peñalver-Martinez, I., Garcia-Sanchez, F., Valencia-Garcia, R., Rodríguez-García, M. Á., Moreno, V., Fraga, A., & Sánchez-Cervantes, J. L. (2014). Feature-based opinion mining through ontologies. *Expert Systems with Applications*, *41*(13), 5995–6008.

Pietranik, M., & Nguyen, N. T. (2014). A multi-attribute based framework for ontology aligning. *Neurocomputing*, *146*, 276–290.

Prinsloo, P., Archer, E., Barnes, G., Chetty, Y., & Van Zyl, D. (2015). Big (ger) data as better data in open distance learning. *The International Review of Research in Open and Distributed Learning, 16*(1).

Rubiolo, M., Caliusco, M. L., Stegmayer, G., Coronel, M., & Gareli Fabrizi, M. (2012). Knowledge discovery through ontology matching: An approach based on an Artificial Neural Network model. *Information Sciences*, *194*, 107–119.

Sachs, L. (1982). *Applied statistics: A handbook of techniques.* New York: Springer-Verlag.

Shvaiko, P., & Euzenat, J. (2013). Ontology matching: State of the art and future challenges. *IEEE Transactions on Knowledge and Data Engineering*, *25*(1), 158–176.

Tabaa, Y., & Medouri, A. (2013). LASyM: A learning analytics system for MOOCs. *International Journal of Advanced Computer Science and Applications, 4*(5), 113-119.

Tang, J., Li, J., Liang, B., Huang, X., Li, Y., & Wang, K. (2006). Using Bayesian decision for ontology mapping. *Web Semantics: Science, Services and Agents on the World Wide Web*, *4*(4), 243–262.

Thangaraj, M., & Sujatha, G. (2014). An architectural design for effective information retrieval in semantic web. *Expert Systems with Applications*, *41*(18), 8225–8233.

Vidal, J. C., Lama, M., Otero-García, E., & Bugarín, A. (2014). Graph-based semantic annotation for enriching educational content with linked data. *Knowledge-Based Systems*, *55*, 29–42.

W3C. (2016, March 2). Retrieved from http://www.w3c.org

Walters, T. N., Walters, L. M., Green, M. R., & Lin, L. H. (2016). Rich text, rich teach: Expanding educational horizons with technology in Malaysia. In I. Hussein Amzat, & B. Yusuf (Eds.), *Fast forwarding higher education institutions for global challenges* (pp. 11-24). Singapore: Springer Singapore.

Wang, J., Peng, J., & Liu, O. (2015). A classification approach for less popular webpages based on latent semantic analysis and rough set model. *Expert Systems with Applications*, *42*(1), 642–648.

Wen, C., & Zhang, J. (2015). Design of a microlecture mobile learning system based on smartphone and web platforms. *IEEE Transactions on Education, 58*(3), 203-207.

Yang, K., & Steele, R. (2009). Ontology mapping based on concept classification. *3rd IEEE International Conference on Digital Ecosystems and Technologies, 2009. DEST'09.* (pp. 656–661). IEEE.