Journal of Universal Computer Science, vol. 22, no. 2 (2016), 197-227
submitted: 1/2/15, accepted: 29/1/16, appeared: 1/2/16 © *J.UCS*

Metadata for Recommending Primary and Secondary Level Learning Resources

Jorge Bozo
(Escuela de Ciencias Empresariales
Universidad Católica del Norte, Coquimbo, Chile
and
Computer Science Department
Pontificia Universidad Católica de Chile, Santiago, Chile
jorge.bozo@ucn.cl, jorge.bozo@uc.cl)

Rosa Alarcon, Monserrat Peralta, Tomas Mery
(Computer Science Department
Pontificia Universidad Católica de Chile, Santiago, Chile
{raalarco, mperalt1, tmery1}@puc.cl)

Verónica Cabezas
(Education Faculty
Pontificia Universidad Católica de Chile, Santiago, Chile
vcabezag@uc.cl)

Abstract: Recommender systems have been used in education to assist users in the discovery of learning resources. Unlike product-oriented recommender systems, the goals and behavior of users in education are influenced by their context; such influence may be stronger in formal scenarios such as primary and secondary education since context is highly regulated. Intuitively, we could assume that a biology teacher may be more interested in biology-related content rather than content from other fields. In this paper we explore such assumption by analyzing the impact of educational metadata that is associated to resources and teachers. We apply hierarchical clustering to determine clusters of interest and using a *teacher profile*, we classify new teachers and new items in order to predict their preferences. In order to validate our approach, we used a dataset derived from a repository of learning resources widely used by teachers in primary and secondary school in Chile in the role of *old users*, we also performed an experiment with teachers in training in the role of *new users*. Our results confirm the diverse impact of metadata on the formation of such clusters and on recommendation.
Key Words: Recommender Systems in Education, Collaborative Filtering, Metadata, Hierarchical Clustering, cold start.
Category: L.1.2, L.2.2, L.3.0, L.3.2

1 Introduction

Learning resources are digital documents used for E-Learning [Lehmann et al., 2008]; this definition includes multimedia resources, hypertext, complete courses or websites. Learning resources can refer to a single or several digital documents written in different formats. Learning Objects (LOs), on the other hand, are small instruction pieces or building blocks, that can be shared, re-used in other contexts, and combined into bigger instruction blocks [Koper, 2003, Motelet, 2007]. Both, learning resources and LOs are described with metadata that provides additional information so that they can be discovered from various perspectives. Metadata itself can be standardized, being LOM one of the most heavyweight and used standards for LOs [Ochoa et al., 2011]; or it can follow an open and lightweight strategy where resources are annotated with informal tags, generating non-standardized, user-based taxonomies [Manouselis et al., 2011].

Learning resources can be provided as organized collections administered by communities of teachers, with or without metadata [Tiropanis et al., 2009]; or through specialized repositories that host various items ranging from hundreds to millions [Ochoa and Duval, 2009]. Such repositories can either contain the resources themselves, and their associated metadata, or only the metadata and a referral to the actual resource location, or to the location of a hosting web site where an additional search could be required. Most repositories support mainly keyword-base search (on metadata, data format, and even content), resulting in an extensive list of resources and requiring users to look into such results to find out what they need.

Furthermore, standardized metadata (e.g. IEEE-LOM) have been criticized as complex and ambiguous [Tiropanis et al., 2009], requiring a lot of effort and expertise from resource producers to associate a resource with good quality metadata [Motelet, 2007]; most importantly, it lacks support for representing pedagogical concerns such as the learning needs they are attending [Nitto et al., 2006]. These characteristics make resource discovery a cumbersome task. In addition, the lack of consensus in the metadata itself and its meaning, makes difficult to integrate repositories, and to provide federated search engines which in turn limits the search scope and resource's reuse potential.

Some researchers use recommender systems in order to facilitate learning resource search. However, they are mainly focused on students as the primary resource consumers [Manouselis et al., 2013]; the needs and practices followed by teachers, particularly from primary and secondary education, are often ignored. Pedagogical tools such as curricula, lesson plans, rubrics and so on, are resources and terminology extensively used by teachers and instructors at these levels but are notably absent in both heavyweight and lightweight approaches [Tiropanis et al., 2009]. Recommender systems in education differ from other fields such as movies, news, etc. (e.g. product-oriented recommenders) [Verbert et al., 2012] because of the user goals. For instance, when teachers want to prepare new material for a class they may consider some lesson plan as a basis and select resources that support information seeking, motivate the audience, recall existing knowledge, etc. [Manouselis et al., 2011] or even to choose simpler material as an introductory subject. Teachers may have various interests depending on their specialty and the administrative activities demanded by their job.

Recommender systems are typically classified into Collaborative Filtering (based on the user's preferences regarding resources), Content-based (considers mainly resource's characteristics), Knowledge-based (uses a user profile model to infer user preferences), or a hybrid. An important challenge in recommender systems is the cold start problem, which is, the recommendation for novel users, new resources or both [Lika et al., 2014], since neither users nor resources have recorded preferences. Most approaches face the cold start problem focusing on new items (learning resources in our case); our focus is on novel users and new items for an education community, particularly the teachers; since this community exhibits a specific behavior due to regulations, content structure and activities.

In order to face the cold start problem, we applied hierarchical clustering to identify clusters of interest (teachers) and analyze their associated metadata. That is, we applied collaborative filtering using a linear kNN search algorithm and Euclidean Distance (visits to the users items) to determine neighborhoods that are agglomerated according to shared neighbors (determined by a threshold p), resulting into clusters (users, items and its metadata) of shared interest.

We define a *teacher profile* (Table 10) to classify *new teachers* in the corresponding cluster without requiring detailed previous knowledge of her preferences (such as teaching styles, resources format, etc.) and activities (such as current stage in the curricular plan). Then we used the items' metadata to classify *new items* into the corresponding cluster and predict a rating of the given item for the given new teacher.

We performed our experiments on an educational dataset derived from a repository of learning resources widely used by primary and secondary level teachers in Chile during a time span of five years. Our dataset confirmed the tendency seen in similar educational datasets, that is, it was characterized with high sparsity. We conducted also an experiment with teachers in training in order to obtain new teacher profiles and preferences and determine the impact of metadata on generating accurate predictions.

Our contributions are twofold, on one hand we propose a strategy to face the cold start problem in education by exploiting a *teacher profile* and communities of shared interest; on the other hand, we found out that curricular metadata greatly influences recommendation quality, particularly *subject* whereas *level* have varying influence probably depending on the complexity of the resource. For instance, grade metadata (e.g. 1st grade) weights are greater than other metadata weights; hence they become more relevant when characterizing a cluster. Later such weighted terms are taken into account when determining the cluster for the new users. Our results are promising in recommending novel users and items and they also suggest that teachers form communities of interest based on a combination of their field (e.g. arts) and level (e.g. 1st grade), with diverse emphasis. Teachers in a cluster are interested in resources corresponding to various levels (e.g. first, second, third year, etc.), but the level alone did not predict a teacher's cluster. That is, the metadata alone is not an accurate predictor, but the community behavior (i.e. preferences) determines the nuances of metadata combination.

This paper is organized as follows, section 2 presents related work, section 3 presents the preliminaries for understanding recommender systems and hence our approach, section 4 presents our analysis of the educational dataset, section 5 presents our approach to face the cold start problem in detail and discusses the results, finally section 6 presents our conclusions and future work.

2 Related work

Learning resources can be found in-the-wild by searching open content through Web search engines such as Google and Yahoo search, or even in public sites such as Wikipedia, Youtube, iTunes or the MIT Open Courseware [Tiropanis et al., 2009]. The problem with these approaches is not the lack of content but the over abundance of material and the lack of acknowledgment of the educational culture and practice in the design of the algorithms [Tiropanis et al., 2009]. Teachers, particularly in primary and secondary education, Related workform a community of practice that shares a common (or similar) terminology, tools and skills. This property is reflected in specialized repositories such as, MERLOT (`www.merlot.org`), which contains more than 45.000 resources and provides a resource classification based on a limited list of 23 academic disciplines (e.g. Agriculture and Environmental Sciences), or academic support communities (e.g. Faculty development), among others.

Recommender systems in education have been studied as a means for supporting users in finding appropriate resources, among other tasks [Duval et al., 2009]. In general, recommender systems techniques are classified as *content-based*, *collaborative filter*, knowledge-based, or a hybrid. Content-based techniques generally recommend similar items according to the text or other features (e.g. affective annotations [Canini et al., 2013]) describing the item. Collaborative filter considers the preferences of users with similar interests [Felfernig and Burke, 2008, Segaran, 2007a]. Knowledge based systems, on the other hand a, user model to serve as a basis for inferring user preferences. For instance, in [Carrer-Neto et al., 2012] a system considers the social network to determine users similarity. A movies ontology along with context information such as location, time and crowd is used in [Mandl et al., 2011] to recommend movies. Context information is used to determine users similarity while movies knowledge is used to determine items similarity in this hybrid system. Users preferences elicitation based on the ontology is taken into account to face the *cold start* problem and user behavior (e.g. time spent on browsing) is used as implicit information to calibrate users preferences. Users willingness to contribute explicit or implicit knowledge (e.g. learning resources, advice) is examined in [Zhang et al., 2012] where they find that social incentives (i.e. recognition) positively influences tacit knowledge contribution whereas monetary incentive promotes explicit contribution. Users preferences are further analyzed in [Colombo-Mendoza et al., 2015] where preferences are constructed within a recommending session rather than being statically predefined. The authors found that aspects such as items poor quality, the order of the recommendation (ascending or descending rankings), the formulation of user choices, and the default options may significantly influence users preferences.

In knowledge based systems, on the other hand, a user model serves as a basis for inferring user preferences. For instance, in [Carrer-Neto et al., 2012] a system considers the social network to determine users similarity and a movies ontology to determine items similarity. A movies ontology along with context information such as location, time and crowd is used in [Mandl et al., 2011] to recommend movies. Context information is used to determine users similarity while movies knowledge is used to determine items similarity in this hybrid system. Users preferences elicitation based on the ontology is taken into account to face the cold start problem and user behavior (e.g. time spent on browsing) is used as implicit information to calibrate users preferences. Users willingness to contribute explicit or implicit knowledge (e.g. learning resources, advice) is examined in [Zhang et al., 2012] where they find that social incentives (i.e. recognition) positively influences tacit knowledge contribution whereas monetary incentive promotes explicit contribution. Users preferences are further analyzed in [Colombo-Mendoza et al., 2015] where preferences are constructed within a recommending session rather than being statically predefined. The authors found that aspects such as items poor quality, the order of the recommendation (ascending or descending rankings), the formulation of user choices, and the default options may significantly influence users preferences. Students preferences, elicited during a question-answering exercise using a Markov chain model in [Taraghi et al., 2015], serves as the basis of a learner profile and a hierarchical clustering classification. Other knowledge based system focuses on activities rather than on the user, for instance, in [Rodrıguez et al., 2015] contextualized learning activities along with a multicriteria approach (weight factors) are used to recommend tools, persons and events to teachers.

However, recommender systems in education are quite different than recommender systems of goods or services, because they must consider not only the learners or teachers preferences for certain material, but also how this material may help them to achieve their goals. For instance, the SMART project [Duval et al., 2009] recommends items considering pedagogical characteristics such as previous knowledge along with learner's interests. Nadoslki et al. [Nadolski et al., 2009] evaluate an ontology-based strategy for modeling the learner's profile versus a lightweight (peer-rates) approach and find that an ontology-based technique is costly but more accurate. Manouselis [Manouselis and Costopoulou, 2007] proposes a classic neighborhood-based collaborative filtering algorithm for recommending LOs that considers multi-dimensional ratings on LOs, provided by the teachers (peer-rates). A nice review of the area can be found in Manouselis [Duval et al., 2009].

Shepitsen et al. [Shepitsen et al., 2008] follow a different approach, although not in the educational field. They use tags, freely assigned by users, to indirectly define a profile of preferred resources for each user. Tags are clustered following a hierarchical agglomerative approach; the clusters are used as the basis for a personalized algorithm that starting with a single tag query, finds the most similar resources. The results are weighted and ranked based on the user interest, understood as the sum of the products of the user's tag annotations for each cluster and the ratio of resources annotated with a tag in the cluster. In [Chatti et al., 2013], sixteen tag-based collaborative filtering algorithms are tested in terms of precision and recall. They found that item-based hierarchical clustering and item-based k-means clustering obtained the best results whereas user-based clustering techniques showed poor results due to sparseness.

In [Kurilovas et al., 2014], a group of learners assign a set of tags to learning resources in order to create a bottom-up learner context model, whereas expert users (e.g. teachers for the case of pedagogical properties) elicit their own tag ranking (top-down), in order to determine the quality of a learning resource. In [Sieg et al., 2010] a domain ontology (instead of tags) is used as the basis for the user profile in a non-educational collaborative recommender experiment. Concepts in the ontology are hierarchically organized and such relationship is considered when determining the user interest in a concept through ratings. Users similarity is established from the ratings contained in the users' profiles and the recommender algorithm considers the similarity based on the distance between users' interest. Collaborative tagging assisted by a concepts ontology is also used in multimedia search in [Gayo et al., 2010] in order to improve responses search algorithm. User queries are enriched with the ontology information so that the video most similar to the users interest (expressed by the query) is retrieved.

On the other hand, the *cold start* problem is a typical challenge of recommender systems (even in education), it occurs when the recommender cannot derive information for new users or new items since they are unknown (i.e. there are no record of preferences for either of them). Various approaches have been proposed outside the area of education in order to deal with this issue, and most are focused on new items rather than on new users. Regarding the later, in [Lam et al., 2008], based on probabilistic aspect models, both collaborative filtering and content-based techniques are combined in order to recommend known items to new users. Another approach is proposed in [Shaw et al., 2010], where the lack of information on user preferences is expanded through association rules, however, for datasets with large multi-valued attributes the solution is often intractable. In [Lika et al., 2014], the cold start problem is faced through a set of classifying algorithms that considers a demographic vector (users characteristics) and produces a set of neighbors. New users are associated to a neighbor using similarity metrics and ratings for the new user are produced collaboratively.

Users characterization in the educational recommending area have been proposed for other purposes than facing the cold start problem. For instance, [Drachsler et al., 2008] proposes the use of stereotypes for learners, which are categories based on demographics data or learning objectives, however the authors do not describe how such approach would be implemented. Tang et al. [Tang et al., 2014] propose multidimensional recommendation and the acknowledgement of the learner context in order to differentiate learner groups. They found that recommendations from within the same learning group are more effective. Song and Gao [Song and Gao, 2014] propose to identify tags from learning resources' content (text) along with collaborative filtering in order to face the cold start issue with promising results. Indicators such as diversity and similarity of users interests in a group and intra-group are used to determine learning groups in [Dascalu et al., 2014]. Learners are grouped automatically, by teachers' choice, or according to a specialized algorithm based on such indicators. Diversity and similarity are determined from a learner profile (explicit user interests) and his or her history during a learning process.

In this paper we model user preferences by means of demographic profile, a teacher profile. We also use learning resources metadata in order to define a hybrid method combining content-based recommending and collaborative filtering to face the cold start problem for new users and new items.

3 Recommender systems

3.1 Definitions and notation

Learning resources a finite set of *ítems* denoted by $\mathscr{I} = \{i_1, i_2, \ldots, i_m\}$. Teachers correspond to *users* who search for educational resources, is denoted by $\mathscr{U} = \{u_1, u_2, \ldots, u_n\}$. Users assign *ratings* to the items, denoted by $r_{u,i}$, representing the level of satisfaction of the user regarding the item. In our case, since such ratings are very scarce, we considered instead an implicit rating. Hence, we consider the number of *visits* (i.e. the number of times the user requested the learning resource) as a measure of interest in an item [Herlocker et al., 2004].

A *preferences profile* for a user u, is denoted by PP_u, and is defined as the set of items visited by u together with the number of visits for each item, that is, $PP_u = \{\langle i, r_{u,i}\rangle \mid i \in \mathscr{I}$ and $r_{u,i}$ is the number of visits of u to $i\}$.

The *preferences matrix* is a *user-item* matrix comprising the number of visits r performed by each user u of $\mathscr{U}$, for each item i that belong to $\mathscr{I}$ (Table 1). If a user u has not yet visited an item i; missing values are predicted by the recommender system [Adomavicius and Tuzhilin, 2005]. $P_{u,i}$ denotes the prediction of a user u visiting an item i.

Table 1: The *preferences matrix* has values $r_{u,i}$ that can be predicted by a recommender system.

	i_1	i_2	$\cdots$	i_m
u_1	r_{u_1,i_1}	r_{u_1,i_2}	$\cdots$	r_{u_1,i_m}
u_2	r_{u_2,i_1}	r_{u_2,i_2}	$\cdots$	r_{u_2,i_m}
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
u_n	r_{u_n,i_1}	r_{u_n,i_2}	$\cdots$	r_{u_n,i_m}

The user *similarity* is a metric that establishes how similar are two users, u and v, regarding their preferences and is denoted by $sim(u,v)$. This value depends on the metric applied (see Section 3.2) and the number of visits performed by u and v to $\mathscr{I}$, the set of items.

3.2 Similarity metrics

Let be C a subset of items $\mathscr{I}$ where all items have been visited at least once by u and v, and let r_u and r_v be the vectors with the number of visits performed by u and v to the items in C, the similarity metrics between the two users are defined through the following equations [Herlocker et al., 2002, Herlocker et al., 2004, Segaran, 2007b]:

– *Pearson* correlation:

$$sim(u,v) = \frac{\sum\limits_{i\in C}[(r_{u,i} - \overline{r}_u)(r_{v,i} - \overline{r}_v)]}{\sqrt{\sum\limits_{i\in C}(r_{u,i} - \overline{r}_u)^2}\sqrt{\sum\limits_{i\in C}(r_{v,i} - \overline{r}_v)^2}} \tag{1}$$

– *Cosine* similarity:

$$sim(u,v) = \frac{\sum\limits_{i\in C} r_{u,i} \cdot r_{v,i}}{\sqrt{\sum\limits_{i\in C} r_{u,i}^2} \cdot \sqrt{\sum\limits_{i\in C} r_{v,i}^2}} \tag{2}$$

- *Spearman* correlation. In equation (3) the *ord* function refers to the ordinal associated to the element of the vectors r_u and r_v.

$$sim(u,v) = 1 - \frac{6 \sum_{i \in C} [ord(r_{u,i}) - ord(r_{v,i})]^2}{|C| * (|C| - 1)^2} \tag{3}$$

- *Manhattan* distance correlation; similarly to the previous correlations, the maximum correlation between u and v is reached when the evaluations of u and v coincide.

$$sim(u,v) = \frac{1}{1 + \sum_{i \in C} |r_{u,i} - r_{v,i}|} \tag{4}$$

- *Euclidean distance* is a metric related to *Pearson* correlation but less sensitive, since it is the square root of the sum of the squared differences between ratings. In many cases, Euclidean distances and Pearson Correlation yield similar results.

$$sim(u,v) = ||r_u, r_v|| \tag{5}$$

3.3 Content-based recommenders

Content-based recommendation techniques use resource information such as a resource description to determine items similarity and users past evaluation of similar items to determine the recommendations [Jannach et al., 2003]. In this case the user's visits to similar items are considered as the preferences profile $PP_u = \{i \in \mathscr{I} \mid u \text{ visited item } i\} = \mathscr{I}_u$.

Information retrieval techniques offer a variety of algorithms [Drachsler et al., 2010, Nadolski et al., 2009] that are often used to determine items similarity. It is assumed, in general, that items' content is plain text and for the case of non-text based items (e.g. images, videos, etc.), a set of plain text metadata is associated. Content-based techniques exploit either item's content or its metadata to implement similarity algorithms. In our study, items are digital resources with a variety of formats such as photos, spreadsheets, pdf documents, music files, web sites or multimedia applications, hence we consider items' content whenever possible as well as their metadata.

3.4 Collaborative Filtering

Collaborative filtering techniques make automatic predictions about the interest of a user, by collecting information about the preferences of other similar users (collaborative) in order to recommend items [Jannach et al., 2003, Segaran, 2007b]. Collaborative filtering can be also understood as the problem of predicting values for the *preferences matrix* [Adomavicius and Tuzhilin, 2005, Herlocker et al., 2002, Segaran, 2007b]. Collaborative filtering techniques typically follow a process of calculating the similarity between users, classifying users, predicting values for the missing rating in the preferences matrix and presenting the recommendation to the user; finally some recommending quality metrics are defined.

In this paper we followed a 4 step approach for generating such recommendations. These steps are applied for each one of the 5 similarity metric considered in this research, namely, Pearson correlation, Cosine, Spearman, Euclidean distance, and Manhattan [Herlocker et al., 2004, Manouselis and Costopoulou, 2008, Segaran, 2007b]:

Step 1: Calculating similarity. In our approach, given a similarity metric (of the five considered in this paper), we calculated all the values of a similarity matrix S containing all the similarity measures between users u and v and is denoted as $sim(u, v)$ $\forall u, v \in \mathscr{U}$.

Step 2: Search. For calculating the predictions or recommendations for a user we determine the users with similar interests. The set of k users that belong to $\mathscr{U}$ and are similar to u is called a neighborhood and is denoted by N_u. We calculate the best-k-neighbors [Herlocker et al., 2002] using a linear kNN search algorithm, with values of k ranging from 5 to 10. As mentioned before, these neighborhoods are calculated for each similarity metric considered in this research, in order to find the biggest neighbor with the highest similarity.

Step 3: Calculating the prediction values for the preferences matrix. We calculated the prediction values $P_{u,i}$, for each similarity metric and neighborhood size (k) combinations described in the previous step (section 4.4). The prediction values $P_{u,i}$ for the items that the user u has not yet visited are determined based on the visits defined in PP_u and the visits of the neighbors in N_u, through the equation 6 [Herlocker et al., 2002, Herlocker et al., 2004].

$$P_{u,i} = \overline{r}_u + \frac{\sum\limits_{v \in \mathscr{U}_i \cap N_u} [r_{v,i} - \overline{r}_u] \cdot sim(u, v)}{\sum\limits_{v \in \mathscr{U}_i \cap N_u} sim(u, v)} \tag{6}$$

Step 4: Finding the prediction error. In our approach, the last step consists of calculating the prediction error using metrics such as MAE (Mean Absolute Error), NMAE (Normalized MAE), Precision, Recall and Coverage (Section 3.5), in order to determine from there the best combination of k value and similarity metrics in terms of MAE.

Finally, we evaluate the results. In our experiments, out of the five metrics and k neighborhoods combinations, we found that the smallest MAE value is 0.643978 for a k of value 10 and Euclidean Distance metric.

3.5 Prediction validation

Recommenders quality is determined through measures such as the mean absolute error (MAE) defined in (7), that is, the difference between the prediction and the actual user preference, and the normalized mean absolute error (NMAE) defined in (8) [Goldberg et al., 2001, Herlocker et al., 2004, Herlocker et al., 2002].

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |r_{u,i} - P_{u,i}| \tag{7}$$

$$\mathrm{NMAE} = \frac{\mathrm{MAE}}{\overline{r}_{max} - \overline{r}_{min}} \tag{8}$$

Other measures such as *precision* (P) is defined as the ratio between the recommended relevant items and all the items that were recommended (9). *Recall* (R) represents the probability that a relevant item is selected (10); and finally *Coverage* (C) is the percentage of the items that the system can make predictions for (11) [Herlocker et al., 2004].

Table 2: Recommenders quality depend on the system capacity for addressing the proper items subset

	Selected	Not Selected	Total
Relevant	I_{rs}	I_{rn}	I_r
Irrelevant	I_{is}	I_{in}	I_i
Total	I_s	I_n	I

$$P = \frac{I_{rs}}{I_s} \tag{9}$$

$$R = \frac{I_{rs}}{I_r} \tag{10}$$

$$C = \frac{I_s}{I} \tag{11}$$

4 Determining the configuration of the Recommender model

Our work is based on a dataset from the EducarChile website[1]. This website contains learning resources and is widely used by teachers and students in Chile. We took a snapshot of the dataset from 2007 to 2011 (5 years). The dataset contained information from registered users such as, students, teachers, administrative staff, researchers, students' tutors, and general public (i.e. non registered users). In this research we considered registered users only, in the teacher category. They were forced to register in order to interact with the system. The information collected from the users was in two categories, the personal category (name, gender, nationality, document DNI) and the teacher category (subject: e.g. mathematics, language, sciences, biology, etc. Level: e.g. 1st primary, 1st high school, etc.). This information is not publicly available but was made available for this research confidentially.

4.1 Analyzing the educational dataset

The dataset contained learning resources produced by EducarChile as well as resources uploaded by its users. Resources metadata were annotated by EducarChile experts or its users and in the latter case were revised and normalized by EducarChile experts. The dataset contained also information about users' interests (i.e. learning resources visits and learning resources ratings), learning resources, learning resources metadata, etc. All metadata in the dataset was reviewed and modified if necessary by a set of educational experts in EducarChile as part of a regular curation process. EducarChile performed no specific curation process on the snapshot used in this research. EducarChile experts defined a specific set of metadata categories, however, when we analyzed the metadata associated to the clusters obtained from the hierarchical clustering process, most metadata categories were missing, so that, we do not believe that such categorization process affects the results of our approach.

The dataset originally weighted 3TB on average, including data that was not relevant to our study. We filtered the dataset and considered only information regarding learning resources, learning resources metadata, registered users (teachers), and user's visits and rating. The information considered from learning resources is: ID, name, description, keywords, title, relevance, resource Web identifier (url), resource metadata identifier (code), and resource metadata value. The information considered for learning resources metadata were: ID metadata, ID value, and value which is a description of the ID value field. Samples of the metadata values (and a brief statistic) are shown in Table 5. The information requested from the users was detailed before in this Section. The dataset characteristics can be appreciated in table 3.

[1] www.educarchile.cl

Table 3: Dataset characteristics

Characteristic	Number	Percentage
Registered teachers (users)	11.260	100%
Learning resources (items)	43.092	100%
Total items' *visits*	60.651	100%
Total visited *items*	5.621	13%
30-top visited items	12.736	21%
Users who visited at least 10 different items (U_{10})	1.376	12.22%
Items visited by U_{10}	4.538	7.48%
Number of visits generated by U_{10}	25.544	42.11%

The dataset contained 60.651 visits from teachers (registered users); however, such users visited only 13% of the learning resources. Furthermore, the top 30 visited learning resources (from now on *items*) represented 21% of such visits. In order to reduce the dataset sparsity and to perform training and testing using a *ten-fold cross-validation* approach [Devijver and Kittler, 1982], we limited the users set to those who visited at least 10 different resources, resulting in 1.376 users, 4.538 learning resources, and 25.544 visits. Table 4 the distribution of visits to items.

Table 4: U_{10} visits distribution per item

Number of visits	Visited items	Percentage
1	16.933	66.3%
2	6.113	23.9%
3	1.164	4.6%
4	732	2.9%
5	602	2.4%

Notice that the number of visits range from 1 to 5; for greater number of visits (only 3 occurrences) we changed them to 5. The items that were visited by teachers 1 or 2 times represent the 90.2% of the dataset, however, single visits represent 66.3% of the total number of visits. For these reasons we considered visits as a unary implicit rating [Herlocker et al., 2004] where each visit is accounted as a users interest on a resource.

4.2 Metadata

Learning resources were annotated with 81 different metadata (see table 5) each one with a fixed list of possible values for a total of 6.957 values, all of them in use in the dataset. We classified metadata into three categories: Website administration (included 19 metadata elements), socio-pedagogical metadata (62 metadata elements) and curricular metadata (level and subject), which is a subcategory of the socio-pedagogical category. We also included the free, plain text, written in natural language, used to describe resources in DB fields such as name, title and description. Notice that the EducarChile platform allows teachers to submit diverse material and later a group of experts revise, complete, discuss, annotate or modify the metadata and all the descriptive elements. Table 6 presents samples of the metadata values for two resources in their corresponding categories.

Table 5: Metadata classification

Category	Metadata entries	Metadata Values	Description
Website administration metadata	19	774	References to other websites, website administrative keywords, material classifications to be used by the page renderer, etc.
Curricular metadata	2	2.994	Includes level (e.g. 1st primary, 3rd secondary, etc.) and subject (e.g. language, communication, etc.)
Socio Pedagogical metadata	62	6.183	Includes the curricular metadata (2.994 values) and adds socio-pedagogical metadata (3.189 values) such as educative usage, Dewey taxonomy, parents academic background, etc.
Content	3	-	Name, title, description.

For the case of curricular metadata, the Chilean educational system (not considering university or tertiary education) includes three levels: preschool (left out of this study), primary, and secondary levels. Primary level includes students from 6 to 13 years old, enrolled in 1st to 8th grade respectively. Secondary level comprehends 4 years (1st to 4th grade) for students from 14 to 17 years old. The primary level curriculum is uniform for all schools, but the secondary level is divided into vocational-technical and science-humanities schools. Both systems share the same curriculum for the first two years. Teachers, on the other hand, are specialized according the three levels, that is, preschool, primary and secondary teachers. Primary teachers are divided into two groups, the generalists that are responsible for a group of students from 1 to 4th grades and teach all the required subjects, and those who are specialized in certain subject (e.g. mathematics) and teach diverse students groups from 5 to 8th grades. Secondary teachers are also specialized in a subject that is taught to diverse students groups from 1st to 4th grades.

Table 6: A sample of Metadata values per category for two items

Item name	Website metadata	Curricular metadata	Socio Pedagogical metadata
Jose de Espronceda, biography	site; education; biographies; literature; student; education; c-h (humanist scientist); t-p (professional technician); international	1st high school; communication; oral; reading	site; Jose; Espronceda; biography; page; spend; poet; Spanish; feature; work; author; writer; romantic; romanticism; Spanish; poetry; poems;
Chronicle of the twentieth century	software; education; student; 2nd grade; c-h (humanist scientist); t-p (professional technician)	8th grade; history; geography; science; social; 4th grade; 4th grade; America; Latin; contemporary; world; current	sw (software); chronic; century; xx; equipment; reference; form; archive; journalism; news; event; discovery; relevant; kind; world; contemporary; war; history; revolution; Russian; Vietnam; science; social; biography; communism; EducarChile; calendar; America;

4.3 Text Processing

In order to perform text-based analysis based on the information describing the learning resources we pre-processed the text in order to eliminate terms such as articles, conjunctions, etc. and reduce verb variations (e.g. run, ran and running is reduced to run). This preprocessing task was performed on the metadata words as well as on the whole text for the case of the content category. We preprocessed the text using stemming and lemmatization algorithms provided by the *Freeling*[2] *analyzers* [Padró and Stanilovsky, 2012], since our resources and metadata are written in Spanish. Finally we ran a mapping algorithm in order to reduce the set of possible values for curricular metadata due to different naming conventions (e.g. NB1 -basic level 1-, 1st primary, 1 primary, NB-1, etc. to 1st primary).

4.4 Applying collaborative filtering on the dataset

We applied the collaborative filtering process defined in Section 3.4 and produced recommendations using all the similarity metrics defined in Section 3.2 in order to determine the strategy and neighbor size that produced the best results (i.e. minimal MAE). To validate our results we used a *ten-fold cross-validation* approach. That is, we divided the dataset into 10 equivalent segments randomly, then we run 10 iterations and at each iteration we pick one (different) segment to serve as a gold standard. We called this segment the *Testing set* (TS) and called the remaining 9 *Training set* (TrS). We average quality metrics at each iteration, and then averaged the metrics for the 10 iterations. We perform this approach in order to use the dataset itself as gold standard (since it contains users actual interest on the resources) and we perform cross validation in order to minimize the noise of the data through randomization. The TrS and TS sets must satisfy the following conditions:

1. $|\mathscr{I}_u| \geq 10, \forall u \in \mathscr{U}$. That is, to guarantee that for each iteration at least one prediction must be realized for each user, and
2. $S = (TrS \bigcup TS)$ and $(TrS \bigcap TS) = \emptyset$

The recommendations were analyzed using MAE (equation 7), NMAE (equation 8), Recall (equation 10), Precision (equation 9), and Coverage (equation 11) metrics. Table 7 presents a summary of the top 10 best results considering the similarity metrics and neighbor sizes (kNN search) previously defined, ordered by MAE. Notice that the table does not include *Cosine, Spearman* and *Pearson* results since they perform worst than the top ten. Since the best results in terms of MAE are yielded by Euclidean distance similarity with a value of $k = 10$ for kNN, we choose this configuration for the remainder of the process.

In our case the differences between the various strategies are minimal when considering any of the applied metrics, but the recall (probability that a relevant item is selected) is very low; this may be explained by the sparsity of the visits

[2] `http://nlp.cs.upc.edu/freeling/`

Table 7: The best ten algorithms in terms of MAE, Recall and Coverage.

Rank	Similarity Metric	kNN search	MAE	NMAE	Recall	Precision	Coverage
1	Euclidean	10	0.643978	0.12880	0.04570	0.18830	0,86239
2	Euclidean	9	0.644287	0.12886	0.04570	0.18659	0.86239
3	Manhattan	10	0.644506	0.12890	0.04437	0.18416	0.71152
4	Euclidean	8	0.644702	0.12894	0.04836	0.19326	0.86239
5	Euclidean	7	0.644773	0.12895	0.04925	0.19072	0.86239
5	Manhattan	9	0.644774	0.12895	0.04392	0.18066	0.71152
7	Manhattan	7	0.644969	0.12899	0.04703	0.18435	0.71152
8	Manhattan	8	0.645114	0.12902	0.04614	0.18705	0.71152
9	Euclidean	6	0.645302	0.12906	0.05235	0.19799	0.86239
10	Manhattan	6	0.645558	0.12911	0.05013	0.19120	0.71152

[Ghazarian et al., 2014, Herlocker et al., 2004] (90.2% of items received 1 or 2 visits, see Table 4). Results, however, are comparable to similar experiments in education [Manouselis and Costopoulou, 2007, Manouselis and Costopoulou, 2008].

For instance in [Manouselis and Costopoulou, 2008] an experiment based on a multiattribute evaluation of learning resources by teachers is performed. The dataset was created for the task and the sparseness problem is avoided since all the available teachers evaluate a limited set of resources. In this case, the best metric is MAE=0.57 whereas Coverage remains as 69.08% for a Cosine algorithm and $k = 4$. The use of a multiattribute evaluation instead of a discrete (0 or 1) one is accounted for the improvement of the error. Precision and recall metrics are not reported. In [Verbert et al., 2011] MAE metrics for a dataset in education are very similar for Cosine and Pearson similarity, again, precision and recall metrics are not reported. In [Zhao et al., 2015] an item based collaborative filtering algorithm is used to recommend learning resources (video) in a distance learning setting, the dataset is characterized by a growing sparsity (40 million users, 9 thousand items). In this case the best MAE value falls into 0.8336 whereas Coverage remains as 99.8% for a sparsity of the training set of 94.86%. In our case, only considering the subset where users have visited at least 10 items, we have a sparsity level of 99.59%. Precision and recall metrics are not reported.

5 Facing the cold start problem

In order to test the impact of the diverse metadata, we used vector terms describing resources and hierarchical clusters of teachers to face the cold start problem.

First, we agglomerated the kNN search neighborhoods into hierarchical clusters of teachers, and then we described such clusters by term vectors. The terms are pulled from the metadata associated to the learning resources visited by the users in the clusters. New users are described also through terms contained in a teacher profile and both kinds of terms are used to determine the set of clusters to which the user belongs. A new item of a new user is compared to the items of the user cluster based on the new item's term vector. The most similar items the users belonging to a cluster are used to predict a rating for a new item. We detail our approach in this section and we also conducted an experiment with teachers in training (gold standard), considered as the new users, and present our prediction results in terms of MAE. In our approach, we differentiate terms into three categories according to the metadata categories defined in Table 5, so that our procedure and results follows such differentiation.

5.1 Hierarchical clustering of teachers

In this section, we perform an agglomerative hierarchical clustering of users [Hastie et al., 2009], which is a technique that groups elements together (or them) into clusters based on certain criteria. The agglomerative or ascending method consists of forming clusters progressively by adding members until a conglomerate is created. To agglomerate members it is necessary to define a metric that determines which elements, which do not belong yet to a cluster, are located at distance. Metrics such as *Pearson*, or *Euclidean distance* are typically used.

Remember that in Section 4 we applied Euclidean distance and kNN search ($k = 10$) to recommend learning resources based on users visits. In this section, we agglomerate the neighborhoods of similar users found in Section 4 based on a threshold p of shared users. We determined a metric of *similar neighborhoods*. That is, two neighborhoods N_i and N_j are *similar*, denoted by $N_i \sim N_j$, when they share at least certain percentage p of neighbors (teachers). Equation (12) formally describes the similarity relation.

$$N_l \sim N_m \Longleftrightarrow |N_l \cap N_m| \geq p \cdot |N_l| \text{ and } |N_l \cap N_m| \geq p \cdot |N_m| \tag{12}$$

Based on equation (12) we can form the set of hierarchical clusters $\mathscr{C}$, denoted by equation (13).

$$\mathscr{C} = \bigcup_{i \neq j} N_i \text{ where } N_i \sim N_j \tag{13}$$

Table 8 presents the distribution of *clusters* according to p, the percentage of shared members, calculated with our dataset. The table shows, for instance, that for a value of $p = 0.85$ there are 1.338 clusters with a single element (just one neighborhood), 24 clusters group together 2 neighborhoods, 3 clusters group 3 neighborhoods, etc. As we can see the more restrictive the similarity threshold the less the chances to agglomerate clusters in many levels. When considering the highest values of p as seen in Table 8, for a value of $p = 0.85$ we can obtain clusters containing $1, 2, \geq 7$ agglomerated neighborhoods, but 0 clusters containing 5 agglomerated neighborhoods. A higher value, $p = 0.9$ produces more single neighborhood clusters and clusters with 2, 3 and 4 agglomerated neighborhoods but 0 clusters with 5,6 7 or more agglomerated neighborhoods. Hence, we choose a value of $p = 0.85$ since it guarantees a high number of shared users but also various agglomerated clusters.

Table 8: Number of neighborhoods agglomerated for various values of p

p	$\#\mathscr{C} = 1$	$\#\mathscr{C} = 2$	$\#\mathscr{C} = 3$	$\#\mathscr{C} = 4$	$\#\mathscr{C} = 5$	$\#\mathscr{C} = 6$	$\#\mathscr{C} \geq 7$
$p = 0.75$	1282	38	13	5	5	1	32
$p = 0.80$	1312	36	2	5	2	2	17
$p = 0.85$	1338	24	3	3	0	1	7
$p = 0.90$	1354	17	3	2	0	0	0

5.2 Identifying clusters metadata

The users of a cluster, denoted by $\mathscr{U}_{\mathscr{C}}$, comprehend all the users belonging to the cluster's neighborhoods (14). Similarly, the items in a cluster, denoted by $\mathscr{I}_{\mathscr{C}}$, comprehend the items visited by the users of a cluster (15).

$$\mathscr{U}_{\mathscr{C}} = \{u \in \mathscr{U} \mid \exists N \in \mathscr{C} \wedge u \in N\} \tag{14}$$

$$\mathscr{I}_{\mathscr{C}} = \{i \in \mathscr{I} \mid \exists u \in \mathscr{U}_{\mathscr{C}} \wedge i \in \mathscr{I}_u\} \tag{15}$$

The information describing the items (words or *terms*) are differentiated according to the *curricular* and *socio pedagogical* metadata, as well as *content.* The terms describing a resource (e.g. text-based, videos, etc.) are obtained from the metadata values (which could be a small sentence of up to 100 terms), resources title, description and name as detailed in Table 5.The terms describing a cluster, denoted by $\mathscr{T}_{c,\mathscr{C}}$, correspond to the union of all the terms describing the items of a cluster $\mathscr{I}_{\mathscr{C}}$ (equation 16), according to the corresponding category.

$$\mathscr{T}_{c,\mathscr{C}} = \{t \in i \mid i \in \mathscr{I}_{\mathscr{C}}\} \tag{16}$$

The relevance of a term t describing a cluster $\mathscr{C}$ according to a category, denoted by $w_c(t)$, is based on a) the number η of occurrences of the term t describing an item i ($\eta(i,t)$), and b) the number of users u that have visited an item i ($\kappa(\mathscr{U}_{\mathscr{C}}, i)$) containing the term t (17). In equation 17, the subscript c in the expression $w_c(t)$ denotes a metadata category for the term (e.g. curricular, socio pedagogical, etc.).

$$w_c(t) = \sum_{i \in \mathscr{I}_{\mathscr{C}}} \eta(i,t) \cdot \kappa(\mathscr{U}_{\mathscr{C}}, i) \tag{17}$$

A vector of weighted terms $w_c(t)$ is associated with each cluster per category. Table 9 presents a snippet of the vectors associated to two clusters $C1$ and $C2$ that are representative of the analysis. Each cluster aggregates 3 or more neighbors with a $p = 0.85$ threshold of shared members. To facilitate the analysis of the information, we considered only the top-40 terms associated to the cluster and we separate the terms into three additional subcategories: subject, grade and other terms; the separation of terms was performed manually by educational experts. We present only the top five terms in each subcategory.

As we can notice, the terms with heavier weight for the curricular metadata correspond to the subcategories subject and grade, whereas for the socio-pedagogical category, the heavier sub-category is *other terms*, describing mainly technical aspects. The weight of the terms in the content category are heavier for the subject and *other terms* sub-category, however, for the case of the subject generic terms such as *secondary*, *science-humanities*, and *vocational-technical*, instead of the subject details.

In order to test the impact of the diverse information category on the ability to deal with the cold-start problem, we used the vector terms and cluster to face the cold start problem. We followed a hybrid approach; relating new users to the users in the predetermined clusters, and new items to the items in the cluster chosen for the new user. The users of the analyzed dataset were considered as the *old users* and we conducted an experiment with teachers in training that we considered as the *new users*; the experiment is described as follows.

5.3 Experimental setup

We assumed that new teachers have associated a basic *teacher profile* whose features are detailed in Table 10. The new users set was comprised of 39 teachers in training, that is, senior students of Education that work as teachers in schools. They were specialized in various subjects as follows: 7 of them specialized in Mathematics, 7 in language and communication, 8 in language and mathematics, and 18 generalists.

Following a methodology similar to [Manouselis et al., 2010], we used a questionnaire and ask the new users to provide a *teacher profile* and to rate 6 items

Table 9: A sample of the top 5 terms of clusters C1 y C2 is presented. The terms are followed by weight and are classified into the subcategories *Subject*, *Grade* and *Other terms* for cluster analysis.

Cluster	Subject	Grade	Other terms
Curricular metadata			
C1	communication: 12.8 history: 4.5 science: 4.5 social: 4.2 language: 4	5th primary: 12.4 1st secondary: 9 8th primary: 7.9 6th primary: 5.9 3rd primary: 2	readings: 2.7 environment: 2.1 verbal: 2 interaction : 1.7 organism: 1.5
C2	science: 7.3 environment: 6.1 interaction: 4.8 function: 3.6 biology: 2.4	1st secondary: 11.9 5th primary: 6.8 2nd secondary: 2.7 8th primary: 2.6 1st primary: 2.2	natural: 8.5 chemical: 4.6 organism: 4.4 living_organism: 4.2 structure: 4.2
Socio Pedagogical metadata			
C1	language: 0.23 mathematics: 0.19 history: 0.18 communication: 0.05 music: 0.04	primary: 0.02	type: 53.55 site: 21.22 site_type: 8.47 educarchile: 3.69 article: 3.66
C2	human: 0.35 chemistry: 0.34 physics: 0.34 arts: 0.16 biology: 0.1		type: 54.55 site: 17.17 site_type: 8.24 article: 4.16 educarchile: 2.93
Content			
C1	secondary: 19.94 science-humanities: 10.90 vocational-technical: 7.55	2nd primary: 6.99 1st primary: 2.23	student: 36.35 education: 13.07 text: 1.68 activity: 0.49 article: 0.34
C2	secondary: 22.04 science-humanities: 13.23 vocational-technical: 9.60	2nd primary: 3.88 1st primary: 1.76	education: 17.65 student: 29.34 text: 1.89 teacher: 0.19 activity: 0.28

from a set of 22 learning resources available in the `EducarChile` website that were not previously visited by any old user. The items corresponded to Mathematics (11) and Language and Communication (11) subjects. 4 items were mandatory

Table 10: Teachers' Profile features

Category	Examples
Subject	Mathematics, Language, ...
Grade	1st primary, 1st high school, ...
School type	Public, private or mix
Geographic location of the school	Urban or rural

for all the participants (were assigned to the 39 new users), whereas the remainder items (18) were randomly assigned. We assigned the evaluations in this way since the number of participants was small and we wanted to guarantee the existence of more than one evaluation for at least 18% of the dataset. The learning resources assigned and evaluated items are shown in Table 11. We pre validated the design of the experiment with educational experts and we found that teachers may rate the resources in a scale broader than 0 to 1 (i.e. I would visit the resource) since the resources may fall out of the teachers expertise, become not interesting at all, or even considered as bad resources. For this reason we defined a 5-point scale (see Table 12) in order to capture the teacher perception and later defined a mapping scale as explained in Section 5.6.

Table 11: Number of evaluations per learning resource

Resource	Subject	Assigned	Evaluated	Resource	Subject	Assigned	Evaluated
1	Language	39	39	12	Maths	39	37
2	Language	39	38	13	Maths	39	39
3	Language	4	4	14	Maths	4	4
4	Language	5	5	15	Maths	5	5
5	Language	4	3	16	Maths	4	4
6	Language	3	3	17	Maths	3	3
7	Language	5	5	18	Maths	6	6
8	Language	4	2	19	Maths	5	5
9	Language	2	2	20	Maths	2	2
10	Language	2	2	21	Maths	2	1
11	Language	8	8	22	Maths	8	8

Table 12: Evaluation scale followed by the new teachers to evaluate new items

Rate	Meaning
1	Useless
2	Somehow useful
3	Regularly useful
4	Useful
5	Very Useful

5.4 Classifying new users

We define $\mathscr{U}_N$, the set of new users, and $\mathscr{U}_{DS}$, the set of all the users in our dataset; $\mathscr{I}_N$, the set of new items and $\mathscr{I}_{DS}$, the set of all the items in the dataset. Let be $\mathscr{T}$ the set of terms contained in the teachers' profile and the dataset items differentiated by category.

Each cluster $\mathscr{C}$, defined in section 5.1, is represented as a vector of weighted terms $\overrightarrow{\mathscr{C}} = \langle w(t_1), w(t_2), \ldots, w(t_n)\rangle$, where $w(t_i)$ is calculated by equation 17. For each new user $nu \in \mathscr{U}_N$, a vector of weighted terms, as detailed in their corresponding teacher profiles, is defined as $\overrightarrow{nu} = \langle w_u(t_1), w_u(t_2), \ldots, w_u(t_n)\rangle$, where $w_u(t_k)$ $(k = 1, \ldots, n)$ is defined by equation (18), for the content of the user profile.

$$w_u(t_k) = \frac{P_u(t_k)}{|P_u|} \tag{18}$$

In equation 18, $P_u(t_k)$ denotes the number of occurrences of the t_k term in the *teacher profile* for a user u, whereas $|P_u|$ denotes the total number of terms for such profile. To define the terms of a teacher's profile, we considered all the terms contained in all the profile fields and we pre-processed such text in the same way we did with the resources metadata to facilitate matching (i.e. we used Freeling analyzers).

The similarity between the new user and the old users, classified in the corresponding clusters ($\mathscr{C}$), is calculated using the cosine equation. The resulting tuples (19) allow us to define the set of top-3 most similar clusters for a new user denoted by $\mathscr{C}^*_{c,nu}$ (20), where the subscript c denotes the metadata category and $k = 1...3$ the top-three clusters.

$$\{(\mathscr{C}_1, sim(\overrightarrow{nu}, \overrightarrow{\mathscr{C}}_1)), (\mathscr{C}_2, sim(\overrightarrow{nu}, \overrightarrow{\mathscr{C}}_2)), \ldots, (\mathscr{C}_m, sim(\overrightarrow{nu}, \overrightarrow{\mathscr{C}}_m)))\} \tag{19}$$

$$\mathscr{C}^*_{c,nu} = \{\mathscr{C}_{k=1\ldots3} \mid sim(\overrightarrow{nu}, \overrightarrow{\mathscr{C}}_k) > sim(\overrightarrow{nu}, \overrightarrow{\mathscr{C}}_j)\ \forall j \neq k\} \tag{20}$$

5.5 Predicting a rating for the new items

For the case of the new items $ni \in \mathscr{I}_N$, we find the most similar items $i_v \in \mathscr{I}_{\mathscr{C}}$, that have been visited by the users $v \in \mathscr{C}$. We use various similarity thresholds consisting on a percentage p of terms shared by the similar items i_v and ni. That is, $\mathscr{I}^p_{\mathscr{C},ni} = \{i_v \in \mathscr{C} \mid \text{number of terms of } i_v > p\cdot(\text{number of terms of } ni)\}$. Then, for each new item (ni) we calculate the rating $R_{\mathscr{C}c,ni}$ given from the old users (v) as defined in equation (21). The old user ratings are defined as the number of users visits (r_v) for the items (i_v) that are similar to the new item ni within the boundaries of the top-3 clusters $\mathscr{C}_c$ to which the new user nu belongs (see equation 21). Notice that if two items share various terms but such terms differ on their weights (in the diverse metadata categories), we can take into account such difference in equations 18 to 21.

$$R_{\mathscr{C}c,ni} = \frac{\sum\limits_{\substack{v \in \mathscr{U}_{\mathscr{C}} \\ i_v \in \mathscr{I}^p_{\mathscr{C},ni}}} r_{v,i_v}}{|\mathscr{I}^p_{\mathscr{C},ni}|} \tag{21}$$

Finally, we consider only the top-3 clusters most similar to the new user $(\mathscr{C}^*_{c,nu})$ and we averaged the ratings given by the clusters' users in order to predict a rating R for a given item ni and a given user nu as defined in equation 22.

$$R_{\mathscr{C}^*_{c,nu},i} = avg(R_{\mathscr{C}c,i})\ \forall, \mathscr{C} \in \mathscr{C}^*_{c,nu} \tag{22}$$

Since the nu's ratings correspond to a scale from 1 to 5 and they differ from the dataset rating criteria (visits that range from 0 to 1), we defined an equivalence rule. Hence, for a user $nu \in \mathscr{U}_N$ that evaluates a new item ni with rate $r_{nu,ni}$ the equivalence rule $e(r_{nu,ni})$ is defined by equation (23). This equation is used to calculate the corresponding values for the teachers in training experience in order to compare the predictions of the system (visits) with the actual user preferences (5-point scale) and hence to calculate the difference (MAE value).

$$e(r_{nu,ni}) = \begin{cases} 0 & \text{, if } r_{nu,ni} = 1 \vee r_{nu,ni} = 2 \\ 0.5 & \text{, if } r_{nu,ni} = 3 \\ 1 & \text{, if } r_{nu,ni} = 4 \vee r_{nu,ni} = 5 \end{cases} \tag{23}$$

5.6 Analyzing the predicted ratings

In order to determine the effect of our approach, we implemented a collaborative filtering strategy, considering only the results obtained in the experiment described in section 5.3. We divided the dataset in two subsets, one (76 evaluations) containing the evaluation of two common objects (remember that users evaluated 4 objects in common) and the other one containing the remainder (149 evaluations). We use the former to make predictions and the latter as gold standard in order to determine the error impact (MAE [Herlocker et al., 2004]) of recommending items. Table 13 shows the results considering four different similarity metrics. As we can observe the results are really poor, and probably is due to the size of the sample.

Table 13: MAE results when recommending resources based on collaborative filtering and the new users data only, considering various similarity metrics.

Ranking	Similarity metric	MAE
1st	Spearman	1,227935814
2nd	Pearson	1,44987662
3rd	Manhattan	1,472972216
4th	Euclidean Distance	1,507429683

Remember that in order to face the cold start problem, we considered the users and the items in the experiment as new users and new items in relation to the EducarChile dataset presented in section 4.1, and that in section 5.4 we classified new users into the neighbors found in section 5.1. Here, we calculated $R_{\mathscr{C}^*_{c,nu},i}$ for the items $ni \in \mathscr{I}_N$ that were evaluated by the new users ($nu \in \mathscr{U}_N$), following the approach presented in section 5.1. We obtained results for the top 3 clusters for each metadata category and averaged the results across clusters and metadata categories. We compared the calculated ratings with those actually assigned by the users in the teachers-in-training experiment considering the whole set of answers as gold standard) divided into three categories, ratings ≥ 4, ratings ≥ 3 and ratings ≥ 1, for each category (metadata and ratings) we varied the similarity threshold p.

Table 14 presents our results. We can notice that when considering *socio-pedagogical* and *content* metadata is not even possible to find similar objects that fulfills a threshold stronger than 30%.

Furthermore, in figure 1 we notice that there are no significant differences between metadata category predictions and 20-30 % (except for *content* metadata for 30% threshold). The curricular metadata category presents the best results. For the items with good evaluations from the new users, the voting estimation based on the old users (dataset) is very close to the ratings given by the users (MAE = 0.4814), on a strict threshold (e.g. 70%). However, when we include

Table 14: MAE obtained when predicting ratings considering various similarity thresholds

Metadata Category	Voting given by nu to ni	Similarity thresholds						
		20%	30%	40%	50%	60%	70%	80%
Curricular	>=4	0.4974	0.4748	0.4775	0.4723	0.4936	0.4814	
	>=3	0.6100	0.5936	0.5910	0.5776	0.5988	0.6085	
	>=1	1.0133	0.9886	0.9930	0.9825	1.0105	1.1638	1.4606
Socio Pedagogical	>=4	0.5546	0.4139					
	>=3	0.6906	0.5773					
	>=1	1.1431	0.8915					
Content	>=4	0.5440						
	>=3	0.7019						
	>=1	1.0784	2.2915					

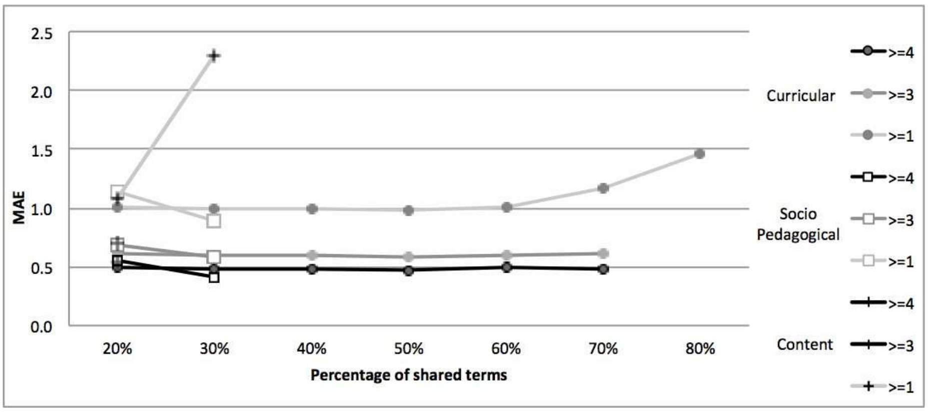

Figure 1: MAE comparison according metadata category and similarity threshold

ratings indicating a lack of interest in a resource the MAE worsens (1.1638).

6 Conclusions and Future Work

Recommender systems can be used as a means to face the over abundance of resources available in the Web, by acknowledging the goals of specific communities of practice, such as teachers, if the recommender algorithms consider the community properties and values. The metadata associated with the resources become fundamental to deal with overabundance of material and also the diversity of users' preferences. Nadolski et al. [Nadolski et al., 2009] evaluate the effects of using an ontology that models user preferences and finds that an ontology-based strategy is more accurate although is costly and complex. Other researchers [Tiropanis et al., 2009] aim for a light-weight approach when faced with the complexity of IEEE-LOM metadata. In this paper we explore the impact of metadata considering primary and secondary levels in the Chilean educational system. The metadata was thoroughly categorized and reviewed by educational experts along 5 years. Such metadata differ from standardized approaches such as IEEE-LOM in the sense that places more emphasis on pedagogical and curricular aspects.

We expect to contribute with the design of search engines for learning resources, mainly by recognizing the need of clusters of shared interests, rather than mere discrete facets, that mix metadata, particularly curricular metadata. Our approach allowed us to recognize teachers' interest for structuring their communities and learning resources and suggests the relevance of curricular metadata as sources that represent teachers' interests. Notice, however, that even though the subject is a good discriminator, the grade. For instance, when considering *grade* in curricular metadata, all the clusters presented the same terms describing almost all educational level but with different emphasis. In Table 9, for cluster C1 the most important metadata for grade is *5th primary*, whereas for grades above (6th primary, 8th primary, 1st secondary) the relevance decreases. C1 has a strong emphasis on 5th grade of primary whereas cluster two placed an emphasis on 1st grade of secondary. The same behavior repeated in all clusters. This is different to the *subject* metadata in the same category since the terms are semantically related giving a sense of coherence, it is clear in Table 9 that C1 main interest is related to humanities (perhaps social science) whereas in C2 the interest is oriented to life sciences (perhaps biology). This may suggest that a single resource can be reused in various grades for different purposes, could be fundamental in some cases but introductory in another. This has a direct effect on search, since most learning repositories' search engines use faceted search or keywords considering either the subject or the grade but do not take into account that resources may be used in mixed categories with weights depending on the searcher interests which can be taken into account at a low cost (i.e. asking him or her to fill up a profile). So that, it is important to discover the proper combination of relevance of subject and grade instead of consider them as absolute discriminators (e.g. considering that a teacher may be interested in one grade only).

The metadata and our approach allow us also to estimate ratings for new users and new items in order to face the cold start problem. Again, the curricular metadata becomes fundamental for minimizing the rating prediction error. That is, the recommendation results in terms of MAE, following a pure collaborative filtering approach, are significantly bad as seen in Table 14, whereas recommendations using our approach for cold start (i.e. considering teachers in training as new users) are much better. Furthermore, when calculating recommendations based on learning resource content (title, description and resource name) our approach is better but makes recommendations on resources that share only 20% of shared terms with the new item. However, when considering curricular metadata, the results are better in terms of MAE even though we become stricter regarding the percentage of shared terms.

Our approach is limited to the dataset provided by EducarChile; nevertheless, we believe these results may be extended to other primary and secondary educational datasets since they were collected under a natural setting. Event hough educational experts curated the metadata, the inconsistencies in the categorization naming as well as the low significance of the categories themselves demonstrate that our approach does not depend on such curation task. A strong limitation of our approach resides on the size of the teachers-in-training experiment to determine new users and new items. We believe the size of the experiment shall be increased for eliminating biases and improving the results. As for future work we plan to test various dimension reduction techniques on the dataset in order to reduce its sparseness degree as well as to try machine learning techniques in order to derive an automatic classification of metadata. We will evaluate the LDA (Latent Dirichlet Allocation) algorithm particularly since it is a generative technique that may allow an automatic classification of metadata (and hence items and users).

Platforms aimed at training and education generate virtual shared spaces for learning, foster the development of interactive teaching/learning activities and opens the opportunity to various actors in the education field to participate, interact and collaborate. Search and recommender systems aimed at primary and secondary school level teachers must consider the structure of such users practice. Recommender systems integrated in these environments may assist users with shared interests if we take into account that they form a community with particular characteristics. For the case of primary and secondary school level teachers, the discovery of such communities must consider curricular metadata in order to account for the practices and interests of such users. A proper search or recommender engine becomes fundamental to achieve such goals.

References

[Adomavicius and Tuzhilin, 2005] Adomavicius, G. and Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Trans. Knowl. Data Eng.*, 17(6):734–749.

[Canini et al., 2013] Canini, L., Benini, S., and Leonardi, R. (2013). Affective recommendation of movies based on selected connotative features. *Circuits and Systems for Video Technology, IEEE Transactions on*, 23(4):636–647.

[Carrer-Neto et al., 2012] Carrer-Neto, W., Hernandez-Alcaraz, M. L., Valencia-Garcia, R., and Garcia-Sanchez, F. (2012). Social knowledge-based recommender system. application to the movies domain. *Expert Systems with Applications*, 39(12):10990 – 11000.

[Chatti et al., 2013] Chatti, M., Dakova, S., Thus, H., and Schroeder, U. (2013). Tag-based collaborative filtering recommendation in personal learning environments. *IEEE Transactions on Learning Technologies*.

[Colombo-Mendoza et al., 2015] Colombo-Mendoza, L. O., Valencia-Garcia, R., Rodriguez-Gonzalez, A., Alor-Hernandez, G., and Samper-Zapater, J. J. (2015). Recommetz: A context-aware knowledge-based mobile recommender system for movie showtimes. *Expert Systems with Applications*, 42(3):1202 – 1222.

[Dascalu et al., 2014] Dascalu, M.-I., Bodea, C.-N., Lytras, M., de Pablos, P. O., and Burlacu, A. (2014). Improving e-learning communities through optimal composition of multidisciplinary learning groups. *Computers in Human Behavior*, 30:362 – 371.

[Devijver and Kittler, 1982] Devijver, P. A. and Kittler, J. (1982). *Pattern Recognition: A Statistical Approach*. Prentice Hall; First Edition edition.

[Drachsler et al., 2010] Drachsler, H., Bogers, T., Vuorikari, R., Verbert, K., Duval, E., Manouselis, N., Beham, G., Lindstaedt, S. N., Stern, H., Friedrich, M., and Wolpers, M. (2010). Issues and considerations regarding sharable data sets for recommender systems in technology enhanced learning. *Procedia CS*, 1(2):2849–2858.

[Drachsler et al., 2008] Drachsler, H., Hummel, H. G., and Koper, R. (2008). Personal recommender systems for learners in lifelong learning networks: the requirements, techniques and model. *International Journal of Learning Technology*, 3(4):404–423.

[Duval et al., 2009] Duval, E., Vuorikari, R., and Manouselis, N. (2009). Special issue on social information retrieval for technology enhanced learning. *J. Digit. Inf.*, 10(2).

[Felfernig and Burke, 2008] Felfernig, A. and Burke, R. D. (2008). Constraint-based recommender systems: technologies and research issues. In Fensel, D. and Werthner, H., editors, *ICEC*, volume 342 of *ACM International Conference Proceeding Series*, page 3. ACM.

[Gayo et al., 2010] Gayo, J. E. L., de Pablos, P. O., and Lovelle, J. M. C. (2010). Wesonet: Applying semantic web technologies and collaborative tagging to multimedia web information systems. *Computers in Human Behavior*, 26(2):205–209.

[Ghazarian et al., 2014] Ghazarian, S., Shabib, N., and Nematbakhsh, M. A. (2014). Improving sparsity problem in group recommendation. In Cena, F., da Silva, A. S., and Trattner, C., editors, *Hypertext 2014 Extended Proceedings: Late-breaking Results, Doctoral Consortium and Workshop Proceedings of the 25th ACM Hypertext and Social Media Conference (Hypertext 2014), Santiago, Chile, September 1-4, 2014.*, volume 1210 of *CEUR Workshop Proceedings*. CEUR-WS.org.

[Goldberg et al., 2001] Goldberg, K. Y., Roeder, T., Gupta, D., and Perkins, C. (2001). Eigentaste: A constant time collaborative filtering algorithm. *Inf. Retr.*, 4(2):133–151.

[Hastie et al., 2009] Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Friedman, J., and Tibshirani, R. (2009). *The elements of statistical learning*. Springer.

[Herlocker et al., 2002] Herlocker, J., Konstan, J. A., and Riedl, J. (2002). An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms. *Inf. Retr.*, 5(4):287–310.

[Herlocker et al., 2004] Herlocker, J. L., Konstan, J. A., Terveen, L. G., and Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 22:5–53.

[Jannach et al., 2003] Jannach, D., Zanker, M., Felfernig, A., and Friedrich, G. (2003). *Recommender Systems, An Introduction*. Cambridge Press.

[Koper, 2003] Koper, R. (2003). Combining reusable learning resources and services to pedagogical purposeful units of learning. *Reusing online resources: A sustainable approach to eLearning*, pages 46–59.

[Kurilovas et al., 2014] Kurilovas, E., Serikoviene, S., and Vuorikari, R. (2014). Expert centred vs learner centred approach for evaluating quality and reusability of learning objects. *Computers in Human Behavior*, 30(0):526 – 534.

[Lam et al., 2008] Lam, X. N., Vu, T., Le, T. D., and Duong, A. D. (2008). Addressing cold-start problem in recommendation systems. In *Proceedings of the 2nd international conference on Ubiquitous information management and communication*, pages 208–211. ACM.

[Lehmann et al., 2008] Lehmann, L., Hildebrandt, T., Rensing, C., and Steinmetz, R. (2008). Capture, management, and utilization of lifecycle information for learning resources. *Learning Technologies, IEEE Transactions on*, 1(1):75–87.

[Lika et al., 2014] Lika, B., Kolomvatsos, K., and Hadjiefthymiades, S. (2014). Facing the cold start problem in recommender systems. *Expert Systems with Applications*, 41(4):2065–2073.

[Mandl et al., 2011] Mandl, M., Felfernig, A., Teppan, E., and Schubert, M. (2011). Consumer decision making in knowledge-based recommendation. *Journal of Intelligent Information Systems*, 37(1):1–22.

[Manouselis and Costopoulou, 2007] Manouselis, N. and Costopoulou, C. (2007). Experimental analysis of design choices in multiattribute utility collaborative filtering. *IJPRAI*, 21(2):311–331.

[Manouselis and Costopoulou, 2008] Manouselis, N. and Costopoulou, C. (2008). Experimental analysis of multiattribute utility collaborative filtering on a synthetic data set. *Personalization Techniques and Recommender Systems, Series in Machine Perception and Artificial Intelligence*, 70:111–134.

[Manouselis et al., 2013] Manouselis, N., Drachsler, H., Verbert, K., and Duval, E. (2013). *Recommender systems for learning*. Springer.

[Manouselis et al., 2011] Manouselis, N., Drachsler, H., Vuorikari, R., Hummel, H., and Koper, R. (2011). *Recommender systems in technology enhanced learning*, pages 387–415. Springer.

[Manouselis et al., 2010] Manouselis, N., Vuorikari, R., and Van Assche, F. (2010). Collaborative recommendation of e-learning resources: an experimental investigation. *Journal of Computer Assisted Learning*, 26(4):227–242.

[Motelet, 2007] Motelet, O. (2007). *Improving Learning-Object Metadata Usage during Lesson Authoring*. PhD thesis, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile.

[Nadolski et al., 2009] Nadolski, R., van den Berg, B., Berlanga, A. J., Drachsler, H., Hummel, H. G. K., Koper, R., and Sloep, P. B. (2009). Simulating light-weight personalised recommender systems in learning networks: A case for pedagogy-oriented and rating-based hybrid recommendation strategies. *J. Artificial Societies and Social Simulation*, 12(1).

[Nitto et al., 2006] Nitto, E. D., Mainetti, L., Monga, M., Sbattella, L., and Tedesco, R. (2006). Supporting interoperability and reusability of learning objects: The virtual campus approach. *Educational Technology & Society*, 9(2):33–50.

[Ochoa and Duval, 2009] Ochoa, X. and Duval, E. (2009). Automatic evaluation of metadata quality in digital repositories. *Int. J. on Digital Libraries*, 10(2-3):67–91.

[Ochoa et al., 2011] Ochoa, X., Klerkx, J., Vandeputte, B., and Duval, E. (2011). On the use of learning object metadata: The globe experience. In *Towards ubiquitous learning*, pages 271–284. Springer.

[Padró and Stanilovsky, 2012] Padró, L. and Stanilovsky, E. (2012). Freeling 3.0: Towards wider multilinguality. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012), Istanbul, Turkey, May 23-25, 2012*, pages 2473–2479.

[Rodrıguez et al., 2015] Rodrıguez, A. C., Gago, J. M. S., Rifón, L. E. A., and Rodrıguez, R. P. (2015). A recommender system for non-traditional educational resources: A semantic approach. *Journal of Universal Computer Science*, 21(2):306–325.

[Segaran, 2007a] Segaran, T. (2007a). *Programming collective intelligence - building smart Web 2.0 applications*. O'Reilly.

[Segaran, 2007b] Segaran, T. (2007b). *Programming Collective Intelligence: Building Smart Web 2.0 Applications*. O'Reilly Series. O'Reilly.

[Shaw et al., 2010] Shaw, G., Xu, Y., and Geva, S. (2010). Using association rules to solve the cold-start problem in recommender systems. In *Advances in Knowledge Discovery and Data Mining*, pages 340–347. Springer.

[Shepitsen et al., 2008] Shepitsen, A., Gemmell, J., Mobasher, B., and Burke, R. (2008). Personalized recommendation in social tagging systems using hierarchical clustering. In *Proceedings of the 2008 ACM conference on Recommender systems*, pages 259–266. ACM.

[Sieg et al., 2010] Sieg, A., Mobasher, B., and Burke, R. (2010). Improving the effectiveness of collaborative recommendation with ontology-based user profiles. In *proceedings of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems*, pages 39–46. ACM.

[Song and Gao, 2014] Song, B. and Gao, J. (2014). The enhancement and application of collaborative filtering in e-learning system. In *Advances in Swarm Intelligence*, pages 188–195. Springer.

[Tang et al., 2014] Tang, T. Y., Winoto, P., and McCalla, G. (2014). Further thoughts on context-aware paper recommendations for education. In *Recommender Systems for Technology Enhanced Learning*, pages 159–173. Springer.

[Taraghi et al., 2015] Taraghi, B., Saranti, A., Ebner, M., Müller, V., and Großmann, A. (2015). Towards a learning-aware application guided by hierarchical classification of learner profiles. *Journal of Universal Computer Science*, 21(1):93–109.

[Tiropanis et al., 2009] Tiropanis, T., Davis, H. C., Millard, D. E., and Weal, M. J. (2009). Semantic technologies for learning and teaching in the web 2.0 era. *IEEE Intelligent Systems*, 24(6):49–53.

[Verbert et al., 2011] Verbert, K., Drachsler, H., Manouselis, N., Wolpers, M., Vuorikari, R., and Duval, E. (2011). Dataset-driven research for improving recommender systems for learning. In Long, P., Siemens, G., Conole, G., and Gasevic, D., editors, *LAK*, pages 44–53. ACM.

[Verbert et al., 2012] Verbert, K., Manouselis, N., Ochoa, X., Wolpers, M., Drachsler, H., Bosnic, I., and Duval, E. (2012). Context-aware recommender systems for learning: a survey and future challenges. *Learning Technologies, IEEE Transactions on*, 5(4):318–335.

[Zhang et al., 2012] Zhang, X., Ordonez de Pablos, P., and Zhang, Y. (2012). The relationship between incentives, explicit and tacit knowledge contribution in online engineering education project. *The Knowledge Society Selected Papers on Engineering Education from the 4th World Summit on the Knowledge Society.*

[Zhao et al., 2015] Zhao, J.-c., Liu, S.-h., and Zhang, J. (2015). Research on personalized recommendation system on item-based collaborative filtering algorithm. In *Proceedings of the International Conference on Advances in Mechanical Engineering and Industrial Informatics (AMEII 2015), Zhengzhou, China, Apr 11-12, 2015.*