Chapter 7
Performance Evaluation of Decision-Based Content Selection Approaches in Adaptive Educational Hypermedia Systems

Pythagoras Karampiperis[1,2] and Demetrios G. Sampson[2,3]

[1] National Center of Scientific Research "Demokritos"
P. Grigoriou & Neapoleos Str., GR-15310 Aghia Paraskevi Attikis, Greece
pythk@iit.demokritos.gr
[2] Department of Digital Systems, University of Piraeus
150 Androutsou Street, Piraeus, GR-18534, Greece
[3] Centre for Research and Technology Hellas
6th Klm. Charilaou - Thermi Road, Thermi, Thessaloniki, GR - 57001, Greece
sampson@iti.gr

Abstract. Adaptive content selection is recognized as a challenging research issue in adaptive educational hypermedia systems (AEHS). In order to adaptively select learning objects (LO) in AEHS, the definition of adaptation behavior, referred to as Adaptation Model (AM), is required. Several efforts have been reported in literature aiming to support the AM design by providing AEHS designers with either guidance for the direct definition of adaptation rules, or semi-automated mechanisms which generate the AM via the implicit definition of such rules. The goal of the semi-automated, decision-based approaches is to generate a continuous decision function that estimates the desired AEHS response, aiming to overcome the problems of insufficiency and/or inconsistency in the defined adaptation rule sets. Although such approaches bare the potential to provide efficient AM, they still miss a commonly accepted framework for evaluating their performance. In this chapter, we discuss a set of performance evaluation metrics that have been proposed by the literature for validating the use of decision-based approaches in adaptive LO selection in AEHS and assess the use of these metrics in the case of our proposed statistical method for estimating the desired AEHS response.

7.1 Introduction

AEHS have been proposed as the underlying facilitator for personalized web-based learning with the general aim of personalizing learning experiences for a given learner (De Bra 2006, Knutov et al. 2009).

A. Peña-Ayala (Ed.): Intelligent and Adaptive ELS, SIST 17, pp. 161–182.
springerlink.com

In order to adaptively select and sequence LO in AEHS, that is, content objects described with educational metadata (McGreal 2004, Harman and Koohang 2006), the definition of adaptation behavior is required (Nejdl and Brusilovsky 2004). The AM contains the rules for describing the runtime behavior of the AEHS. In the literature, there exist different approaches aiming to support the AM design by providing AEHS designers with either guidance for the direct definition of adaptation rules, such as Authoring Task Ontology - ATO (Aroyo and Mizoguchi 2004), My Online Teacher - MOT (Cristea and Kinshuk 2003, Cristea 2007) and ACCT (Dagger et al. 2005), or semi-automated mechanisms which generate the AM via the implicit definition of such rules (Karampiperis and Sampson 2005, Huang et al. 2008, Ras and Ilin 2008).

The main drawback of the direct definition of adaptation rules is that there can be cases during the run-time execution of AEHS where no adaptation decision can be made due to insufficiency and/or inconsistency of the defined adaptation rule sets (Wu and De Bra 2001, Brusilovsky et al. 2007). This is due to the fact that, even if appropriate resources exist in the media space, the absence of a required rule (insufficiency problem) or the conflict between two or more rules (inconsistency problem), prevents the AEHS to select and use them in the generated learning resource sequence. As a result, either less appropriate resources are used from the media space, or required concepts are not covered at all by the resulting sequence (Wu and De Bra 2001).

The goal of the semi-automated approaches is to generate a continuous decision function that estimates the desired AEHS response, overcoming the above mentioned problem (Karampiperis and Sampson 2004). To achieve this, they use data from the implicit definition of sample adaptation rules and attempt to fit the response function on these data. Although such approaches bare the potential to provide efficient AMs, they still miss a commonly accepted framework for evaluating their performance.

This chapter is structured as follows: First, we discuss issues related with the AM design in AEHS focusing on the different approaches used in the literature for the definition of content selection rules. Then, we discuss the performance evaluation metrics that have been proposed by the literature for validating the use of decision-based approaches. Moreover, we present a performance evaluation methodology for decision-based content selection approaches in AEHS, and set up and report simulation-based experiments, following the above mentioned methodology, which aim to validate these evaluation metrics within the framework of our previously proposed statistical method for estimating the desired AEHS response. Finally, we discuss our findings and the conclusions that can be offered.

7.2 Overview of AEHS

Current state-of-the-art AEHS such as AHA! (Stash et al. 2007), OntoAIMS (Aroyo et al. 2003), The Personal Reader (Dolog et al. 2004), WINDS (Kravcik and Specht 2004), ACCT (Dagger et al. 2005) follow an architectural approach that fully implements the core structural elements defined by (Henze and Nejdl 2004) their AEHS definition.

This architecture is a variation of the Adaptive Hypermedia Application Model (AHAM) (De Bra et al. 1999) and consists of two main layers, namely, the run-time layer which contains the adaptation engine that performs the actual adaptation and the design layer. AM design (Brusilovsky and Henze 2007) involves defining:

- Concept selection rules which are used for selecting appropriate concepts from the domain model to be covered,
- Content selection rules which are used for selecting appropriate resources from the media space,
- Sequencing rules which are used for generating appropriate "learning paths" (that is, sequences of LO) for a given learner.

Typically, adaptive educational hypermedia sequencing is based on two main processes, namely, the *concept selection process* and the *content selection process*. In the concept selection process, a set of learning goals from the learning goals hierarchy is selected by the learner e.g. the AIMS (Aroyo and Mizoguchi 2004), or in some cases by the designer of the AEHS e.g. INSPIRE (Papanikolaou et al. 2003). For each learning goal, related concepts from the domain concept ontology are selected. In the content selection process, learning resources for each concept are selected from the media space based on the content selection rules. Typical AEHS examples that utilize this process are the MOT (Cristea and Kinshuk 2003, Cristea 2007), the ApeLS (Conlan et al. 2002), and the ELM-ART (Brusilovsky 2007).

The most commonly used approach for the definition of content selection rules by the AEHS designers team is the direct definition. In this approach, the content selection rules are set by the instructional designer during the design process and they are based on the items of the user model and the resource description model.

As already discussed, the main drawback of the direct definition of adaptation rules is that there can be cases during the run-time execution of AEHS where no adaptation decision can be made due to insufficiency and/or inconsistency of the defined adaptation rule sets. To this end, in the literature, two main approaches have been proposed to overcome these problems.

The first approach uses adaptation patterns (or templates) that have been a priori defined by an instructional designer during the design phase of the AEHS. These patterns hold the content selection rules of the AM. Typical examples of these systems are MOT (Cristea and Kinshuk 2003, Cristea 2007) and ACCT (Dagger et al. 2005).

Although this approach provides a solution to the inconsistency problem, it does not tackle with the problem of insufficiency, since that would require a huge set of patterns, which is difficult to be a priori defined. The problem of defining adaptation rules is a combinatorial problem, which means that in order to design sufficient and consistent adaptation rule sets, all the combinations of the adaptation decision variables should be covered. However, these combinations can be millions (Karampiperis and Sampson 2005), leading to huge rule sets that is difficult to author, manage and verify their sufficiency and/or consistency.

An alternative approach is the use of semi-automated decision based mechanisms (Karampiperis and Sampson 2005, Alfonseca et al. 2007, Huang et al. 2007, Hsieh et al. 2008), which generate a continuous decision function that estimates the desired AEHS response. To achieve this, they use data from the implicit definition of sample adaptation rules and attempt to fit the response function on these data. This definition of implicit adaptation rules, is given in the form of model adaptation decisions, over which the adaptation response function should be fit. This approach overcomes both the problems of sufficiency and consistency; however it introduces decision errors that result from the decision function fitting errors during the machine learning process (Karampiperis and Sampson 2005).

Sect. 7.3 presents the evaluation metrics given in the literature for evaluating the performance of decision-based adaptive content selection and discusses them.

7.3 Performance Evaluation Metrics for Decision-Based AEHS

We focus on the performance evaluation metrics used in semi-automated decision-based approaches for adaptive content selection. Performance evaluation in this context means: measuring how well a semi-automated approach fits the decision function to the provided model adaptation decisions (training data), and how well this decision function responds to decision cases not known during the training process (generalization capacity). As a result, model adaptation decisions are divided into two sets: the training dataset, which is used for evaluating the performance during the training of the semi-automated approach, and the generalization dataset, which is used for measuring the generalization capacity of the approach. Performance evaluation is the comparison result between the expected system output and the estimated AEHS response over the above mentioned datasets.

In adaptive content selection several approaches are proposed in the literature. The most commonly used are the following (Sampson and Karampiperis 2011):

Concept/keyword-based selection: In this approach, searching is made based on keywords representing the desired concepts to be covered from the retrieved LO. In AEHS, these keywords are set over the domain concept ontology at the concept selection process. The ranking of LO is done using a concept/keyword-based similarity formulae (Lee et al. 2006, Biletskiy et al. 2009), which evaluates the relevance of each LO, by comparing the desired concepts/keywords with the classification metadata used for describing the LO in hand.

The main assumption of this approach is that the domain concept ontology and the classification metadata used for the LO share the same concept/keyword terms. However, this is not always true, especially in domains where there exist a variety of classification models which use different terminology for describing a concept depending on the context of use, i.e. in the medical domain there exist many classification systems such as Medical Subject Headings (MeSH), the International Classification of Primary Care (ICPC) etc. targeting different end-users. An alternative approach proposed by (Kiu and Lee 2007), uses unsupervised data-mining techniques for estimating the match between the desired concepts/keywords with the classification metadata used for describing the LO in hand. This approach provides better results from the use of keyword-based similarity formula when different classifications models are used, but it requires significantly more time for the content selection process.

Preference-based selection: In these approaches, selection is performed based on the comparison of the learner profile in hand with the metadata description of the LO. In this case, the ranking of LO is performed using a preference score (Karampiperis and Sampson 2004, Wang et al. 2007, Dolog et al. 2008), which evaluates the utility/suitability of each LO for the learner profile in hand.

In both techniques, the concept/keyword-based and the preference-based selection, general purpose evaluation metrics are used from the field of information extraction (Ochoa and Duval 2008). More specifically, *precision* and *recall* measures are applied in order to evaluate the effectiveness of the LO selection technique, in terms of accuracy and completeness respectively. Precision is the ratio of correct responses to the sum of correct and incorrect responses, and is defined by the equation (7.1) (Wang et al. 2007, Biletskiy et al. 2009):

$$\text{Precision} = \left(\frac{\#\, retrieved\ relevant\ LOs}{\#\, retrieved\ LOs} \right) \tag{7.1}$$

Recall is the number of correct system responses divided by the sum of correct, incorrect and missing system responses, and is defined by the equation (7.2) (Wang et al. 2007, Biletskiy et al. 2009):

$$\text{Recall} = \left(\frac{\#\, retrieved\ relevant\ LOs}{\#\, relevant\ LOs} \right) \tag{7.2}$$

In order to have a single evaluation metric, *F-measure* is used, which is a weighted combination of recall and precision, and is defined by the equation (7.3) (Biletskiy et al. 2009):

$$\text{F - measure} = \left(\frac{2 * precision * recall}{precision + recall} \right) \tag{7.3}$$

However, AEHS implement a content selection strategy which limits the number of retrieved LO, aiming to restrict the amount of information provided to learners at a given time instance, due to the problem of learners' cognitive overload (Brusilovsky 2007). As a result, the precision should be measured not on the entire media space, but only on the desired sub-space which represent a set of the *n* most preferred LO, where *n* is the number of the desired LO. If not, the resulting precision would be higher or equal to the real one, since the number of retrieved LO is less or equal to the number of desired LO at a given time instance.

Moreover, since the resulting LO space is restricted, the recall measure should also be measured over the space of the *n* most relevant LO, and not over the space of all relevant LO. This introduces the need for an alternative evaluation metric in adaptive content selection. In (Karampiperis and Sampson 2004), the *Selection Success* (SS) evaluation metric has been proposed as follows in (7.4):

$$\mathrm{SS}\,(\%) = 100 * \left(\frac{\#\, correct\ ranked\ LOs}{\#\, requested\ LOs} \right) \tag{7.4}$$

Although this metric seems similar to the precision metric (PM) in information retrieval systems, its difference is critical. It evaluates the precision of selecting LO not on the entire space of the Media Space, but only on the desired sub-space, and also takes into consideration the ranking of the selection process. This means that the proposed metric is stronger, since it measures the precision over a smaller value space.

7.4 Evaluation Methodology for Decision-Based AEHS

The underlying hypothesis of the design of a decision-based approach for content selection in AEHS is that it is feasible to construct a semi-automated algorithm, which generates a continuous decision function that estimates the desired AEHS response, aiming to overcome the above mentioned problems of insufficiency and inconsistency of the defined adaptation rule sets.

Thus, the goal of evaluating such an approach is twofold: first, to examine whether a proposed semi-automated decision based approach is capable of extracting decision models which replicate the AM of existing AEHS; and second, to verify via performance evaluation that this approach can be applied in cases where large-scale adaptation rule sets are needed to describe the desired AEHS response. To this end, the evaluation should be performed in two phases:

Phase A: Extracting the AM of existing AEHS. In this evaluation phase, the AM rules of existing AEHS are used for generating sample adaptation decisions. These decisions have the form of combinations of LO mapped to learner profiles, and are used to train the intelligent mechanism that fits the response function on these data.

The goal of this phase is to examine whether the proposed semi-automated decision based approach is capable of extracting the decision model of the AEHS in hand. In our experiments, we will try to extract the AM rules for content selection used in the INSPIRE (Papanikolaou et al. 2003) system.

Phase B: Scaling up the experiments. As already discussed, the problem of defining adaptation rules is a combinatorial problem, which means that in order to design sufficient and consistent adaptation rule sets, all the combinations of the adaptation decision variables should be covered. However, these combinations can be millions (Karampiperis and Sampson 2005), leading to huge rule sets that is difficult to author, manage and verify their sufficiency and/or consistency. To this end, in order to keep the adaptation rule set human-maintainable, existing AEHS in the literature use few adaptation variables, typically 2-4 variables for describing learners' behavior and 2-3 variables for describing educational content. The goal of this evaluation phase is to verify that the proposed approach can be applied in cases where large-scale adaptation rule sets are needed to describe the desired AEHS response. In order to do this, we simulate the existence of an AEHS that uses as many adaptation variables as possible. The variables learner profile properties and educational description model properties are selected from the items of wide-spread learning technology standards. However, special attention is given in generating learner profiles and educational content metadata records that simulate real-life conditions. Details on how such datasets are set are stated in Sect. 7.5.

7.5 Setting Up the Experiments

Before executing our experiments for measuring the performance of adaptive selection of LO, we need to design the media space and the learner model as the way explained in the next subsections.

7.5.1 Designing the Media Space

In the evaluation, we extract the AM of the INSPIRE system (Papanikolaou et al. 2003). INSPIRE system uses two variables in the educational resource description model, namely, the performance level and the learning resource type.

In the second evaluation phase, we simulate the existence of an AEHS where large-scale adaptation rule sets are needed to describe the desired AEHS response. To do so, we have used as educational resource description model a subset of the IEEE Learning Object Metadata (LOM) standard elements (IEEE 2002), illustrated in Table 7.1. The aggregation level and the relation/kind elements are used for structuring the media space.

Table 7.1 Educational resource description model used in evaluation phase B

IEEE LOM Category	IEEE LOM Element	Explanation
General	Structure	Underlying organizational structure of a LO
	Aggregation Level	The functional granularity of a LO
Educational	Interactivity Type	Predominant mode of learning supported by a LO
	Interactivity Level	The degree to which a learner can influence the aspect or behavior of a LO
	Semantic Density	The degree of conciseness of a LO
	Typical Age Range	Developmental age of the typical intended user
	Difficulty	How hard it is to work with or through a LO for the typical intended target audience
	Intended End User Role	Principal user(s) for which a LO was designed, most dominant first
	Context	The principal environment within which the learning and use of a LO is intended to take place
	Typical Learning Time	Typical time it takes to work with or through a LO for the typical intended target audience
	Learning Resource Type	Specific kind of LO. The most dominant kind shall be first
Relation	Kind	Nature of the relationship between two LO

In both evaluation phases, we need to simulate real-life conditions. This means that the simulated LO metadata records should have a distribution over their value spaces similar to the metadata value distribution found in real-life LO repositories.

(Najjar and Duval 2006) presented a statistical analysis of the actual use of IEEE LOM metadata elements in the ARIADNE LO repository. The results were derived from analyzing the empirical data (usage logs) of 3,700 ARIADNE metadata instances. Table 7.2 presents the percentage of times each ARIADNE data element was filled in by indexers during the indexing process.

Table 7.2 Usage percentage of data elements in ARIADNE repository

IEEE LOM Element	Value Provided (%)	Most used Vocabulary value (M)	% of M (filled-in)	%M among all cases
Aggregation Level	91.9	Lesson	92.7	85.2
Context	53.5	University Degree	69.7	37.2
Interactivity Level	53.2	Medium	67.7	36.1
Semantic Density	52.4	Medium	76.4	40.0
Difficulty Level	52.2	Medium	72.8	38.0
Restrictions	5.2	Contact Author	90	5.2
Source	1.3	-	-	-
Version Information	7.0	-	-	-
Description	11.2	-	-	-
OS Version	0.5	-	-	-
Installation Remarks	24.3	-	-	-
Other Constraints	0.15	-	-	-

From the data shown in Table 7.2, we notice that only one data element is almost always used: the aggregation level element. Other elements are used in about 50 % of the descriptions and the rest are rarely used in the indexing process. For the values of data elements, we can see that indexers often use just one value.

As a result, in order to simulate in our experiments the metadata of a real-world repository, we will generate metadata records with normal distribution over the metadata elements value space, simulating that not all metadata elements and their corresponding vocabulary terms are used equally. Normal distribution is a continuous probability distribution that is often used to describe random variables that tend to cluster around a single mean value.

7.5.2 Designing the Learner Model

In the first phase of the evaluation, we will extract the AM of the INSPIRE system (Papanikolaou et al. 2003). The INSPIRE system uses two variables in the learner model, namely, the learner's knowledge level and the learner's learning style.

In the second evaluation phase, we simulate the existence of an AEHS where large-scale adaptation rule sets are needed to describe the desired AEHS response. To do so, for the design of the learner model in our simulations, we have used an overlay model (Martins et al. 2008) for representing the learners' knowledge space and a stereotype model (Rich 1979) for representing learners' preferences. More precisely, for the learners' knowledge level we assume the existence of a related certification for each node of the learners' knowledge space, the evaluation score in testing records and the number of attempts made on the evaluation.

For modeling of learners' preferences we use learning styles according to (Honey and Mumford 1992), as well as modality preference information consisting of four modality types, namely, the visual modality, the textual modality, the auditory modality and any combination of the three modality preferences (Razmerita 2005). Each element of the learner model was mapped to the IMS Learner Information Package (LIP) specification (IMS 2001), as shown in Table 7.3.

In order to simulate in our experiments the profiles of real learners we generated profile records using truncated standard lognormal distribution with [sigma] = 1 and reduced by factor 1/5. This distribution is often used in the literature for simulating learner behavior (McCalla 2005).

7.5.3 Simulating the AM of an AEHS

The goal of our experiments is to evaluate the suitability of the set of performance evaluation metrics, presented in Sect. 7.3, for validating the use of decision-based approaches for adaptive LO selection in AEHS, and assess the use of these metrics in the case of our previously proposed statistical method for estimating the desired AEHS response.

Performance evaluation in this context means measuring how well our semi-automated approach fits the decision function to the provided model adaptation decisions (training data), and how well this decision function responds to decision cases not known during the training process (generalization capacity).

Table 7.3 Learner model used in evaluation phase B

Learner Model Element	IMS LIP Element	Explanation
Learning Style	Accessibility/ Preference/typename	The type of cognitive preference
	Accessibility/ Preference/prefcode	The coding assigned to the preference
Modality Preference	AccessForAll/ Context/Content	The type of modality preference
Knowledge Level	QCL/Level	The level/grade of the QCL
	Activity/ Evaluation/ noofattempts	The number of attempts made on the evaluation
	Activity/Evaluation / result/interpretscope	Information that describes the scoring data
	Activity/Evaluation/ result/score	The scoring data itself
Relation	Accessibility/ Preference/typename	The type of cognitive preference

As a result, we need to produce model adaptation decisions and compare them with the corresponding response of our decision-based approach. Some of these model adaptation decisions will be used for training our method, and some will be used for measuring its' generalization capacity.

In the first evaluation phase, the AM rules of an existing AEHS are used for generating sample adaptation decisions. In the second evaluation phase, we need to simulate the existence of an AEHS that uses as many adaptation variables as possible. Since such AEHS does not exist, we will simulate model adaptation decisions via the use of simulated instructional designers' preference models. These models have been selected in such a way that the preference surface is complex, thus, it would be a difficult task for the decision based algorithm to fit the training data.

To achieve this, we use as an instructional designers' preference model a multivariable function, with 18 variables (k). These variables model the eleven (11) elements of the educational resource description model in use (that is, the elements used from the "general" and the "educational" IEEE LOM categories) and the seven elements of the learner model in use (Karampiperis and Sampson 2005). We assume that the response of this function expresses the utility of a given LO for a given learner profile (preference-based selection problem).

In our experiments, we simulate the preference models of five instructional designers, using multivariable non-convex functions. In our previous work (Karampiperis and Sampson 2004), we have defined the suitability/utility function of a learning object LO_i for the learner L_j as a function which varies from 0 to 1. This means that before we can use the non-convex functions as instructional designers' preference models, we need to scale them in the same value space. The normalisation equation that we use for this purpose is the (7.5):

$$F_{(f_{(x)})} = \frac{f_{(x)}^2}{1 + f_{(x)}^2} \tag{7.5}$$

where $f_{(x)}$ is the testing function. This formula, scales $f_{(x)} \in \Re$ to a new function $F_{(x)} \in [0,1)$, where $F_{(f_{(x)}=0)} = 0$, and $\lim_{f_{(x)} -> \pm\infty} F_{(f_{(x)})} = 1$.

For evaluating the performance, we have generated a set of 1.000 LO metadata records and a set of 100 learner profiles. In each experiment, 50% of the available LO metadata records, randomly selected, were used for algorithmic training and the rest 50% for measuring the generalisation, that is, the estimation capacity, of the algorithm. Similarly, in each experiment 50% randomly selected of the available learner profiles were used for algorithmic training and the rest 50% for measuring the generalisation of the algorithm.

7.6 Experimental Results and Discussion

We present experimental results from the execution of the above mentioned evaluation methodology for the case of our previously proposed statistical method for estimating the desired AEHS response (Karampiperis and Sampson 2005). The results are presented per evaluation phase.

7.6.1 Extracting the AM of Existing AEHS

Our first experiment was the application of our decision-based approach for replicating the AM of an existing AEHS. To this end, we simulated the AM of the INSPIRE (Papanikolaou et al. 2003), produced sample adaptation rules in the form of combinations of LO mapped to learner profiles, and applied our methodology to extract the AM. The INSPIRE system uses two variables from the learner model (namely, the learner's knowledge level and the learner's learning style) and two variables from the educational resource description model (namely, the performance level and the learning resource type), for performing adaptation decisions according to Table 7.4.

Fig. 7.1, presents the INSPIRE's AM dependencies of the learning style and learning resource type in the LO utility space, whereas Fig. 7.2 presents the same dependencies of the produced AM when our decision based approach is applied. From these figures we can observe that the produced AM is a super class of the INSPIRE's AM, since it contains more adaptation rules (dependencies between LO and learner characteristics). Moreover, we can observe that the produced AM has a continuous contour in the utility space, which means that this AM has the ability to always propose LO.

Table 7.4 INSPIRE AM rules (Papanikolaou et al. 2003)

Learner Attributes		Proposed LO	
Knowledge Level	Inadequate	Performance Level	Remember
	Mediocre		Use
	Advanced		Find
	Proficient		-
Learning Style	Activist	Learning Resource Type	Activity-oriented
	Reflector		Example-oriented
	Theorist		Theory-oriented
	Pragmatist		Exercise-oriented

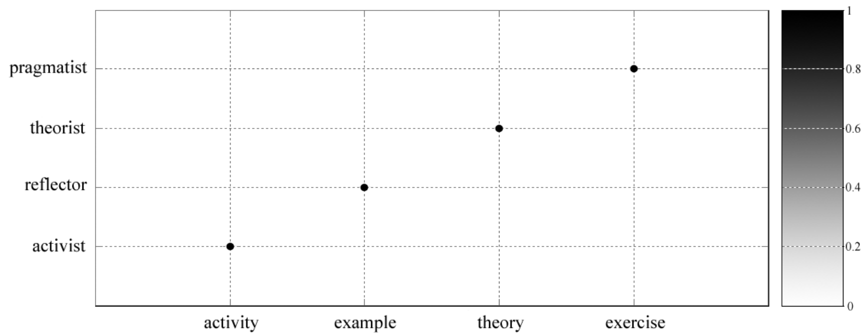

Fig. 7.1 INSPIRE: learning style and learning resource type utility space

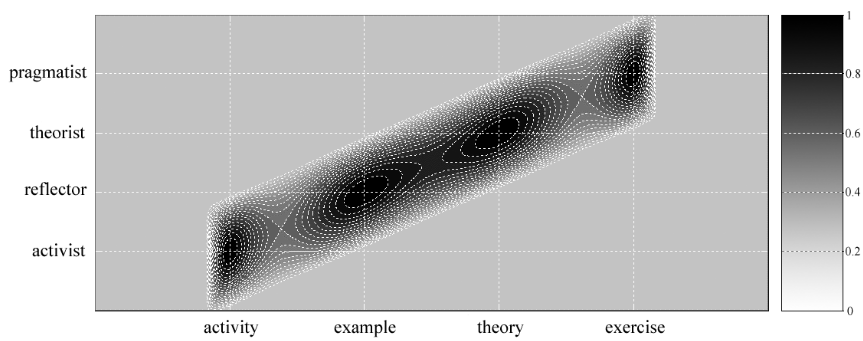

Fig. 7.2 Generated learning style and learning resource type utility space from INSPIRE

The designers of INSPIRE recognize as a problem when designing the INSPIRE system, the required effort for producing LO which cover all the combinations introduced by the INSPIRE AM Rules (Papanikolaou et al. 2003). This is due to the fact that the INSPIRE adaptation rules does not cover all the combinations of the free variables value space, e.g. what happens when a learner has knowledge level equal to “advanced” and learning style equal to “theorist”, but no theory-oriented LO with performance level equal to “find” exist in the LO repository. In this case, the INSPIRE system fails to provide a response, whereas by using our proposed decision based approach, the INSPIRE would respond with a suboptimal solution which would select the LO with the maximum utility for the given learner from the available ones.

After the above experiment, the research question was to investigate if the proposed decision based approach has the capacity of learning more complex AMs, consisting of many free variables (such as the adaptation variables presented in Table 7.1 and Table 7.3), with complex preference surfaces, thus, it would be a difficult task for the decision based algorithm to fit the training data. This is the goal of the second evaluation phase, which is presented in Sect. 7.6.2.

7.6.2 Scaling Up the Experiments

Before proceeding with the performance evaluation of our decision-based approach, we have conducted an additional experiment, aiming to assess the use of the performance evaluation metrics proposed by the literature.

Our semi-automated approach for adaptive content selection uses a preference-based LO selection mechanism based on the use of a suitability function that estimates the utility of a given LO for a given learner.

In order to compare the performance evaluation metrics discussed in Sect. 7.3, we evaluate the performance using randomly generated datasets which serve as model adaptation decisions and vary in size. The size of these datasets depends on the number of ranked LO for a given number of learner profiles. In real conditions, these rankings would be requested from an instructional designer. In our experiments, these rankings are the result of the application of the simulated instructional designers' preference models.

The datasets were divided into two subsets: the training dataset, which was used for algorithmic training and for evaluating the performance during the training process, and the generalization dataset, which was used for measuring the generalization capacity of the algorithm. Each experiment was executed 100 times using a randomly selected instructional designers' preference model.

Fig, 7.3 presents average selection performance results during algorithmic training, when using different simulation parameters regarding the number of learner profiles and the number of LO metadata records used. In each experiment, the selection performance was measured when using different values of the parameter n (varying from 10 to 500), which expresses the maximum number of requested LO from the Media Space. In this figure the performance evaluation was measured using the typical PM, the proposed alternative metric for SS, as well as, by applying the PM metric only on the desired sub-space of the media space (partial precision metric, PPM). From these results we observe the following:

1. Precision when measured with PM metric is independent from the maximum number of requested LO from the media space (selection space), as well as, from the ranking of the selected LO,
2. Precision when measured with PPM metric is independent from the ranking of the selected LO, but depends on the volume of the selection space,
3. The PPM metric tends to be equal to the PM metric when the selection space becomes bigger (n increases),
4. Performance evaluation using the PM metric is higher or equal to the performance when using the PPM metric. Also performance evaluation using the PM metric is higher or equal to the performance when using the SS metric,
5. The SS metric tents to be lower as the searching space increases, whereas PPM metric becomes higher as the searching space increases. This is due to the fact that, when the searching space increases the probability of introducing ranking errors also increases. Since the PPM metric is not dependent by the ranking of the selected LO, the PPM metric behaves differently from the SS metric.

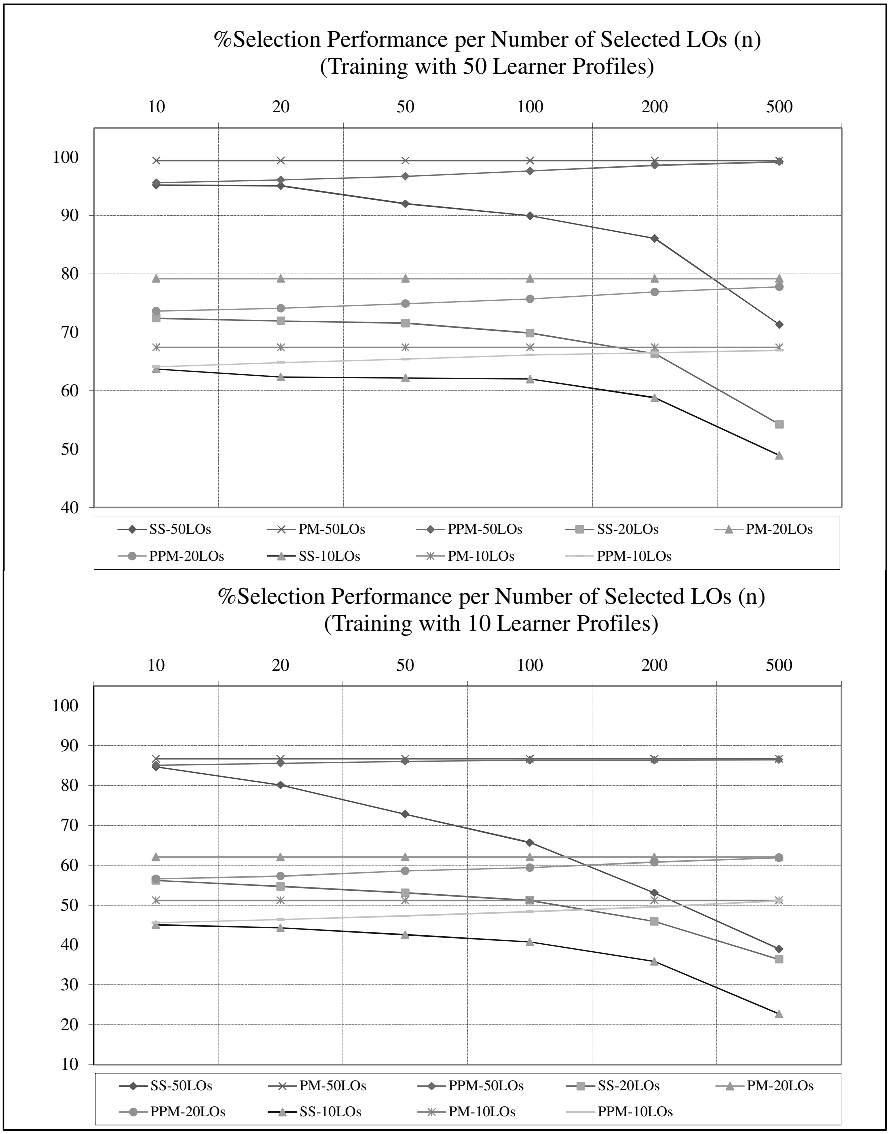

Fig. 7.3 Adaptive selection of LO - training results

The same observations apply also when measuring the generalization capacity, as depicted in Fig, 7.4. These observations verify the hypothesis that by definition the SS metric is stronger than the PM or the PPM metric. The PP and PPM metrics do not capture the precision errors resulting from ranking errors in the selected LO set.

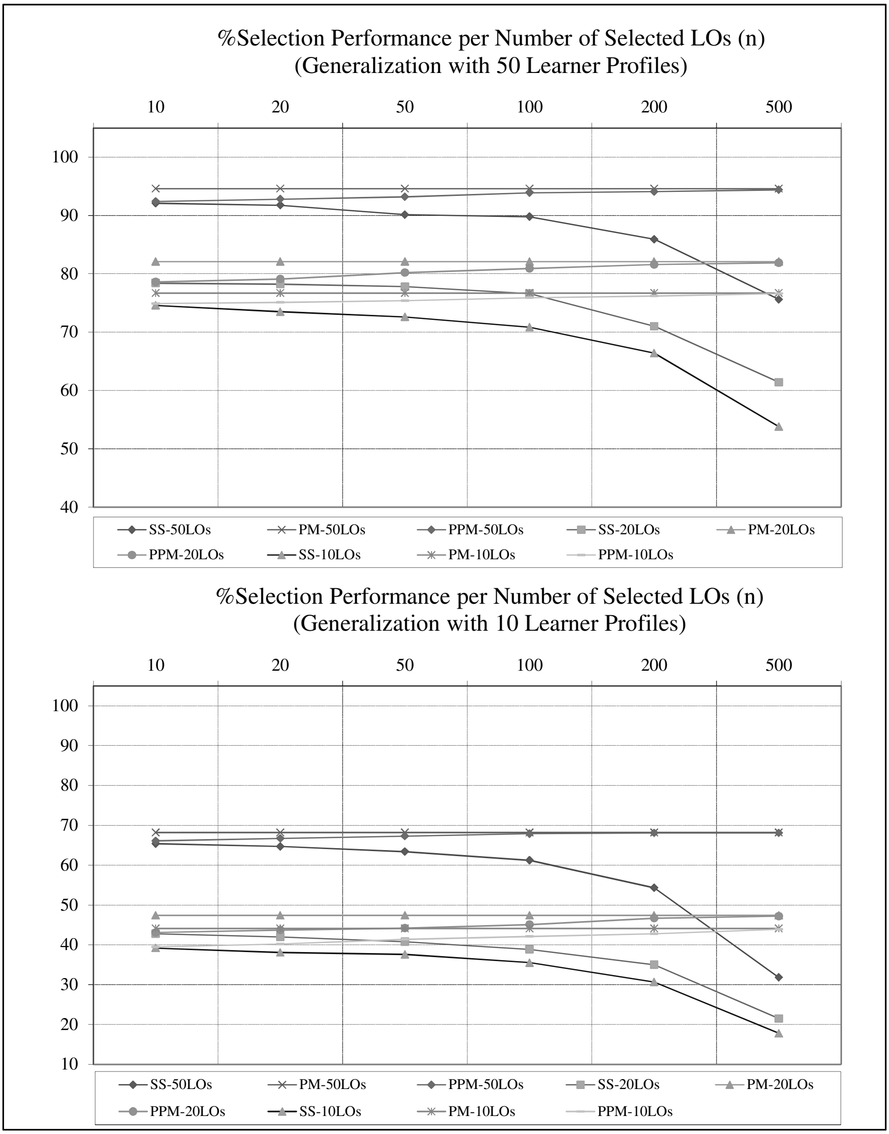

Fig. 7.4 Adaptive selection of LO - generalization results

In the case where the resulting LO sets hold the same LO but differ in the ordering of LO, the PP and PPM metrics remain the same (seam stable); whereas, the SS metric provides a realistic measurement of the precision. So in the case of AEHS, where the ranking of the selected LO is critical, the SS metric is used.

From these results we also observe that the SS depends on the requested LO from the media space (n), as well as the number of the LO and learner instances used for algorithmic training.

Additionally, for the same number of requested objects and the same number of learner profiles used, using more LO metadata records produces higher SS rates. Accordingly, for the same number of requested objects and the same number of used LO metadata records, using more learner profiles produces higher SS rates.

More analysis on the results presented in Fig, 7.3 and Fig, 7.4 show that, when the desired number of LO (n) is relatively small (less than 20), the selected LO by the decision model are close to those the instructional designer would select (with success rate over 70%), when using an input set consisting of more than 500 combinations of LO mapped to learner profiles (calculated as the multiplication of the LO with the learner profiles used).

By using the presented performance evaluation metrics, we can additionally investigate the influence of the explicit combinations required from the instructional designer (which are directly equivalent to the design effort required). To this end, we have executed additional experiments measuring the SS gain per number of requested combinations. This metric provides evidences about the trade-off that an instructional designer should make between the required effort and the improvement of the SS rate.

Fig, 7.5 presents simulation results of the design trade-off for combinations of LO metadata records with learner profiles that produce SS over the threshold of 70% for different values of the desired number of LO (n). From these results we observe that using a configuration of 500 combinations (which means classifying 50 LO metadata records over 10 learner profiles or vice versa) the gain in the SS rate is higher than using configurations with more combinations. The machine learning algorithm uses input knowledge in order to generate a continuous decision function that estimates the desired AEHS response. This knowledge comes in the form of combinations of LO mapped to learner profiles. When more input knowledge is provided, the machine learning algorithm fits better the response function on these data. However, there is a limit where this fitting process fails. If the algorithm is fed with too many input data, then it will over fit the response function over these data, losing its generalization capacity. Furthermore, we can observe that using the combination of 10 LO metadata records classified over 50 learner profiles leads to higher gain in the generalization success rate, whereas, using the opposite combination, that is, 50 LO metadata records classified over 10 learner profiles, leads to better results during the algorithmic training.

This is due to the fact that our decision based approach uses an interpolation method over the LO metadata space and an extrapolation mechanism over the learner profile space. This means that our approach learns from LO sequences associated with known learner profiles and generalizes its results to cover unknown learner profiles. Thus, using combinations with more LO leads to higher success rates during the training process, whereas, using combinations with more learner profiles leads to higher success rates during the generalization process. As a result, in order to minimize the required design effort and at the same time to maximize the SS rate, the combination of 10 LO metadata records classified over 50 learner profiles would be preferred.

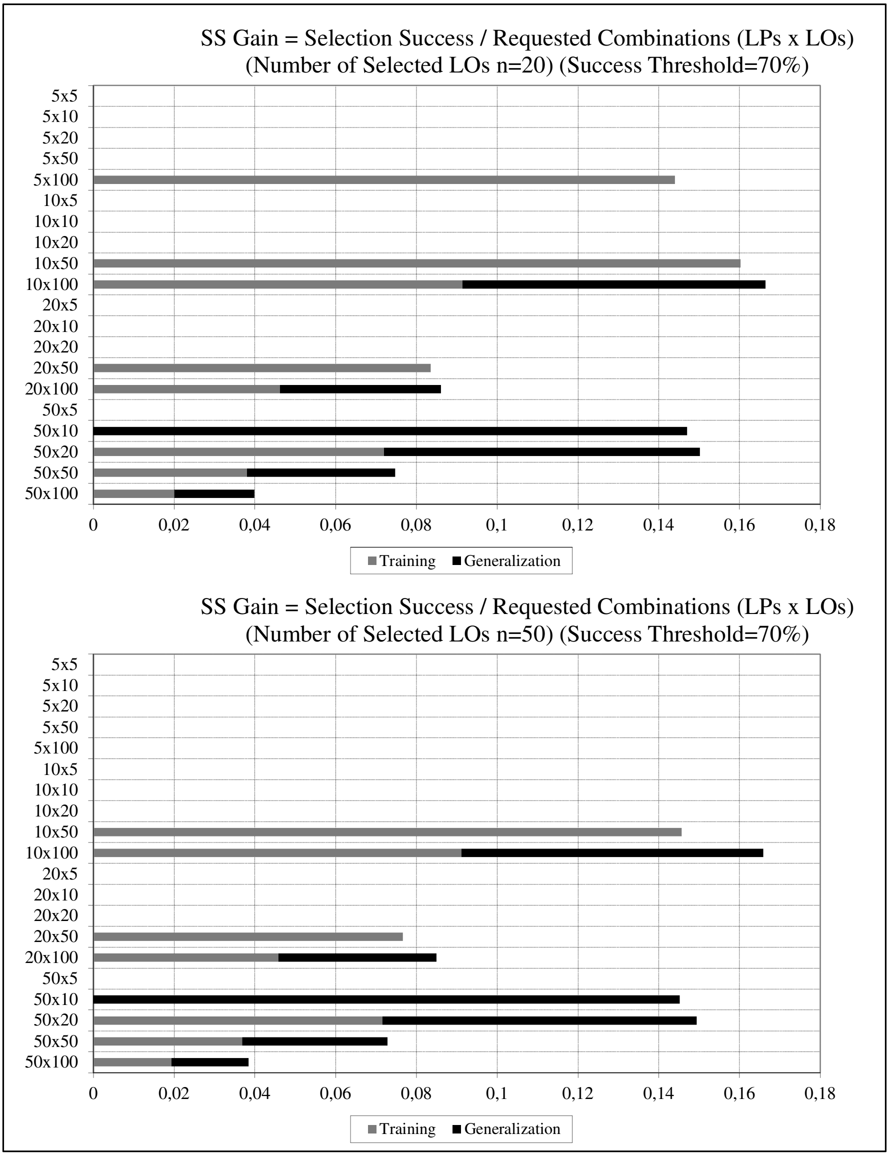

Fig. 7.5 Adaptive SS gain per requested input combinations

7.7 Summary and Future Research Directions

Adaptive LO selection is recognized as a challenging research issue in AEHS. In order to adaptively select LO in AEHS, the definition of adaptation behavior is required.

Several efforts have been reported in literature aiming to support the AM design by providing AEHS designers with either guidance for the direct definition of adaptation rules, or semi-automated mechanisms which generate the AM via the implicit definition of such rules.

The main drawback of the direct definition of adaptation rules is that there can be cases during the run-time execution of AEHS where no adaptation decision can be made. This is due to the fact that, even if appropriate resources exist in the media space, the absence of a required rule (insufficiency problem) or the conflict between two or more rules (inconsistency problem), prevents the AEHS to select and use them in the generated learning resource sequence. As a result, either less appropriate resources are used from the media space, or required concepts are not covered at all by the resulting sequence.

The goal of the semi-automated, decision-based approaches is to generate a continuous decision function that estimates the desired AEHS response, aiming to overcome the above mentioned problem. To achieve this, semi-automated approaches use data from the implicit definition of sample adaptation rules and attempt to fit the response function on these data. Although such approaches bare the potential to provide efficient AMs, they still miss a commonly accepted framework for evaluating their performance.

In this chapter, we discussed a set of performance evaluation metrics that have been proposed by the literature for validating the use of decision-based approaches in adaptive LO selection, and assessed the use of these metrics in the case of our previously proposed statistical method for estimating the desired AEHS response.

More precisely, we discussed the limitations of the performance metrics used by the literature for the problem of adaptive content selection, introduced the need for an alternative evaluation metric and presented a metric, which although seems similar to the PM in information retrieval systems, its difference is critical. This metric evaluates the precision of selecting LO not on the entire space of the media space, but only on the desired sub-space, and also it takes into consideration the ranking of the selection process.

Future research includes the study of variations of the presented performance evaluation metrics, as well as, the investigation of a comparison metric between rule-based and decision based AEHS. In the context of AEHS an interesting research question is the separation of the learning scenario from the AM. By this way, we anticipate, on one hand, to support the sequencing of unstructured raw media, and on the other hand, to facilitate the support of different pedagogical strategies without redesigning the AM rules. Moreover, the decomposition of LO from existing courses, allowing reuse of the disaggregated LO in different educational contexts is considered as an important research question.

The intelligent selection of the disaggregation level and the automatic structuring of the atoms (raw media) inside the disaggregated components in order to preserve the educational characteristics they were initially designed for, is a key issue in the research agenda for LO (Duval and Hodgins 2003).

References

Alfonseca, E., Rodriguez, P., Perez, D.: An approach for automatic generation of adaptive hypermedia in education with multilingual knowledge discovery techniques. Computers & Education 49, 495–513 (2007)

Aroyo, L., Mizoguchi, R.: Towards evolutional authoring support systems. Journal of Interactive Learning Research 15(4), 365–387 (2004)

Aroyo, L., Mizoguchi, R., Tzolov, C.: OntoAIMS: Ontological approach to courseware authoring. In: Proceedings of ICCE, pp. 1011–1014 (2003)

Biletskiy, Y., Baghi, H., Keleberda, I., Fleming, M.: An adjustable personalization of search and delivery of learning objects to learners. Expert Systems with Applications 36(5), 9113–9120 (2009)

Brusilovsky, P.: Adaptive Navigation Support. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 263–290. Springer, Heidelberg (2007)

Brusilovsky, P., Henze, N.: Open Corpus Adaptive Educational Hypermedia. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 671–696. Springer, Heidelberg (2007)

Brusilovsky, P., Wade, V., Conlan, O.: From learning objects to adaptive content services for E-learning. In: Pahl, C. (ed.) Architecture Solutions for E-Learning Systems, pp. 243–261. Idea Group Inc. (2007)

Conlan, O., Hockemeyer, C., Wade, V., Albert, D.: Metadata driven approaches to facilitate adaptivity in personalized eLearning systems. Journal of the Japanese Society for Information and Systems in Education 1(1), 38–44 (2002)

Cristea, A.I.: Authoring of adaptive educational hypermedia. In: Proceedings of ICALT, pp. 943–944 (2007)

Cristea, A.I., Kinshuk: Consideration on LAOS, LAG and their Integration in MOT. In: Proceedings of ED-MEDIA, pp. 511–518 (2003)

Dagger, D., Wade, V., Conlan, O.: Personalisation for all: Making adaptive course composition easy. Education Technology & Society 8(3), 9–25 (2005)

De Bra, P.: Web-based educational hypermedia. In: Romero, C. (ed.) Data Mining in E-Learning, pp. 3–17. WIT Press (2006)

De Bra, P., Houben, G., Wu, H.: AHAM: A Dexter based reference model for adaptive hypermedia. In: Proceedings of HT, pp. 147–156 (1999)

Dolog, P., Henze, N., Nejdl, W., Sintek, M.: The Personal Reader: Personalizing and Enriching Learning Resources Using Semantic Web Technologies. In: De Bra, P., Nejdl, W. (eds.) AH 2004. LNCS, vol. 3137, pp. 85–94. Springer, Heidelberg (2004)

Dolog, P., Simon, B., Nejdl, W., Klobucar, T.: Personalizing access to learning networks. ACM Transactions on Internet Technology 8(2), Article 8 (2008)

Duval, E., Hodgins, W.: A LOM research agenda. In: Proceedings of WWW (2003)

Harman, K., Koohang, A.: Learning objects: standards, metadata, repositories & LCMS. Informing Science Press, California (2006)

Henze, N., Nejdl, W.: A logical characterization of adaptive educational hypermedia. New Review of Hypermedia and Multimedia 10(1), 77–113 (2004)

Honey, P., Mumford, A.: The manual of learning styles. Maidenhead (1992)

Hsieh, T.C., Chiu, T.K., Wang, T.I.: An approach for constructing suitable learning path for documents occasionally collected from internet. In: Proceedings of ICMLC, pp. 2138–3143 (2008)

Huang, M.J., Huang, H.S., Chen, M.Y.: Constructing a personalized e-learning system based on genetic algorithm and case-based reasoning approach. Expert Systems with Applications 33, 551–564 (2007)

Huang, Y., Chen, J., Huang, T., Jeng, Y., Kuo, Y.: Standardized course generation process using dynamic fuzzy Petri nets. Expert Systems with Applications 34, 72–86 (2008)

IEEE. IEEE learning object metadata standard (IEEE LOM). Technical specification. IEEE (2002)

IMS. Learner information package (LIP). Technical specification. IMS Global Learning Consortium Inc. (2001)

Karampiperis, P., Sampson, D.G.: Adaptive learning object selection in intelligent learning systems. Journal of Interactive Learning Research 15(4), 389–409 (2004)

Karampiperis, P., Sampson, D.G.: Adaptive learning resources sequencing in educational hypermedia systems. Educational Technology & Society 8(4), 128–147 (2005)

Kiu, C.C., Lee, C.S.: Learning objects reusability and retrieval through ontological sharing: A hybrid unsupervised data mining approach. In: Proceedings of ICALT, pp. 548–550 (2007)

Knutov, E., De Bra, P., Pechenizkiy, M.: AH 12 years later: a comprehensive survey of adaptive hypermedia methods and techniques. New Review of Hypermedia and Multimedia 15(1), 5–38 (2009)

Kravčík, M., Specht, M.: Flexible Navigation Support in the WINDS Learning Environment for Architecture and Design. In: De Bra, P., Nejdl, W. (eds.) AH 2004. LNCS, vol. 3137, pp. 156–165. Springer, Heidelberg (2004)

Lee, M., Tsai, K., Wang, T.: An ontological approach for semantic-aware learning object retrieval. In: Proceedings of ICALT, pp. 208–210 (2006)

Martins, A.C., Faria, L., Vaz de Carvalho, C., Carrapatoso, E.: User modeling in adaptive hypermedia educational systems. Educational Technology & Society 11(1), 194–207 (2008)

McCalla, G.: Smart recommendation for an evolving e-learning system: architecture and experiment. International Journal on E-Learning 4(1), 105–129 (2005)

McGreal, R.: Online education using learning objects. Routledge, London (2004)

Najjar, J., Duval, E.: Actual use of learning objects and metadata: An empirical analysis. IEEE Technical Committee on Digital Libraries Bulletin 2(2) (2006)

Nejdl, W., Brusilovsky, P.: Adaptive hypermedia and adaptive web. In: Singh, M. (ed.) Practical Handbook of Internet Computing, pp. 1.1–1.12. CRC Press (2004)

Ochoa, X., Duval, E.: Relevance ranking metrics for learning objects. IEEE Transactions on Learning Technologies 1(1), 34–48 (2008)

Papanikolaou, K., Grigoriadou, M., Kornilakis, H., Magoulas, G.: Personalising the interaction in a web-based educational hypermedia system: The case of INSPIRE. International Journal of User Modeling and User-Adapted Interaction 13(3), 213–267 (2003)

Ras, E., Ilin, D.: Using Decision Models for the Adaptive Generation of Learning Spaces. In: Nejdl, W., Kay, J., Pu, P., Herder, E. (eds.) AH 2008. LNCS, vol. 5149, pp. 153–162. Springer, Heidelberg (2008)

Razmerita, L.: User modeling and personalization. In: Sherry, Y.C., Magoulas, G.D. (eds.) Adaptable and Adaptive Hypermedia Systems, Idea Group Inc. (2005)

Rich, E.: User modeling via stereotypes. Cognitive Science 3(4), 329–354 (1979)

Sampson, D., Karampiperis, P.: Decision models in the design of adaptive educational hypermedia systems. In: Graf, S., Lin, F., Kinshuk, McGreal, R. (eds.) Intelligent and Adaptive Learning Systems: Technology Enhanced Support for Learners and Teachers. Idea Group Inc. (2011)

Stash, N., Cristea, A.I., De Bra, P.: Adaptation languages as vehicles of explicit intelligence in adaptive hypermedia. International Journal of Continuing Engineering Education and Lifelong Learning 17(4-5), 319–336 (2007)

Wang, T.I., Tsai, K.H., Lee, M.C., Chiu, T.K.: Personalized learning objects recommendation based on the semantic-aware discovery and the learner preference pattern. Educational Technology & Society 10(3), 84–105 (2007)

Wu, H., De Bra, P.: Sufficient Conditions for Well-Behaved Adaptive Hypermedia Systems. In: Zhong, N., Yao, Y., Ohsuga, S., Liu, J. (eds.) WI 2001. LNCS (LNAI), vol. 2198, pp. 148–152. Springer, Heidelberg (2001)

Abbreviations

AEHS	Adaptive Educational Hypermedia Systems
AHA	Adaptive Hypermedia Architecture
AHAM	Adaptive Hypermedia Application Model
AM	Adaptation Model
ATO	Authoring Task Ontology
ICPC	International Classification of Primary Care
LIP	Learner Information Package
LO	Learning Object
LOM	Learning Object Metadata
MeSH	Medical Subject Headings
MOT	My Online Teacher
PM	Precision Metric
PPM	Partial Precision Metric
SS	Selection Success