A user-profile semantic modelling methodology for the composition of training groups

Laurie Acensio
Laurie.acensio@lexiane.fr

The societal context of LLP\(^1\) encourages the adult-learner to engage in learning programmes throughout his professional career. In continuing education, the adult-learner engagement is mainly focused on the need for highly individualized and contextualized training and this must be reconciled with the classroom training group situation. The adult-learners are included in a random manner which may hinder cohesion and group dynamics. Although the training group is short-lived, peer interaction remains an essential component stimulating the learning process. In this paper, we propose a methodology to facilitate the creation of homogeneous training groups. Our approach is to represent and enrich the initial adult-learner profile formalized by an ontology. For the purposes of this analysis, grouping criteria must first be established.

**Keywords.** Group Formation, User profile, Ontology, Social and semantic similarity

### 1 Introduction

Continuing education implies a hybrid approach (blended-learning) in order to articulate the training according to the mode, face-to-face or distance. In this context, the IT environment helps to regulate and mediate, both socially and intellectually.

Collaborative activity in a learning context has been widely theorized in the scientific literature, in particular offering communication tools to promote synchronous and/or asynchronous interaction between learners. However, the strengthening of face-to-face peer interactions remains a preferred method for maintaining the adult-learner's commitment to the learning process. According to [1], the essential question is no longer whether peer interactions *generally* are or are not conducive to learning but *under what conditions* these interactions occur. The quality of the collaborative process between learners can depend on the composition of the training group. Therefore, the theme is closely linked to the problem of the heterogeneity of participant profiles. Currently within our training organization, the composition of the training groups is mainly mode makes opportunistically according to a general and limited training offer by the trainers and

\(^{1}\) Lifelong Learning Programme
allowance for certain adult-learners' preferences (e.g. availability, budget, geographical proximity ...) to facilitate synchronous meetings.

Our research work consists in optimizing the organization of training groups composed of adult-learners whose learning objectives only partly converge. Indeed, the main reason adult-learners remain committed to the training process is to fulfil their specific learning objectives. Individually, the need for training is highly individualized and contextualized according to the specific problems of the adult-learner.

Moreover, the training needs can be formed imprecisely during the initial request and evolved throughout the training course. Our experience shows that the need for training arises from an individual cognitive impulse which is then formed gradually through feedback from others: the trainer, peers and the training organization. The training organization is faced with the difficulty of adapting the training need of each learner related to his personal situation in a training group. Therefore, we can summarize the problem as follows:

**How compose homogenous training groups from the heterogeneity learners profiles?**

Related research questions are:

**Question 1 : how to model the scalability of the adult-learner profile?**

Indeed, one of the main limitations is for enriching “dynamic” characteristics (indications of activities) in the user profile. The main difficult is to identify the thematic progression of indications of activities taking into account his temporal context.

**Question 2 : how to calculate the similarity between different actors of the learning situation (learners profiles/trainers) from the adult-learner profile?**

This question focuses on extracting of the common characteristics between learners profiles in the automation process. Our main hypothesis states that the homogeneity of the group combining on the similar socio-semantic criteria, defined in advance by a human. We assume that the need for training and the adult-learner's profession are key factors in ensuring homogeneity within the training group. The profession is a relatively stable indicator whereas the need for training reflects a changing body of knowledge and by nature imprecise.

## 2 State-of-the-art approach

Here we present an overview of existing approaches for automatic composition of training groups and introduce the concept of learner profiles.

### 2.1 Automatic group composition methods

In socio-cognitive psychology research, the model of [2], offers four methods of constituting groups: random, autonomous, opportunist and rational. The “random” mode is based on chance combinations. The “autonomous” mode leaves the choice to the participants based on individual preferences while the “opportunist” mode makes allowance for certain learners' preferences. We are particularly interested in the last
category, the one called “rational”, which aims to encourage group dynamics involving learners based criteria previously defined.

The majority of scientific work uses classification algorithms for automatic composition of training groups like [3, 4]. The state-of-the-art [5] computing techniques implemented and used are primarily probabilistic algorithms (e.g. genetic algorithms), followed by data-mining techniques (k-means). However, the author notes a lack of source code for the reuse of these algorithms and therefore the community has no means to compare and evaluate different approaches to forming training groups.

Some studies use an approach based on an ontology [6, 7] advocating a hybrid approach for better results. Following [8], the use of a semantic ontology-based model can help to better understand users’ interests in relation to the vector model. This method allows managing terminology ambiguities by refining the concepts.

To our knowledge, the experiments were mainly carried out in a context of initial training, so the freedom of participation of the learner profile within the group is less significant.

2.2 Learners profiles

While the methods differ, we can highlight the common factor in all work: calculation of the similarity is based primarily on explicit features of the learner profile (e.g. learning styles, preferences, interests). In the computing environments for Human Learning research, the modelling of the user profile is a method that has been widely used in several research areas mainly from the perspective of personalization and adaptation. Out of many variants of definitions will be used the one, formulated by [9] in which as information concerning a learner or a group of learners, collected or deduced from one or several pedagogical activities, computerized or not. Information contained in the learner profile can concern his knowledge, abilities, conceptions or his behaviour. From the perspective of composing a group training, the criteria based on the level of knowledge and skills are prioritized according to [10] while the learning style, personality and gender are also criteria that can be taken into account in heterogeneity [11]. Regarding interpersonal relationships, the work of [12, 13] show that the interactions are more effective for learning when they are part of a symmetrical relationship between partners. More recently, [14] advance the argument that social networks are built around people who are united by thematic content they share. Similarities between peoples’ speech gives rise to networks called socio-semantic networks and these can be established from written or oral records of the topics (...) discussed.

This state-of-the-art shows that mostly of the studies explore group formation strategies relying on learners profile. This position leads us to our assumption that the selection of beforehand predefined characteristics on learners profil influence collaboration in a training group.

Subsequently, we present our approach based on semantic modelling of adult-learner profiles.
3 Proposal

Our proposal is to calculate proximity using ontological modeling of adult-learner profile to extract a rational composition of a training group.

![Simplified conceptual schema](image)

In order to perform group composition, we use the following socio-semantic characteristics learner’s profil: we pay special attention to the adult-learner's profession which we believe is a relevant indicator of the associated level of expertise, as well as his areas of interest. Similarly, the symmetry of the socio-professional relationship characterized by the hierarchy level is also a key criterion to ensure bilateral cohesion in peer interactions.

The aim is to define an adult-learner model which takes into account the so-called static characteristics and the so-called dynamic characteristics, being the changing nature of the learner profile's interests over time.

3.1 Methodology

We defined three steps for create a group formation: our approach is to (1) build the initial profile of the adult-learner via an ontology (2) enhance the user profile from his interactions and (3) define the rules for inserting conditions.

* Representation of the user profile

The adult-learner profile includes the explicit personal and professional information entered by the adult-learner. The semantic profile of the user profile is represented as a set of concepts within an ontology that we have qualified as "Named entities". The ontology extends the FOAF\(^2\) ontology and describes people for building formal and informal training group. The main advantage of this approach is a structuring via a subsumption hierarchy according to several levels of generality / specificity.

A semantic extension of this ontology is provided by an RTO that aims to determine the terminology variations of domain knowledge, in particular the names of professions and training interests.

\(^2\) Friend Of A Friend
• **Enhancement of the user profile**

The user profile is progressively enhanced and enriched from the traces of social interactions and activities with the system.

The main idea is to supplement the interests provided by the adult-learner explicitly. Our analysis is limited to a thematic extraction of synchronous and asynchronous social interactions. This is to detect language descriptors (or tokens) via manual annotation. The selected annotations are stored as instances and subsequently categorized into previously-defined concepts of the ontology. The main limitations are linked to the relevance of text content stored via annotations. A decision based on annotated fragments is a subjective process originating from a compromise in the relevance of the annotation.

In addition, in our case, training interests change over time and are inherently uncertain. To refine thesis semantic annotations it is necessary to associate a time constraint to analyse the difference between an initial topic and a future topic into which it may have drifted. Similarly, the accuracy of the values should be taken into consideration. We incorporate an extra layer of annotations called "fuzzy" which allows us to associate each value with a level of confidence we attribute to this value.

• **Reasoning**

The aim is to produce a rationale for a given adult-learner profile previously conceptualized, in order to detect associations. We are considering the use of a reasoning layer called abductive as used by the rule language SWRL. Based on observations, this reasoning attempts to produce explanations that allow us to extract instances.

Regarding the explicit characteristics of the adult-learner, domain experts have defined the following conditions:

**Condition 1**: the group must have at least 4 participants

**Condition 2**: all participants must belong to the same group of professions

**Condition 3**: all participants must have the same hierarchy level

**Condition 4**: all participants must work in the same field of activity

A first test was conducted using methods of direct and declarative handling using query language (SPARQL). Descriptive logic such as "the condition for the minimum number within a group" can be easily programmed using cardinality conditions and therefore expressed as a SPARQL query. However, the methods have limitations in that they cannot infer knowledge previously defined by the ontology. In addition, rule definition is a manual task which could incorporate previously known parallels, at the same time not detecting new cases of potential and unknown associations.

To overcome the limitations of SWRL rules, we plan to incorporate a semantic similarity computation between concepts. The typical approach is to use a measure that is calculated based on the length of the path by considering the concepts as vertices and edges as semantic relationships. The major disadvantage of this measure lies in the strong assumption that all the ontology subsumption relations have the same

---

3 Semantic Web Rule Language

4 Syntax Protocol and RDF Query Language
weight in calculating similarity. This limitation can be managed by using a measure that associates a numeric value with the degree of similarity between two essential concepts.

4 Conclusion and perspectives

In continuing education, uniform composition of adult-learners within the training group will help create favourable conditions for peer collaboration.

This paper presents a methodology to model the adult-learner profile with the aim of composing training groups based on reasoning rather than chance. Our approach is to adopt a semantic method via an ontology which combines the adult-learner profile guided mainly by the profession with a semantic analysis of the social interactions underlying collaboration.

The main advantage of ontological modeling is the treatment of semantic and heterogeneous data. From the perspective of to compose a training group, the optimization depends on the distance between all group members. However, if we respect the principles of construction of ontology, the measures are calculated according to the length of the path by considering the concepts as vertices and edges as semantic relations. The main disadvantage of this measure lies in the strong hypothesis that all the subsumption relations of the ontology involved in the same proportion during the calculation of similarity.

The main perspective of our work will be to test a similarity computation based on conceptual distances and comparing the results in order to determine the optimal method. Our experimental project is to confirm whether a small, homogeneous training group may affect the observable interactions between peers. Our next step is to analyse the performance of the group according to different methods of selection (random, self-selection and reasoned selection) in order to confirm whether or not our hypothesis of research.

References


