

Article

A Hybrid Adaptive Educational eLearning Project based on Ontologies Matching and Recommendation System

Vasiliki Demertzis¹, Konstantinos Demertzis²

¹ International Hellenic University, Department of Computer Science, University Campus, Kavala, Greece, vademer@teiemt.gr;

² Democritus University of Thrace, School of Civil Engineering, University Campus, Kimmeria, Xanthi, Greece, kdemertz@fmenr.duth.gr;

* Correspondence: kdemertz@fmenr.duth.gr;

Abstract: The implementation of teaching interventions in learning needs has received considerable attention, as the provision of the same educational conditions to all students, is pedagogically ineffective. In contrast, more effectively considered the pedagogical strategies that adapt to the real individual skills of the students. An important innovation in this direction is the Adaptive Educational Systems (AES) that support automatic modeling study and adjust the teaching content on educational needs and students' skills. Effective utilization of these educational approaches can be enhanced with Artificial Intelligence (AI) technologies in order to the substantive content of the web acquires structure and the published information is perceived by the search engines. This study proposes a novel Adaptive Educational eLearning System (AEeLS) that has the capacity to gather and analyze data from learning repositories and to adapt these to the educational curriculum according to the student skills and experience. It is a novel hybrid machine learning system that combines a Semi-Supervised Classification method for ontology matching and a Recommendation Mechanism that uses a hybrid method from neighborhood-based collaborative and content-based filtering techniques, in order to provide a personalized educational environment for each student.

Keywords: Adaptive Educational System; E-Learning; Machine Learning; Semantics; Recommendation System; Ontologies Matching.

1. Introduction

The world wide web (www) today is an unruly construct, with a wide variety of styles. Specifically, last decade, the amount of www content dramatically increased that implies the need to manage and analyze big data volumes, which come from heterogeneous and often non-interoperable sources [1]. The semantic modeling of the www content in order to be perceived by the search engines is achieved with the Semantic Web (SWeb) technologies [2]. In addition, the management of these big volumes is further complicated by the need for high-security policies and privacy under the recent General Data Protection Regulation (GDPR) [3]. As the web evolves, the need for semantics technologies that focuses on the importance of the content are an important priority for the research communities.

Generally, the SWeb technologies “enable people to create data stores on the web, build ontologies, and write rules for handling data. Linked data are empowered by technologies such as RDF, SPARQL, OWL, and SKOS” refers to W3C’s vision of the web of linked data [4]. Ontologies are a complex, and possibly quite a formal collection of terms. Used to define and exemplify an area of concern and to organize the terms that can be used in a domain, characterize possible relationships, and define probable restrictions on using those terms [5]. With this approach, the search engines will

contribute to their more efficient collection and processing of useful web content to the setting up a new global educational system [6].

Modern education promotes teaching and learning through sophisticated methods. The precipitous evolution of the web and mobile devices has made eLearning adaptable, time-saving, and cost-effective in education process. Besides, since the early days of eLearning, its advantages and have significantly overshadowed those of face-to-face training, making distance education an crucial pillar of every new education and training system [7].

Also, the pandemic of Covid-19 that disrupted the education and training of an entire generation makes necessary the use of eLearning platforms for distance education. The distance education systems use modern communication and information technologies to achieve the essential two-way interaction to accelerate and support the educational process [8]. But the new trends in eLearning philosophy such as interactive videos, learning analytics, mobile-friendly online course platforms, virtual conferences, etc. [9], marks the transition to a new era, that needs to expand the learning process with more sophisticated educational opportunities throughout the life of individuals. The ternary relationship that develops between the instructor, the trainee, and the educational material replaces the dual relationship between the instructor and the trainee that until now characterized conventional education [10].

At the same time, the rapid development of the cloud computing, the SWeb methodologies, and especially the AI technologies, offer new opportunities in the future development of innovative systems that will allow the smarter management of learning content, for providing personalized educational environments [11].

The SWeb technologies are as much about the data as they are about reasoning and logic but does not deal with unstructured content. It is about representing not only structured data and links but also the meaning of the underlying concepts and relationships. For example, the RDF is the foundational technology in the SWeb stack, which is a flexible graph data model that does not involve logic or reasoning in any way. Even the parts of the SWeb technology stack that deal with reasoning and inference are grounded in well-understood formal semantics and can usually be expressed via straightforward sets of rules [5]. As such, they lack both the complexity and the opacity of AI approaches that are based on machine learning and neural models.

AI defined as "a system's ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation" [12]. Also, an AI system includes capabilities to learn from experience and connectivity and can adapt according to the current situation.

The most important developments concerning the combination of AI and SWeb in education and more specifically in the modern eLearning systems focus on:

1. In information management with appropriate ontologies for optimized performance. The use of ontologies in collaborative environments where collective content are produced, will allow correlations between heterogeneous sources (documents, emails, etc.) in order to easily retrieve all the absolutely relevant information.

2. In the digital libraries where they need to comply with the semantic ontologies and organize their librarian catalogs in a semantic way so that search engines can locate the appropriate content.

3. In the development of innovative applications and eLearning platforms, which using semantic ontologies, will allow the transform of distance education, creating friendly in search engines semantic "maps" of learning material and content.

AES, accepting the above wording, are new technologically supported education systems that adapt the provided educational content to the specific educational needs of each trainee or group of trainees in order to achieve sophisticated learning [6]. They also provide specialized support to the trainees taking into account the learning needs, the special characteristics of learners in addition to their evolution during their study [9].

The contribution of the SWeb and ontologies matching technologies, and especially the artificial intelligence in the development of a novel eLearning architecture, is the motivation of this paper. Specifically, this paper proposes a novel AEeLS, which with extensive use of AI methods, allows the

modeling of the process of retrieval and management of information based on semantic criteria, for the needs of individualized education of each student.

The sections appear in the rest of the paper in the following prescribed order as follows: Section 2 presents the related work about the relevant AES that have used machine learning methods. Section 3 describes the proposed model. Section 4 defines the methodology and finally, section 5 contains the conclusions.

2. Related Work

Online collaborative has highlighted the eLearning approaches as an essential part of modern educational system. Universities, organizations, and companies have adopted eLearning as a more flexible and effective way to train their students, executives, or employees. However, the current and future trends in eLearning prove that it is a field for continuous innovation and research.

There are some scientific works, related to several topics relevant to the development AEeLS of the present paper. For example, the work [13] discovers several tactics for educational metadata mining, whose one of the most important open challenges is the recognition of Learning Objects and the metadata that can be gained from them. Also, both Mao et al. [14] and Liu et al. [15] show how Ontology Matching can be specified as a binary classification problem, forcing use of most well known machine learning algorithms. In the former work, an approach for locating relationships between two ontologies using Support Vector Machines (SVM) is presented. The experimental results show promising are remarkable when contrasted against other mapping methods.

In addition, the paper [16] propose a novel ontology matching method that uses again SVMs, demonstrating a precision of the order of 95% in their investigational results.

Other research work [17], explore the ontology mapping problem based on concept classification by decision trees algorithms that introduces a similarity measure among two portions fitting to distinct ontologies. Nonetheless, the effort does not give analytical precision results, although claiming that the model produced is faster at execution due to the less evaluations needed.

A different approach presented by the [18] that introduce a graph-based semantic annotation method for enriching educational content with linked data, in order to gain document search with high recall and precision.

Metaheuristics have also had an important role in the vicinity of e-learning. In this sense, Luna et al. [19] propose an association paradigm for finding learning rules applying evolutionary metaheuristic algorithms.

Moreover, Peñalver-Martinez et al. [20] apply some natural language techniques to resources produced for opinion mining with remarkable results.

Also, Wang et al. [21] presents a classification method for less widespread webpages based on suppressed semantic analysis and difficult set patterns for the automated tagging of web pages with related content.

On the other hand, the investigation of smart recommendation systems, have noticed great recognition and usage in e-commerce platforms. Though, authors of [22] introduce an online courses recommendation system, which joins numerous clustering methods in order to prove that machine learning approaches can enhance significant the estimation process of courses immersed in e-learning environments.

Also, Gladun et al. [23], presents a multi-agent recommendation system for automatic feedback concerning knowledge obtained by students in e-learning platforms, taking advantage of the SWeb technologies.

Finally, other research methods on distance learning are focused on proposing a novel way of microlecture through mobile terminals and web platforms [24], while others focused on expanding educational horizons (Walters, Walters, Green, & Lin, 2016).

3. Proposed Framework

Since eLearning systems' methodology is an extremely complex process, trainers cannot be based only on the use of pathetic isolated content and products based solely on the old and maybe

obsolete educational materials. The content classification based on the student needs, should not be a manual and time-consuming process, something that will offer an important disadvantage to the education system. Following this point of view, the use of more effective methods of education supervision, with capabilities of automated control the educational content and use of specific materials for every student is important to every modern educational system.

It is also important the update the eLearning philosophy and its transformation into an Adaptive Educational eLearning System. The ideal AEeLS includes advanced AI solutions for real-time analysis of the educational needs both known and unknown students, instant reports, data visualization of progress, and other sophisticated solutions that maximize the education experience alongside with fully automated content evaluation process by semantic technologies.

Unlike other techniques that have been proposed in the literature focused on static approaches [16-17], the dynamic model of AEeLS produce a evolving educational tool without special requirements and computer resources.

The algorithmic approach of the proposed AEeLS includes in the first stage an Ontologies Matching process from www in order to find the relevant educational content as you can see in the depiction of the proposed model, in Figure 1. In the second stage, the content checked for the precision and accuracy and a Recommendation Mechanism proposes new relevant material in order to produce an extremely fitted curriculum for each student (stage 2 in Figure 1).

The following Figure 1 is a representation of the algorithmic approach of the proposed AEeLS model:

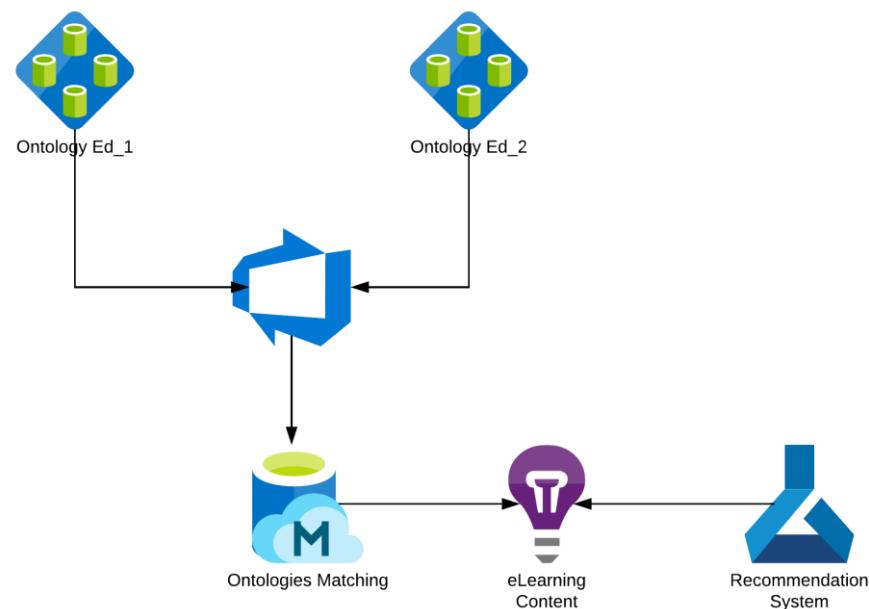


Figure 1. AEeLS model.

4. Methodology

4.1 Ontologies Matching

The ontologies are a formal structured information framework and a clear definition of a common and agreed conceptual formatting of properties and interrelationships of the entities that really exist in a particular domain of interest. The main components of the ontologies are classes, properties, instances and axioms. Classes exemplify adjusts of entities within a specific domain. Properties define the various attributes of concepts and constraints on these attributes. Both of them can be formed into separate hierarchies. Instances represent the concepts and axioms are assertions in the form of logic to constrain values for classes or properties [25].

Officially an ontology can be defined as below [26]:

$$O = \{C, P, H^C, H^P, I, A^O\} \quad (1)$$

where C and P denote classes and properties, H^C and H^P are the hierarchy of them, I is a set of instances and A^O is a set of axioms.

The proposed Ontologies Matching Mechanism (OMM) based on advanced computational intelligence and machine learning techniques. The aim is to develop a fully automated method for extracting information and controlling the effectiveness of student needs [27]. In particular, this subsystem automates the extraction, analysis, and interconnection of educational web content material based on relevant ontologies for further processing. It also allows for the effective detection of conflicting rules or content related to the transmission of personal data to ensure that they cannot be used to create a user profile or privacy leakages. To achieve this, ontology matching techniques using AI methods used.

Ontology matching is a hopeful method to the semantic heterogeneity dilemma. It uncovers correspondences among semantically linked entities of the ontologies. These correspondences can be applied for various tasks, such as ontology merging, query answering and data translation. Thus, matching ontologies allows the knowledge and data expressed in the paired ontologies to interoperate [28].

The aim of ontology matching is the procedure of establishing correspondences between concepts in ontologies to derive an alignment between two ontologies, where an alignment consists of a set of correspondences between their elements so that significant similarity can be equivalent. Given two ontologies O_s (source ontology) and O_t (target ontology) and an entity e_s in O_s , the procedure ontology matching M denoted as a process that find the entity e_t in O_t , that e_s and e_t are deemed to be equivalent [29].

It should be emphasized that the ontology matching process it can be subsumption, equivalence, disjointness, part-of or any user specified relationship. The most significant matchings or alignments can be categorized in three particular sections [30]:

1. Similarity vs Logic: This category concerns the similarity and logical equivalence among the ontology terms.
2. Atomic vs Complex: With regard to that category the alignment considers if it is "one-to-one", or "one-to-many".
3. Homogeneous vs Heterogeneous: In the third category, the alignments examines if it is on terms of the same type or not (e.g., classes to classes, individuals to individuals, etc.).

Usually, an ontology matching tactic applies several and different categories of matchers such as labels, instances, and taxonomy forms to recognize and calculate the similarity between ontologies. The easiest strategy is to aggregate the similarity values of each entity pair in a linear weighted fashion and decide on a suitable threshold to recognize matching and non-matching pairs. Though, given a matching condition, it is difficult to define the right weights for each matcher [30]. In recent past, many ontology matching methods and weighting strategies have been suggested to adaptively verify the weights such as Harmony [31] and Local Confidence [32], but there is no single strategy.

Against, the machine learning based ontology matching methods have been proved to get more accurate and reliable matching results [33]. Specifically, the supervised machine learning methods use a set of validated matching pairs as training examples, in order to apply a learning patterns strategy that can be find the right matches from all the candidate matching pairs. On the other hand, the unsupervised machine learning methods uses arbitrary and heuristic strategies to matching pairs without orderly and modeled methodology. Comparing the machine learning approaches, supervised methods usually get better results [33].

However, the main weakness of the methods with full supervision is that they need a substantial amount of labeled training examples to create a predictive model with acceptable performance. The training dataset is mostly accomplished manually by the trainer, which is a difficult and time-consuming procedure. In addition, the current method only give the similarity values purely as numeric features, without taking their critical characteristics into account [34].

As an alternative, the key characteristic of training with Semi-Supervised method is the creation of the robust model with the use of pre-classified along with unlabeled instances. This approach operates on the condition that the input patterns with and without labels, belong to the similar

marginal distribution, or they follow a common formation. Largely, unclassified data offer useful information for the discovery of the whole dataset data structure, while separately the sorted data are presenting in the learning procedure. Thus, even the most serious real-world problems can be developed successfully, based on the crucial oddities that describe them [34].

The OMM uses a semi-supervised learning ontology matching innovative approach. Provided a slight set of labeled matching entity pairs, the technique first utilizes the central relationships in the similarity area to enhance positive training instances. After receiving more training instances, a graph based semi-supervised learning algorithm is engaged to classify the rest applicant entity pairs into matched and non-matched classes. Finally, the suggested method define several constrictions to adapt the probability matrix in label propagation algorithm, which help to increase the performance of matching results [35].

The semi-supervised learning method is suitable for the OMM as ensures high-speed, vigorous and efficient classification performance. Moreover, it is easily adjustable and applicable method. Also, it is a pragmatic machine learning technique that can model the ontologies matching challenge based on a section of few pre-classified data vectors, exposing the relationships amongst the taxonomy constructions of ontologies [34-35].

Specifically, the OMM applies a hybrid model which employs well-established algorithms, optimally combined in order to create a faster and more flexible integrated Fuzzy Semi-Supervised Learning system. The most important innovation and advantage of the proposed approach is the easy validation of the classification process for a first time seen data, based on robust measurable factors. The theoretical background of the system's core is presented in the next paragraphs.

The naive Bayes classifier [36] is a practical learning method based on a probabilistic representation of a data structure, representing a set of random variables and their hypothetical independence, in which complete and combined probability distributions are substantiated. The objective of the algorithm is to classify a sample X in one of the given categories C_1, C_2, \dots, C_n using a probability model defined according to the theory of Bayes. These classifiers make probability assessment rather than forecasting, which is often more useful and effective. Here the projections have a score and the purpose is the minimization of the expected cost. Each category is represented by a prior probability.

We make the assumption that each sample X belongs to a class C_i and based on the Bayes theory we estimate the posteriori probability. The quantity P describing a naive Bayes classifier for a set of samples, expresses the probability that c is the value of the dependent variable C , based on the prices $x=(x_1, x_2, \dots, x_n)$ of the properties $X=(X_1, X_2, \dots, X_n)$ and it is given by the following relation (2) where the characteristics x_i are considered as independent [36]:

$$P(c|x) = P(c) \cdot \prod_i^n P(x_i|c) \quad (2)$$

The estimation of the above quantity for a set N examples is done by using the relations 3, 4 and 5:

$$P(c) = \frac{N(c)}{N} \quad (3)$$

$$P(x_i|c) = \frac{N(x_i, c)}{N(c)} \quad (4)$$

For a characteristic x_i with discrete values, the Probability is estimated by equation 5.

$$P(x_i|c) = g(x_i, \mu_c, \sigma_c) \quad (5)$$

where $N(c)$ is the number of examples that have the value c for the depended variable, $N(x_i, c)$ is the number of cases that have the values x_i and c for the characteristic X_i and the depended parameter respectively and $g(x_i, \mu_c, \sigma_c)$ is the Gaussian probability density function with an average value μ_c and variance σ_c for the characteristic x_i .

Collective classification [37] is a combinatorial optimization challenge, in which we are provided a set of intersections, $V = \{V_1, \dots, V_n\}$ and a neighborhood function N , where $N_i \subseteq V \setminus \{V_i\}$. Each node in V is a indiscriminate variable that can take a value from an applicable domain. V is additional divided into two sets of nodes: X , the observed variables and Y , the nodes whose values need to be defined. Our task is to label the nodes $Y_i \in Y$ with one of a small number of labels, $L = \{L_1, \dots, L_q\}$; we'll use the shorthand y_i to imply the label of node Y_i .

Also, according to Zadeh [38] every element “ x ” of the Universe of discourse “ X ” belongs to a Fuzzy Set (FS) with a degree of membership in the closed interval $[0,1]$. Thus, the subsequent function 6 is the mathematical base of a FS [38]:

$$S = \{(x, \mu_s(x)/\mu_s: X \{[0,1]: x\} \mu_s(x)\} \quad (6)$$

The next function 7 is a case of a normal Triangular Fuzzy Membership Function (FMF). It must be explained that the “ a ” and “ b ” factors have the values of the lower and upper bounds of the raw data individually [38]:

$$\mu_s(X) = \begin{cases} 0 & \text{if } X < a \\ (X - a)/(c - a) & \text{if } X \in [a, c] \\ (b - X)/(b - c) & \text{if } X \in [c, b] \\ 0 & \text{if } X > b \end{cases} \quad (7)$$

According to the typical (crisp) classification methods, each sample can be assigned only to one class. Thus, the class membership value is either 1 or 0. In general, classification methods reduce the dimensionality of a complex data set by grouping the data into a set of classes.

In fuzzy classification, a sample point can be assigned to many classes with a different degree of membership. The fuzzy c-means clustering algorithm initially gives random values to the cluster centers and then it assigns all of the data points to all of the clusters with varying Degrees of Membership (DoM) by measuring the Euclidean distance.

The Euclidean distance of each data point x_i from the center of each cluster c_1, \dots, c_j is calculated based on equation 8 [39].

$$d_{ji} = \|x_i - c_j\|^2 \quad (8)$$

where d_{ji} is the distance of x_i from the center of the cluster c_j

Then the DOM of each data point to each cluster is estimated based on equation 9:

$$\mu_j(x_i) = \frac{\left(\frac{1}{d_{ji}}\right)^{\frac{1}{m-1}}}{\sum_{k=1}^p \left(\frac{1}{d_{ki}}\right)^{\frac{1}{m-1}}} \quad (9)$$

where m is the fuzzification parameter with values in the interval $[1.25, 2]$ [39]. The values of m specify the degree of overlapping between the clusters. The default value of m is equal to 1.2. The algorithm has the following direct restriction in the DOM of each point [28]. See equation 10 [39]:

$$\sum_{j=1}^p \mu_j(x_i) = 1 \quad i = 1, 2, 3, \dots, k \quad (10)$$

where p is the number of the clusters, k is the number of the data points, x_i is the i -th point and $\mu_j(x_i)$ is a function that returns the degree of membership of point x_i in the j -th cluster $i=1, 2, \dots, k$.

Then the centers are estimated again.

The following equation 10 is used for the re-estimation of the values of new cluster centers [39]:

$$c_j = \frac{\sum_i [\mu_j(x_i)]^m x_i}{\sum_i [\mu_j(x_i)]^m} \quad (11)$$

where c_j is the center of the j -th cluster with $(j=1, 2, \dots, p)$, and x_i is the i -th point [39]. This is an iterative algorithm and the whole process is repeated till the centers are stabilized.

The OMM is an innovative hybrid algorithm based on the combination of soft computing approaches. Let us consider a supervised learning case with a training set of size N $\{X, Y\} = \{x_i, y_i\}_{i=1}^N$, where $x_i \in R^{n_i}$ and y_i is a binary vector of size n_o . It must be clarified that i and n_o are the dimensions of the input and output respectively.

The OMM initially performs Semi-Supervised Clustering (SSC). This means that cluster assignments may be already known for some subset of the data. The final aim is the classification of the unlabeled observations to the appropriate clusters, using the known assignments for this subset of the data. At the same time the algorithm produces the degree of membership of each record to its cluster.

The clustering validation process is performed by employing the “*classes to clusters*” (CL_A_U) method, that adopts SSC. Originally a minimum data sample is used comprising of the clusters derived from the SSC process (labeled data). The remaining unlabeled data are used to dynamically form and adjust the classes based on their DOM.

Actually, the CL_A_U approach assigns classes to the clusters, based on the majority value of the class attribute within each cluster. The class attribute is treated like any other attribute and it is a part of the input to the clustering algorithm.

The objective is the assessment as to whether the selected clusters match the specified class data. In the CL_A_U evaluation, you tell the system which attribute is a predetermined "class."

Then this is removed from the data before passing to the SSC algorithm. The CL_A_U evaluation, finds the minimum error of mapping classes to clusters (where only the class labels that correspond to the instances in a cluster are considered) with the constraint that a class can only be mapped to one cluster.

The emerged classes are fuzzified by assigning them proper Linguistics, in order to obtain a realistic coherence between the associated values of the dataset under study.

The whole process is presented in the Algorithm1 below.

Algorithm 1. The OMM Algorithm

Inputs: Input labeled data D_l , clusters of the labeled data L_l and a set of unlabeled data D_u

Step 1: % Initialization of clusters

Identify the discrete number of clusters based on L_l

For every cluster, create matrices with the mean and standard deviation of all D_l

Step 2: % Calculate the new centers of the clusters

For every cluster, recreate these matrices, based on the testing data D_u

Calculate a variable, based on the formula below:

$$x = (1. / (2 * \pi * ns.^2)) * \exp(-((test-nm).^2) / (2 * sn.^2))$$

where ns is the new standard deviation matrix, nm is the new mean matrix and test D_u

Sum all these variables for each cluster

Step 3: % Calculate the winner cluster for each record

For every testing data D_u , find the minimum value of the summary calculated before.

% Calculate the fuzzy membership values for every cluster for every record

For every testing data D_u and for every class, divide the mean matrix with the sum of the

values calculated before (normalization probability – membership value)

Outputs: Winner cluster for each testing data D_u , C_u and fuzzy membership values for every cluster

for every testing data D_u , $F_M_V_{u,j}$ (j the number of clusters)

Step 5: % Validation of the clustering process

Repeat Steps 1 – 3 from the previous part, only this time from $D_u \rightarrow D_l$, using C_u as labels

Output: Winner cluster for each testing data D_l , L_{2l}

Step 6:

For every initially labeled data D_l :

Compare the initial label L_l with L_{2l}

Create confusion matrix based on these comparisons

Step 7:

Repeat Steps 5 - 6 for every D_w of D_u

% Generalization of the amount of the extreme cases, based on the fuzzy membership values

Inputs: The winner class for every record (C_u) and the fuzzy membership values for each record ($F_M_V_{u,j}$)

Step 8:

For every record:

If max($F_M_V_{u,j}$) = A AND $F_M_V_{u,A} - max2(F_M_V_{u,j}) \leq threshold$, then

% max2($F_M_V_{u,k}$) = k, the second biggest membership value

Change the winner class for this record to k ($C_u = k$)

Outputs: Updated winner cluster for each record C_u

4.1. Recommendation Mechanism

The Recommendation Mechanism (RMm), is a computational intelligence and machine learning mechanism [40] in the AEeLS to create intelligent rules for intervention decisions and offer personalized real-time information for the students educational needs with Collaborative Filtering (CF) [41] technique.

CF is a machine learning method of making filtering about the conception by accumulating preferences or unique information from several users (collaborating). In the more general sense, CF is the method of filtering for information or patterns using procedures affecting collaboration between various agents, opinions, data resources, etc. Usually, a workflow of a CF can be defined as below [41]:

1. A user extracts the predilections by ranking objects of the system. These grades can be considered as an estimated description of the user's importance in the related domain.
2. The system match up this user's rankings compared to other users' and discovers the people with most "related" preferences.
3. With similar users, the system recommends items that the similar users have ranked highly but not yet being ranked by this user.

CF systems are separated in memory-based and model-based methods. Memory-based methods simply memorize the user preferences and issue recommendations based on the relationship between the new rating items and the rest of the ranking matrix. Model-based methods on the other hand fit a parameterized prototype to the given ranking matrix and then issue recommendations based on the tailored model [41].

The most popular and reliable CF methods are neighborhood-based methods, which predict ratings by referring to users whose ratings are similar to the closest training examples in the feature space. The most useful technique for this purpose is to allocate weight to the impacts of the neighbors, so that the nearer neighbors provide more to the average than the more distant ones. This is inspired by the hypothesis that if two users have similar grades on some items they will have similar grades on the remaining items and the opposite [42].

CF methods include cluster-based approaches [43], Bayesian techniques [44], Pearson correlation processes, vector similarity practices, regression strategies and error-based tactics [45].

Currently, CF methods have been applied to many kinds of systems including sensing and monitoring applications, environmental sensing over large areas, financial process and electronic commerce and web applications [42][45].

Traditional CF methods face two major challenges: data sparsity and scalability [42]. In the RMm, we use a hybrid method from neighborhood-based CF and content-based filtering that addressing these challenges and improve quality of recommendations [43].

The aim of this hybrid method trying to achieve more personalized intelligent rules for intervention decisions and personalized recommendation in real-time information for the student's educational needs based on skills. This hybrid method is more versatile, in the sense that they can be applied to heterogeneous ontologies and with some care could also provide cross-domain recommendations. Also, it works best when the user space is large, it is easy to implement, it scales well with no-correlated items and does not require complex tuning of properties [46].

5. Data

The proposed model of pattern classification was validated through tests, which were done on data taken from the Ontology Alignment Evaluation Initiative (OAEI) 2014 [47] campaign, as well as on data taken from two known educative content repositories: ADRIADNE [48] and MERLOT [49]. Thus, two datasets were built, containing patterns representing the relationships between pairs of Learning Objects taken from two different ontologies immersed in the Open and Distance Learning context.

For the first trial test according the [50], the OAEI 2014 data bank was used, for undertaking the problem of Instance Matching Track, more precisely for the Identity Recognition Task [47] and specifically is to find an appropriate similarity function, in order to build pairs of objects which are

actually close in meaning. Through the adequate use of a given similarity function, the ontologies matching problem transformed into a binary pattern classification problem.

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

The ADRIADNE Foundation offered a provision that is the capability to transform the metadata of the objects into known specifications, such as Learning Objects Metadata and Dublin Core.

MERLOT is one of the biggest open access warehouses for educative subjects and is created for use by research communities. Includes a gathering of learning resources and educational materials, such as: animations, case studies, collections, questionnaires, simulators, etc.

In this experiment according the [50], a total of 100 1:1 matching examples were constructed from both ontologies. The features extraction takes into account for the pattern structure: title, description, keywords, and type of resource.

The classification performance is estimated by the usual evaluation measures: Precision (PRE), Recall (REC) and F-Score indices that are defined as in equations 12, 13 and 14 respectively [51-52]:

$$PRE = \frac{TP}{TP + FP} \quad (12)$$

$$REC = \frac{TP}{TP + FN} \quad (13)$$

$$F - Score = 2 \times \frac{PRE \times REC}{PRE + REC} \quad (14)$$

The Precision rate shows what percentage of positive predictions where correct, whereas Recall measures what percentage of positive events were correctly predicted. The F-Score can be interpreted as a weighted average of the precision and recall. Consequently, this measure takes both false positives and false negatives into account. Subliminally it is not as straightforward to comprehend as accuracy, but F-Score is generally more valuable than accuracy and it works best if false positives and false negatives have similar cost, in this case.

Also, the validation method used the 10-fold cross-validation method because the quantity of available examples is relatively larger, which in turn offers statistically sound performance measurements [51-52].

The following table 1, presents an wide evaluation for both datasets, by engaging competitive methods namely: Radial Basis Function Neural Network (RBFNN), Group Method of Data Handling (GMDH), Polynomial Neural Networks (PNN), Feedforward Neural Networks using Genetic Algorithms (FFNN-GA), Feedforward Neural Networks using Particle Swarm Optimization (FFNN-PSO), SVM and Random Forest (RF).

Table 1. Comparison between algorithms (1st experimental test)

OAEI 2014 data bank			
Classifier	PRE	REC	F-Score
OMM	0.904	0.908	0.906
RBFNN	0.710	0.700	0.709
GMDH	0.845	0.846	0.848
PANN	0.813	0.818	0.817
FFNN-GA	0.887	0.888	0.889
FFNN-PSO	0.891	0.889	0.892
SVM	0.895	0.897	0.897
RF	0.900	0.900	0.901

Table 2. Comparison between algorithms (2nd experimental test)

ADRIADNE and MERLOT			
Classifier	PRE	REC	F-Score
OMM	0.981	0.981	0.982
RBFNN	0.888	0.889	0.889
GMDH	0.940	0.942	0.946
PANN	0.901	0.902	0.902
FFNN-GA	0.963	0.962	0.962
FFNN-PSO	0.965	0.964	0.964
SVM	0.976	0.977	0.976
RF	0.975	0.976	0.978

Tables 1 and 2 demonstrates obviously that the proposed method has superior performance for both datasets which is quite promising contemplating the complexities faced in this problem. It is crucial to say that evaluating several factors that can define a type of challenge discussed here is a partly subjective non-linear and dynamic procedure.

6. Conclusions

6.1 Discussion

This work presented a hybrid [53-56], innovative [57], reliable [58-59] and highly effective eLearning system that has the capacity to gather and analyze data from learning repositories and to adapt these to the educational curriculum according to the student skills and experience, based on sophisticated computational intelligence methods [60]. The AEeLS is a clearly innovative effort to effectively analyze and recommend relevant educational content based on semantic ontologies techniques. The proposed method is based on the optimal combination of the OMM and the RMm algorithms, which ensures the adaptation of the system in new situations. It offers high level of generalization, by implementing a robust algorithm capable to respond to high complexity problems. The performance of the proposed algorithm was tested on two multidimensional datasets of high complexity. These data sets emerged as a result of an extensive research on the function of ontologies. They realistically state the operating modes of these devices in normal conditions and in situations where they are subject to modern educational systems and needs. The results have proven the efficiency of the developed hybrid model.

6.2 Innovation

An important innovation of AEeLS is the use of hybrid learning techniques capable to solve a multi-dimensional and complex problem. The proposed system simulates in a realistic way the functioning of biological knowledge, the practical mode of human memory, and more commonly, the ways in which the brain models use the skills and experiences.

Also, an important improvement is the partition of the OMM and the RMm to relocate the expertise in the eLearning system. This method significantly enriches the way in which the learning extraction techniques work, as it generates the likelihood of forming heterogeneous systems to which learning transfer can be applied.

Finally, it should not be overlooked that an similarly valuable invention is the fact of combining AI to the level of an educational eLearning system. This fact considerably improves the performance

of modern educational systems. This innovation provides important solutions and improves the way eLearning systems work and respond to new generation.

6.3 Future Work

Future research will focus in further optimization of the algorithm's parameters that may result in a faster and more accurate performance. We will work on the improvement of the AEeLS complexity in a high understandable and adjustable level.

Further optimization by means of self-improvement and autolearning can be explored to fully automate the process of detecting relevant educational content. Finally, a very important future improvement is the extension of the algorithm for Natural Language Processing (NLP) capabilities, with Recurrent Neural Network (RNN) and specifically with deep architectures such as Long-Short Term Memory (LSTM), in order to approach and model time sequences and their broader dependencies with greater accuracy and efficiency.

Author Contributions: Conceptualization, V.D. and K.D.; methodology, V.D. and K.D.; software, V.D. and K.D.; validation, V.D. and K.D.; formal analysis, V.D. and K.D.; investigation, V.D. and K.D.; resources, V.D. and K.D.; data curation, V.D. and K.D.; writing—original draft preparation V.D. and K.D.; writing—review and editing, V.D. and K.D.; visualization, V.D. and K.D.; supervision, K.D.; project administration, K.D.; funding acquisition, V.D. and K.D. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. 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).
2. 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.
3. <https://eur-lex.europa.eu/eli/reg/2016/679/oj>
4. <https://www.w3.org/standards/semanticweb/>
5. Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American*, 284(5), 34–43.
6. Beydoun, G. (2009). Formal concept analysis for an e-learning semantic web. *Expert Systems with Applications*, 36(8), 10952–10961.
7. Gerber, A. J., Van der Merwe, A., Barnard, A. (2008). A Functional Semantic Web Architecture. *European Semantic Web Conference 2008, ESWC'08, Tenerife, June 2008*.
8. Gooley, A., & Lockwood, F. (Eds.). (2001). *Innovation in open and distance learning: Successful development of online and web-based learning*. London: Routledge, Taylor & Francis.
9. Masud, M. (2016). Collaborative e-learning systems using semantic data interoperability. *Computers in Human Behavior*. Vol. 61, pp.127–135.
10. 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.
11. 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.
12. Kaplan Andreas and Michael Haenlein (2019) Siri, Siri in my Hand, who is the Fairest in the Land? On the Interpretations, Illustrations and Implications of Artificial Intelligence, *Business Horizons*, 62(1), 15-25
13. 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
14. Mao, M., Peng, Y., & Spring, M. (2011). Ontology mapping: As a binary classification problem. *Concurrency and Computation: Practice and Experience*, 23(9), 1010-1025.
15. 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.

16. 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).
17. 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.
18. 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.
19. 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.
20. Peñalver-Martinez, I., Garcia-Sánchez, 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.
21. 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.
22. 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.
23. Gladun, A., Rogushina, J., García-Sánchez, 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.
24. 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.
25. Katifori, A.; Halatsis, C.; Lepouras, G.; Vassilakis, C.; Giannopoulou, E. (2007). "Ontology Visualization Methods - A Survey" (PDF). *ACM Computing Surveys*. 39:10. doi:10.1145/1287620.1287621.
26. Li, J., Tang, J., Li, Y., Luo, Q.: Rimom: a dynamic multistrategy ontology alignment framework. *IEEE Trans. Knowl. Data Eng.* 21(8), 1218-1232 (2009).
27. Min H., Mabahi H., Vukomanovic S., Irvin K., Krasniqi I., Avramovic S., and Wojtusiak J., "Applying an Ontology-guided Machine Learning Methodology to SEER-MHOS Dataset," 2016 Bio-ontology at Intelligent Systems for Molecular Biology (ISMB), Orlando, Florida, July 8-9, 2016.
28. Min H., Mabahi H., Vukomanovic S., Irvin K., Krasniqi I., Avramovic S., Wojtusiak J., "Ontology applications in Machine Learning", 2016 Bio-ontology at Intelligent Systems for Molecular Biology (ISMB), Orlando, Florida, July 8-9, 2016.
29. Tang, J., Li, J., Liang, B., Huang, X., Li, Y., Wang, K.: Using bayesian decision for ontology mapping. *Web Semant.* 4(4), 243-262 (2006).
30. Euzenat, J., Shvaiko, P.: *Ontology Matching*, 1st edn. Springer, New York (2007).
31. Mao, M., Peng, Y., Spring, M.: An adaptive ontology mapping approach with neural network based constraint satisfaction. *Web Semant. Sci. Serv. Agents World Wide Web* 8(1), 14-25 (2010)
32. Isabel, F.P.A., Cruz, F., Stroe, C.: Efficient selection of mappings and automatic quality-driven combination of matching methods. In: *Workshop on Ontology Matching*, pp. 49-60 (2009).
33. Eckert, K., Meilicke, C., Stuckenschmidt, H.: Improving ontology matching using meta-level learning. In: Aroyo, L., Traverso, P., Ciravegna, F., Cimiano, P., Heath, T., Hyvonen, E., Mizoguchi, R., Oren, E., Sabou, M., Simperl, E. (eds.) *ESWC 2009. LNCS*, vol. 5554, pp. 158-172. Springer, Heidelberg (2009).
34. Wang Z. (2014) A Semi-supervised Learning Approach for Ontology Matching. In: Zhao D., Du J., Wang H., Wang P., Ji D., Pan J. (eds) *The Semantic Web and Web Science. CSWS 2014. Communications in Computer and Information Science*, vol 480. Springer, Berlin, Heidelberg.
35. Zhu, X., Ghahramani, Z., Lafferty, J.: Semi-supervised learning using Gaussian fields and harmonic functions. In: *ICML*, pp. 912-919 (2003).
36. Bchir O, Frigui H, Ismail MMB (2013) Semi-supervised fuzzy clustering with learnable cluster dependent kernels. *International Journal on Artificial Intelligence Tools*, 22(3):1-26. Article number 1350013 doi: <http://dx.doi.org/10.1142/S0218213013500139>
37. Sen P, Namata G, Bilgic M, Getoor L, Galligher B, Eliassi-Rad T (2008) Collective Classification in Network Data. *Advancement of Artificial Intelligence*, 29(3):93-106.
38. Kecman V. (2001) *Learning and Soft Computing*. MIR Press. ISBN: 9780262112550

39. Cox E (2005) Fuzzy Modeling and Genetic Algorithms for Data Mining and Exploration. Elsevier Science, USA.
40. Francesco Ricci and Lior Rokach and Bracha Shapira, Introduction to Recommender Systems Handbook, Recommender Systems Handbook, Springer, 2011, pp. 1-35.
41. I. Bartolini, Z. Zhang, and D. Papadias, "Collaborative filtering with personalized skylines," IEEE Trans. Knowl. Data Eng., vol. 23, no. 2, pp. 190203, Feb. 2011.
42. Z. Yang, B. Wu, K. Zheng, X. Wang and L. Lei, "A Survey of Collaborative Filtering-Based Recommender Systems for Mobile Internet Applications," in IEEE Access, vol. 4, pp. 3273-3287, 2016. doi: 10.1109/ACCESS.2016.2573314
43. R. Hu, W. Dou, and J. Liu, "ClubCF: A clustering-based collaborative filtering approach for big data application," IEEE Trans. Emerg. Topics Comput., vol. 2, no. 3, pp. 302:313, Sep. 2014
44. N. Sherif and G. Zhang, "Collaborative filtering using probabilistic matrix factorization and a Bayesian nonparametric model," 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA) (Beijing, 2017, pp. 391-396, doi: 10.1109/ICBDA.2017.8078847
45. Xiaoyuan Su, Taghi M. Khoshgoftaar, A survey of collaborative filtering techniques, Advances in Artificial Intelligence archive, 2009.
46. Ya-Yueh Shih and Duen-Ren Liu, "Hybrid Recommendation Approaches: Collaborative Filtering via Valuable Content Information," Proceedings of the 38th Annual Hawaii International Conference on System Sciences, 2005, pp. 217b-217b, doi: 10.1109/HICSS.2005.302.
47. 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).
48. ADRIADNE. (2020, June 21). Retrieved from <http://www.adriadne-eu.org>
49. MERLOT. (2016, June 21). Retrieved from <http://www.merlot.org>
50. Sergio Cerón-Figueroa, Itzamá López-Yáñez, Yenny Villuendas-Rey, Oscar Camacho-Nieto, Mario Aldape-Pérez, and Cornelio Yáñez-Márquez, Instance-Based Ontology Matching For Open and Distance Learning Materials, International Review of Research in Open and Distributed Learning Volume 18, Number 1
51. Mao, J., Jain, A.K., Duin, P.W.: Statistical pattern recognition: A review. IEEE Trans. on Pattern Analysis and Machine Intelligence 22(1), 4-37 (2000).
52. Fawcett, T., (2006), An introduction to ROC analysis. Pattern Recognition Letters. Elsevier Science Inc. 27(8), 861-874. doi: <http://doi.org/10.1016/j.patrec.2005.10.010>.
53. Anezakis VD., Iliadis L., Demertzis K., Mallinis G. (2017) Hybrid Soft Computing Analytics of Cardiorespiratory Morbidity and Mortality Risk Due to Air Pollution. In: Dokas I., Bellamine-Ben Saoud N., Dugdale J., Díaz P. (eds) Information Systems for Crisis Response and Management in Mediterranean Countries. ISCRAM-med 2017. Lecture Notes in Business Information Processing, vol 301. Springer, Cham. https://doi.org/10.1007/978-3-319-67633-3_8.
54. Iliadis L., Anezakis VD., Demertzis K., Spartalis S. (2018) Hybrid Soft Computing for Atmospheric Pollution-Climate Change Data Mining. In: Thanh Nguyen N., Kowalczyk R. (eds) Transactions on Computational Collective Intelligence XXX. Lecture Notes in Computer Science, vol 11120. Springer, Cham. https://doi.org/10.1007/978-3-319-99810-7_8.
55. Anezakis VD., Demertzis K., Iliadis L., Spartalis S. (2016) A Hybrid Soft Computing Approach Producing Robust Forest Fire Risk Indices. In: Iliadis L., Maglogiannis I. (eds) Artificial Intelligence Applications and Innovations. AIAI 2016. IFIP Advances in Information and Communication Technology, vol 475. Springer, Cham. https://doi.org/10.1007/978-3-319-44944-9_17
56. Anezakis, V., Demertzis, K., Iliadis, L. et al. Hybrid intelligent modeling of wild fires risk. Evolving Systems 9, 267–283 (2018). <https://doi.org/10.1007/s12530-017-9196-6>
57. Demertzis Konstantinos, Iliadis Lazaros, Anezakis Vardis-Dimitrios, Commentary: *Aedes albopictus* and *Aedes japonicus*—two invasive mosquito species with different temperature niches in Europe, Frontiers in Environmental Science, VOLUME 5, YEAR 2017, PAGES 85, DOI:10.3389/fenvs.2017.00085
58. K. Demertzis, L. Iliadis and V. Anezakis, "A deep spiking machine-hearing system for the case of invasive fish species," 2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA), Gdynia, 2017, pp. 23-28, doi: 10.1109/INISTA.2017.8001126.

59. Konstantinos Demertzis, Lazaros S. Iliadis & Vardis-Dimitris Anezakis (2018) Extreme deep learning in biosecurity: the case of machine hearing for marine species identification, *Journal of Information and Telecommunication*, 2:4, 492-510, DOI: 10.1080/24751839.2018.1501542
60. Demertzis, K.; Kikiras, P.; Tziritas, N.; Sanchez, S.L.; Iliadis, L. The Next Generation Cognitive Security Operations Center: Network Flow Forensics Using Cybersecurity Intelligence. *Big Data Cogn. Comput.* 2018, 2, 35.