Who Knows/Uses What of the UML: A Personal Opinion Survey

Gianna Reggio, Maurizio Leotta, Filippo Ricca

Abstract:

UML is a comprehensive notation, offering a very large set of diagrams and constructs covering any possible modelling need. As consequence, on one hand, it is difficult and time consuming to master it, and on the other hand, people tend, naturally, to consider only a part of it. In practice, many UML diagrams/constructs seem scarcely used or even their existence is not known. By means of a study covering any possible source of information (e.g. UML books and tools), we started to assess which part of the UML is considered and used in practice. Here, we present some results about knowledge and usage of the UML diagrams by means of a personal opinion survey with 275 participants from both industry and academy, analysing also the influence of different factors: working environment (academia vs. industry), working role, seniority, education, and gender.

Digital Object Identifier (DOI):

http://dx.doi.org/10.1007/978-3-319-11653-2_10

Who Knows/Uses What of the UML: A Personal Opinion Survey

Gianna Reggio, Maurizio Leotta, Filippo Ricca

DIBRIS, Università di Genova, Italy
{gianna.reggio, maurizio.leotta, filippo.ricca}@unige.it

Abstract. UML is a comprehensive notation, offering a very large set of diagrams and constructs covering any possible modelling need. As consequence, on one hand, it is difficult and time consuming to master it, and on the other hand, people tend, naturally, to consider only a part of it. In practice, many UML diagrams/constructs seem scarcely used or even their existence is not known. By means of a study covering any possible source of information (e.g. UML books and tools), we started to assess which part of the UML is considered and used in practice. Here, we present some results about knowledge and usage of the UML diagrams by means of a personal opinion survey with 275 participants from both industry and academy, analysing also the influence of different factors: working environment (academia vs. industry), working role, seniority, education, and gender.

Keywords: UML Usage and Knowledge, Personal Opinion Survey, Empirical Study

1 Introduction

UML is a comprehensive notation grown up during the years; it is general purpose and so it has a very large number of constructs to cover any possible modelling need in any possible context. Here we report some data to illustrate the size of UML: (1) the specification document for version 2.4.1 (superstructure only) [22] has 732 pages, (2) there are 14 different types of diagrams (the latest, profile diagram, has been added in version 2.1), and (3), each diagram offers a large number of constructs (e.g. 47 for the activity diagram).

Being a so big notation has some drawbacks hindering its acceptance, its adoption and its usage; see, e.g. as this feeling is quite strongly expressed on the SEI (Software Engineering Institute) Architecture Technology User Network blog[1] "*UML is too complex. UML has increased in complexity and size over the years. Today there are 14 different types of diagrams! That's too much for a human being to grasp. Some people shy away from UML because they judge that the effort to climb the learning curve will not pay off*". Trivially, printing the UML specification requires a large number of paper sheets, but it also requires many hours to read and understand it, not to mention maintaining and evolving it by the designers of the notation, and of course learning and teaching the whole UML is a difficult task [25].

[1] http://saturnnetwork.wordpress.com/2010/10/22/five-reasons-developers-dont-use-uml-and-six-reasons-to-use-it/

It is worthwhile noticing that this feeling has been also acknowledged by the OMG[2], indeed the next version (2.5) waiting for the formal release is the result of the "UML Simplification" initiative [19], but in this case the simplification will cover only the way UML is defined without any impact on its constructs; nevertheless as stated by Steve Cook in his blog[3] "*the work done in 2.5 provides an excellent foundation for future simplifications and improvements*".

However, to the best of our knowledge, only few studies have addressed the extent to which UML diagrams are known and used in practice [2, 4, 8, 18], and neither examined why professionals choose to use some diagrams and ignore others. Our research seeks to address the first issue by surveying the UML usage in practice.

UML is so huge that it is not possible to investigate the knowledge and usage level of the whole notation in "only one shot". So we decided to start considering only the UML diagrams and the constructs of use case and activity diagrams.

We have decided to investigate all the possible sources providing information on the knowledge/usage of the UML. Precisely, the considered sources are: books (guides about the UML and books using the UML as a notation, i.e. books where UML is not the primary subject), university courses, tutorials, tools, people (academics and professionals), web resources and model repositories, e.g. ReMoDD[4]. We think to have taken into account all sensible, accessible, and non-biased sources; for example research papers on the UML are not a good choice, since they usually cover the newest and most problematic features; whole models produced in industry are wonderful, but it is extremely difficult to get a sensible number of them, and in many cases it is impossible to examine them, since they have been produced with tools not available to the authors.

We have already published the results of our first survey concerning static sources (only UML books, courses, tutorials and tools) [14–16]. Here, we present the results of a novel personal opinion survey with 275 respondents from both academia and industry aimed to understand: (1) which UML diagrams are known/used, and (2) whether the working environment (academia versus industry), the working role, the UML seniority (years of UML knowledge), the UML education (how the UML was learned), and the gender influence the results. Then, we briefly compare the personal opinion survey results with those of the previous survey on the static sources. The results of our investigations will be useful to anyone that for various reasons has to consider only a part of the UML, for example teachers and instructors with a fixed number of teaching hours, people engaged in self learning and developers of model transformations interested to understand which diagrams/constructs are neglected and why.

We present in Section 2 the design and the procedure of the conducted personal opinion survey, and in Section 3 some results and possible threats to validity. Related works are in Section 4, followed by conclusions and future works in Section 5.

2 Study Definition, Design, and Procedure

The main aim of this work is to collect information on the knowledge and usage of the UML in the industrial and academic reality. In particular, the survey focuses on UML

[2] http://www.omg.org

[3] http://searchsoa.techtarget.com/feature/Steve-Cook-on-what-architects-can-expect-from-UML-25-revision

[4] http://www.cs.colostate.edu/remodd/v1/

diagrams and on the constructs of the use case and activity diagrams. In this paper, for space reason, we focus only on the results about the UML diagrams. To implement this survey, we: (*i*) used the same framework of [20, 21] (based on [6]), (*ii*) followed as much as possible the suggestions given in [7], and (*iii*) adopted an on-line questionnaire to collect information.

The **goal** of the survey is *taking a snapshot of UML knowledge and usage in the industry and academy* taking the perspective of:

– **teachers and instructors**: allowing to offer courses and/or tutorials concentrating on a smaller language made out of the most used UML diagrams/constructs;
– **tool builders/users**: interested to focus the tools on the most used UML diagrams/ constructs, since the tools covering a smaller number of diagrams/constructs will be simpler to implement and to use;
– **notation designers**: interested in discovering scarcely used constructs, and understanding for which reasons they have been added to the UML. Moreover, other interesting questions could arise: are the scarcely used constructs derived[5] or primitive? Can the scarcely used constructs be applied only in specific cases? It will be interesting to investigate whether the metamodel (and subsequently the UML specification) may be easily simplified, without losing expressive power, to cover only the most used parts.

The context consists of a sample of 275 professionals and academics having at least a basic UML knowledge.

Research Questions Given the above goal, the survey aimed at addressing the following *research questions*:

RQ1: *Which of the UML diagrams are the most/less known?*
RQ2: *Which of the UML diagrams are the most/less used in practice?*
RQ3: *Have Professionals and Academics the same level of knowledge and use of UML?*
RQ4: *Have the working role, UML seniority, UML education type, and gender any kind of influence on the level of knowledge and use of the UML?*

Target Population and Sample Identification The target population is the set of individuals to whom the survey applies. In our case the population consists of professionals and academics having at least a basic UML knowledge. Our sample consists of university professors, researchers, students (mainly PhD), and professionals (mainly, project managers, business analysts, software architects, designers, and developers) who work (1) in companies of the IT field; their skills are related to the production, maintenance or management of software systems (the larger part); (2) in companies that do not directly belong to the IT field (a smaller part) but that use information systems to carry out and support the company's business activities; (3) for public agencies, government enterprises; the remainder of the sample perform other kinds of activities.

The sample was obtained in two ways: (1) by convenience, i.e. relying on the network contacts of our research group, and (2) by sending invitation messages on mailing lists and Web groups concerning UML, MD and software engineering. In particular, we have used some lists available at the university (such as former students or people who have participated in previous surveys about other topics) and international professional groups (e.g. LinkedIn). We opted for non-probabilistic sampling methods even if we know all

[5] A derived construct may be replaced by a combination of other constructs.

the problems of this sampling (e.g. the risk of using a sample not representative of the target population) [7] because this survey is exploratory and because we thought that the target population was hard to identify and of limited availability.

In total, we received 275 complete responses to our survey. Unfortunately, we do not know exactly how many people have been reached by our invitation messages and advertisements, so we cannot calculate the response rate. The same problem is present also in other software engineering surveys (e.g. in [10]).

Data Collection Data were collected through the creation of an on line questionnaire. The use of a web-based tool simplifies and speeds up the completion of the questionnaire with clear advantages in terms of the number of responses obtained [6]. The online questionnaire has been developed and published using LimeSurvey[6].

Questionnaire Design The questionnaire is organized into four sections each of them implemented as a Web page. The first section contains the questions designed to get information about the survey participants, while the 2nd, 3rd, and 4th sections contain a series of questions designed to assess the knowledge and usage of the UML, respectively focusing on UML diagrams, use case diagram constructs and activity diagram constructs. This division was necessary to create different paths to complete the questionnaire depending on the type of response given by the participants (e.g. the activity diagram section is skipped if a participant states that (s)he does not know that kind of diagram). Thus, the total number of questions to compile is variable depending on the responses given to previous questions. A partial list of questions (i.e. the "Personal Data" and "UML Diagrams" sections) is shown in Table 1. Some of them are not mandatory (e.g. questions 1.1 and 1.2).

We have decided to formulate the questions about the knowledge and usage of a diagram/construct in terms of "having seen a model containing it" and "having produced a model containing it", to avoid misunderstandings or personal interpretations by the respondents. For example "I know diagram/construct X" may be interpreted as "I know its existence" or "I know its syntax and semantics"; similarly, instead of "I have used diagram/construct X" we have preferred the most precise "I have produced a model containing diagram/construct X". Furthermore, to avoid that someone that knows or has used a diagram/construct but perhaps (s)he is not able to remember its precise name (e.g. composite structure diagram used to represent structured classes and collaborations) we have added an example to the question (see, e.g. Fig. 1).

We have chosen to force the participants to answer all the questions in each section of the questionnaire using a special option provided by LimeSurvey; in this way it was possible to get all the questionnaires filled out correctly. The only optional questions concern full name and email (i.e. questions 1.1 and 1.2); so we allowed the participants to complete the questionnaire anonymously (some studies have shown that the response rate is affected by the anonymity policy of a study [24]).

To harvest more answers, we decided that the questionnaire should take no longer than approximately 15 minutes to complete (long questionnaires get fewer answers than short questionnaires [24]) and we designed it accordingly.

[6] http://www.limesurvey.org

ID	Conditional	Question
1.1		What is your full name? *(optional)*
1.2		Please provide a valid personal email address. *(optional)*
1.3		What is your gender?
1.4		How old are you?
1.5		What is your nationality?
1.6		What is your current occupation?
1.7	Yes[a]	What legal entity does your company fit?
1.8	Yes[a]	What is the main business activity of your company?
1.9	Yes[a]	Are modelling and/or MD* techniques used in your company?
1.10		Are you currently using UML for your (business) activity?
1.11		How did you learned UML?
1.12		How many years have you been using UML?
1.13	Yes[a]	Do you use software tools to produce UML models?
1.14	Yes[a]	List all the software modelling tools you have used.
1.15		Have you read the OMG UML specifications?
1.16		Are you active in OMG?
1.17		Did you take part in OMG UML definition?
1.18		Which one of the following OMG UML certifications do you have?
2.1		Class Diagram[b]
2.2		Object Diagram[b]
2.3		Activity Diagram[b]
2.4		Use Case Diagram[b]
2.5		State Machine[b]
2.6		Sequence Diagram[b]
2.7		Communication Diagram[b]
2.8		Component Diagram[b]
2.9		Composite Structure Diagram[b]
2.10		Deployment Diagram[b]
2.11		Package Diagram[b]
2.12		Timing Diagram[b]
2.13		Interaction Overview Diagram[b]
2.14		Profile Diagram[b]

[a] Shown only if a condition on the previous answers is satisfied

[b] *I have never seen it*, *I examined a model containing it*, *I produced a model containing it*
For each question in this group an image depicting a sample diagram is shown (see Fig. 1)

Table 1. Questionnaire (only "Personal Data" and "UML Diagrams" sections reported).

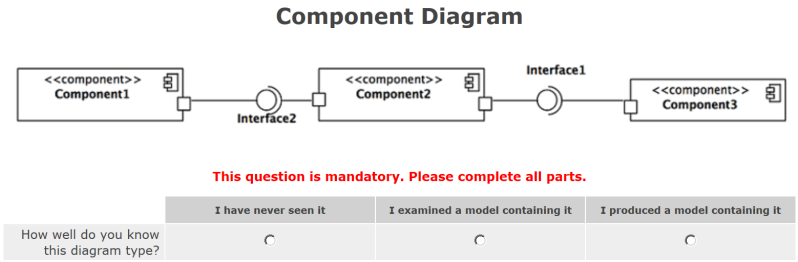

Fig. 1. A question (Q 2.8) from the "UML Diagrams" questionnaire section

The questionnaire was introduced with a brief statement about the purpose of our research (as suggested in [7]), and we added a sentence to clarify that all the collected information had to be considered highly confidential[7].

Survey Execution The data presented in this paper have been collected since 1st of October 2013 until the 10th of March 2014 (about 5 month and half). Approximately

[7] In conformity with privacy Italian law: "D.lgs. n. 196/2003".

every month, we sent to different mailing lists and groups an invitation to participate to the survey. The procedure followed to prepare, administer, and collect the questionnaire data is made up of the following five main steps:

1. *Preparation and Design of the Questionnaire*. Starting from similar questionnaires [20, 21] and tailoring them to our objectives, an initial set of questions was agreed among us.
2. *Pilot Study*. A pilot study was performed before executing the survey (i) to tune the questionnaire and (ii) to reduce the ambiguities contained in the questions. An industrial IT professional and a university professor filled a preliminary version of the questionnaire and provided their judgment on it. Following the suggestions of the two contacted experts, minor changes to the questionnaire were made. After this pilot study we concluded that the survey was well suited for IT professionals and academics and that the questions were clear enough.
3. *On-line Deployment*. Once the questionnaire was refined after the pilot study, it was deployed on-line by using LimeSurvey as explained before.
4. *Monitoring*. During the data capture phase, our research group monitored the progress of the questionnaire submission. A few people reporting difficulties about the questions asked us for clarifications.
5. *Data Analysis*. After questionnaires have been collected, simple analyses were performed with the aim of answering the research questions. Given the nature of this survey, that is mainly descriptive (it describes some conditions or factors found in a population in terms of its frequency and impact [7]), we applied quite exclusively descriptive statistics and showed our findings by means of charts. For space reasons, we employed line charts instead of the more common/adapt column charts. Anonymized raw data are available at: http://sepl.dibris.unige.it/2014-UMLPersonalSurvey.php

3 Results

We first present some information about the background of the respondents, and then some results from the execution of the survey.

We assume that a diagram is known iff the respondent did not answered "I have never seen it" to the corresponding question from 2.1 to 2.14 in the questionnaire (see Table 1), and thus if (s)he has either just examined or produced a model containing it.

3.1 Respondents' Background

From the answers to the first section of the questionnaire we have found that: (Q1.3) the majority (83%) of the respondents are male, and (Q1.4) the most common age groups are 28-37 and 38-47 with respectively the 43% and 30% of the respondents.

We have respondents from many different nationalities (Q1.5, see Fig. 2) with 231 participants from Europe (mainly from Italy with about the half of the survey's respondents, and then Germany and France), 32 from Americas (mainly from US, Brazil and Canada), 10 from Asia, and 2 from Africa. The majority of the respondents are professionals (60%) while the remaining are from the academia (40%). For what

Fig. 2. Nationalities of the Respondents (dark grey means that in our sample we have at least one respondent for that Nation) – (Q1.5)

concerns the role of the respondents (Q1.6), the most frequent are software developer, professor, PhD student, and university researcher, see Fig. 3.

MD*[8] techniques (Q1.9) are used in the company of the 61% of the respondents, while the 54% of the respondents is using UML in their current activities (Q1.10). The majority of the respondents learned UML in university courses 50% or by self-study 38% (Q1.11). The 39% of the respondents use UML by less than 5 years, 35% by 5-10 years, 17% by 10-15 years and the remaining (9%) by more than 15 years (Q1.12). The 80% of the respondents create UML models using specific tools (Q1.13). The 49% of the respondents have read the OMG UML specifications (Q1.15). Only 2 respondents are active in the OMG, and only 7 took part in the OMG UML definition. Finally, only the 6% of the respondents have an OMG UML certification (Q1.18).

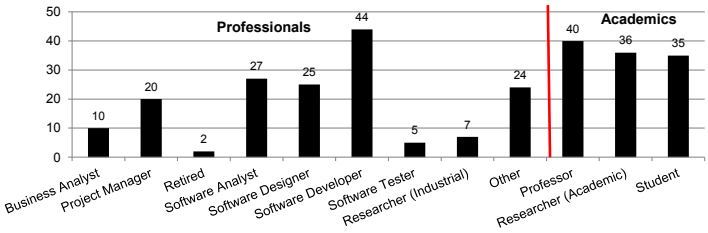

Fig. 3. Current Roles and number of the Respondents – (Q1.6)

3.2 RQ1: UML Diagrams Knowledge

The results about the knowledge of the UML diagrams of all the respondents can be seen in Fig. 4 (straight dark line). The chart shows that the level of knowledge of the various diagrams is quite different, and that we can distribute the diagrams in three main groups. **(K1)** They are diagrams that are without any doubt *widely known*, precisely use case diagram (97%), class diagram, state machine and sequence diagram (96%), and activity

[8] A common name for several model driven approaches and methodologies, such as MDD, MDSD, MDE, MDA [23].

diagram (92%). The most known one is the use case diagram, and this it is not surprising, since this diagram may be used without any other part of the UML, and it is truly useful to complement classical textual use case based requirements specifications, offering a nice way to visually summarize use cases, actors and relationships among them. Also the authors have proposed a method for building "disciplined use case specifications" where only this part of the UML is used [17]. All the diagram types in this group were already present in UML version 1.

(K2) Other diagrams are *known* with good percentage: package diagram (86%), component diagram (82%), object diagram (81%), deployment diagram (77%), and communication diagram (73%).

(K3) The remaining diagrams are *scarcely known*: composite structure diagram (59%), profile diagram (52%), interaction overview diagram (51%), and timing diagram (42%).

The answer to **RQ1** is then that some UML diagrams are very widely known **(K1)**, others are known **(K2)**, and the remaining ones are scarcely known **(K3)**. The least known is the timing diagram.

In Fig. 4 we have also presented the combined results of the previous survey [14] (dotted line) covering static sources about the UML (books, tools, university courses and tutorials). In this case the percentage is relative to the number of sources in which the various diagram types were considered. The lines in Fig. 4 are slight different but their trend is very similar (the Pearson's correlation coefficient between the two polylines is really high, 0.92), thus the most/less known diagrams are more or less the same, and so we have a confirmation of the soundness of the above answer to **RQ1**. The only real visual difference between the two polylines concerns the profile diagram, where people knowledge seems better than its consideration by the static sources. However, it is the newest diagram (indeed it was introduced only in UML version 2.1), and it is sensible to assume that books/tools/courses take more time to integrate the novelties than people. A possible confounding factor about this comparison could be that our previous survey on static sources [14] has been performed about one year ago.

Finally, we have partitioned the knowledge level of the UML diagrams considering question Q1.10 (Are you currently using UML for your business activity?). Fig. 5 clearly

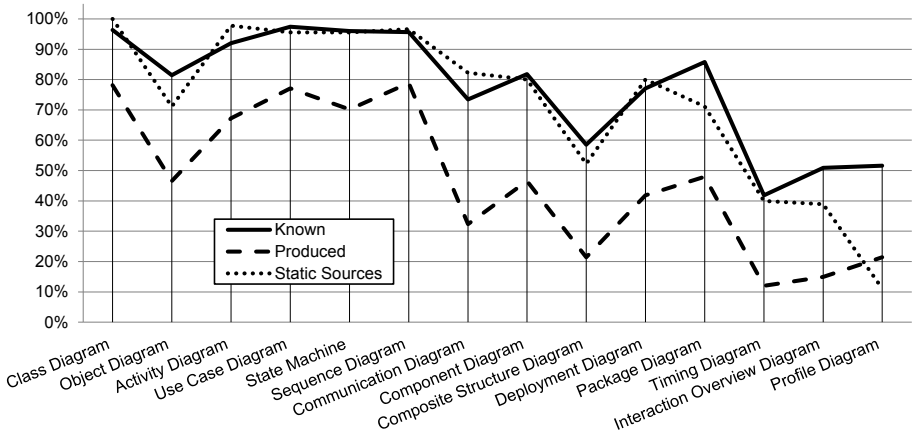

Fig. 4. UML diagrams knowledge and usage whole survey population

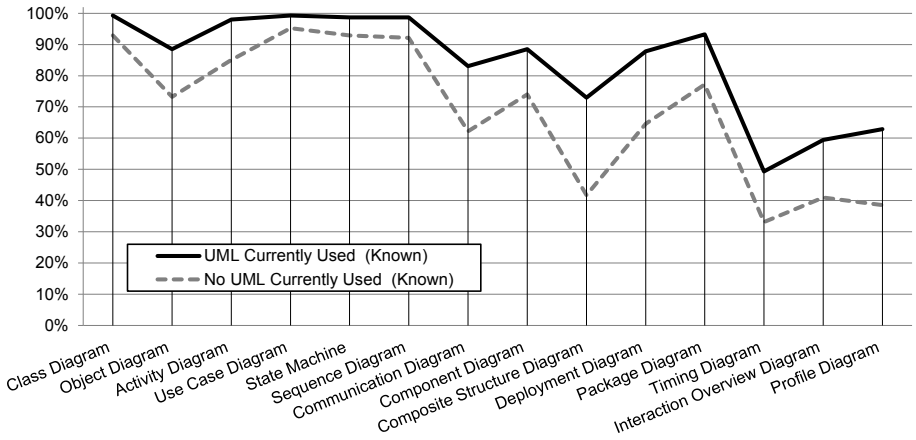

Fig. 5. UML diagrams knowledge: people currently/not currently using UML for their job

shows that the current usage of the UML in the respondents' job activities is correlated with a higher level of UML knowledge. The difference about the two distributions (used/not used) is clear-cut for all the UML diagrams with the exception of the four most known diagrams where it is really small.

It is interesting to note that: (1) using currently UML implies a better knowledge of all the diagrams and not of only a subset of them, and (2) with the reduction of the UML diagrams knowledge increases the difference between the two distributions, in other words, diagrams in group K1 are known by everyone while specific diagrams (groups K2 and K3) are more known by those who are currently using UML in the everyday activities.

3.3 RQ2: UML Diagrams Usage

The data about the real usage of the various UML diagrams are also shown in Fig. 4 (dashed line). We can see that obviously the usage figures are lower than those referring to the knowledge of the various diagrams (on average of about 30%), but the distance is greater for the less known diagrams (e.g. for the composite structure diagram). This fact may be interpreted as either less known are less used because they are less "useful" or they are less used because people do not know them. However, they follow the same trend: most/less known diagrams are also the most/less used ones. This is confirmed by the Pearson's correlation coefficient between the two polylines that is really high (0.96). Thus we can again distribute the diagrams in three groups:

(U1) The diagrams that are without any doubt *widely used* are: sequence diagram (79%), class diagram (78%), use case diagram (77%), state machine diagram (70%), and activity diagram (67%). Here we can observe that some people have produced a sequence diagram without a supporting class diagram (thus they have made an inconsistent UML model, since the classes typing the lifelines have not been defined).

(U2) Other diagrams are *used* but with lower percentage: package diagram (48%), component diagram and object diagram (47%), deployment (42%), and communication diagram (32%).

(U3) The remaining diagrams can be considered *scarcely used*: composite structure and profile diagram (21%), interaction overview diagram (15%), and timing diagram (12%).

The answer to **RQ2** is then quite similar to that to **RQ1** (the correlation between the two RQs is really strong): some UML diagrams are widely used **(U1)**, other are used **(U2)**, and the remaining **(U3)** are scarcely used; and also in this case the least used one is the timing diagram. Future empirical work is needed to understand in which niches of the industry the diagrams in **(U3)** are really used, and whether they could be also useful elsewhere.

3.4 RQ3: Professionals vs. Academics

Fig. 6 presents the data about knowledge and usage of the UML partitioning the sample by category of occupation: professionals vs. academics. For what concerns the knowledge level, academics (straight dark line) seems to know a little more than professionals (dashed light line) — average less than 7%. Only in two cases professionals seem to know a slightly more than academics: for composite structure diagram (+6%), and timing diagram (+10%). Probably, this is due to a specific use of these two types of diagrams in industrial niches (e.g. timing diagrams are heavily used in avionics industries to model guidance and control systems).

The data about the usage of the various UML diagrams are also shown in Fig. 6 (see the dashed polylines). We can see that obviously the usage figures are lower than those referring to the knowledge of the various diagrams, but the distance between industry and academic usage — on average of about 10% — is greater in some cases (e.g. for class, object, state machine, sequence and package diagrams). We believe that this is due to the different nature of the produced models containing the considered diagram type: relative to a real project for professionals and to a toy model used for teaching in the case of academics.

Thus, we can answer to question **RQ3** simply saying that UML is slightly more known and used by the academics than by the professionals except for composite structure diagrams and timing diagrams.

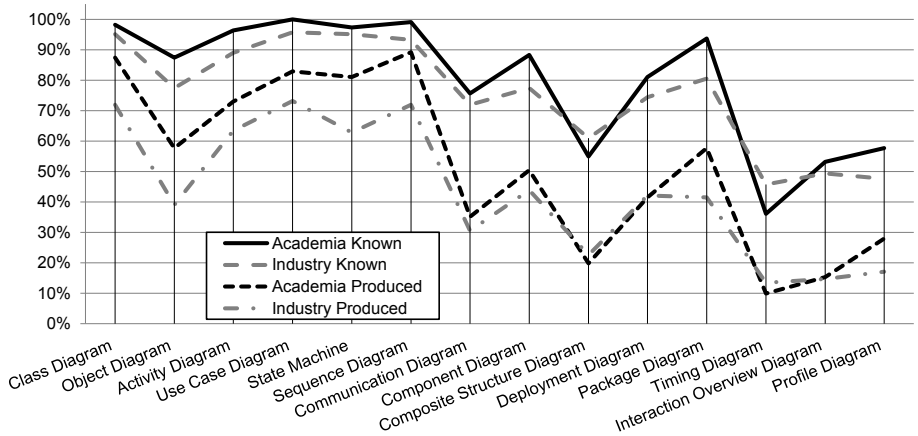

Fig. 6. UML diagrams knowledge and usage: Professionals versus Academics

3.5 RQ4: Working Role, Seniority, Education, and Gender

Fig. 7 (upper part) shows the data about the knowledge of the UML diagrams distinguishing the different working roles. To keep the charts readable we do not report the data about the roles with a very low number of respondents (e.g. software tester, see Fig. 3).

In the case of the professionals, we can see that there is a visual difference between the developer and the other roles in the majority of the diagrams; it seems that developers are those with the less knowledge of the UML diagrams. These results seem to suggest that UML is less relevant for the last software development phases. For the academics, we can see — always in Fig. 7 (lower part) — that, as expected, students (86% of them are PhD) have slightly lower knowledge of the UML with respect to the other academic roles. This is more marked for object and timing diagrams.

For what concerns the UML seniority (i.e. partitioning for question Q1.12, how many years UML has been used?), we can see in Fig. 8, that the range "less than 5 years" presents the lower level of knowledge, the range "between 5 and 10 years" shows a slight better knowledge, whereas people in the ranges "between 11 and 15 years" and "more than 15" have both a higher level of knowledge, in particular concerning the less known diagrams.

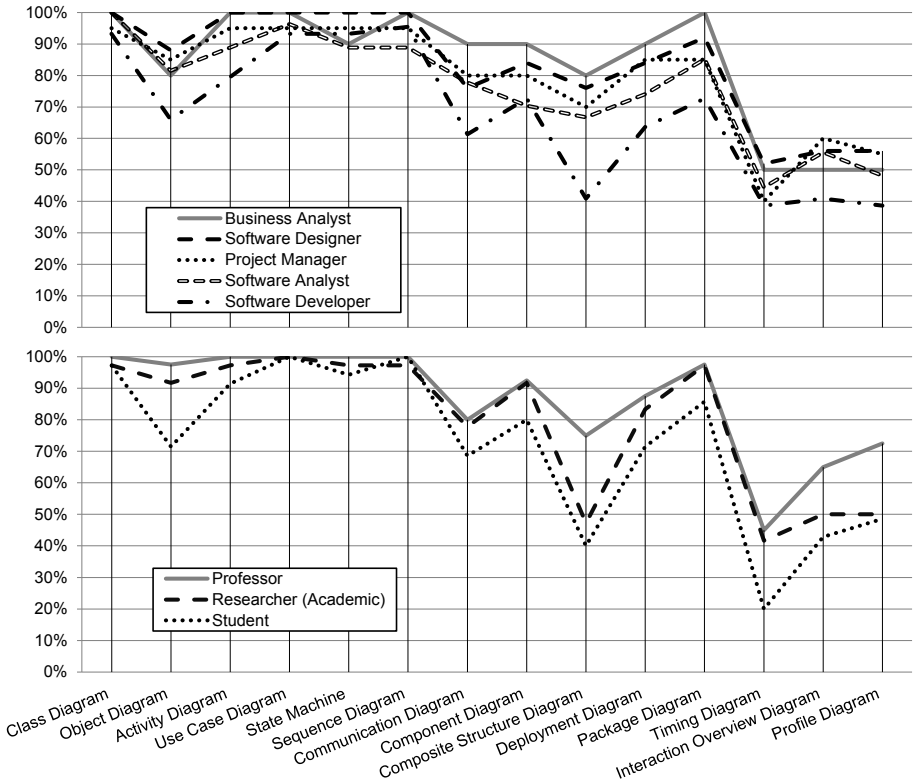

Fig. 7. UML diagrams knowledge: Professional and Academic working roles

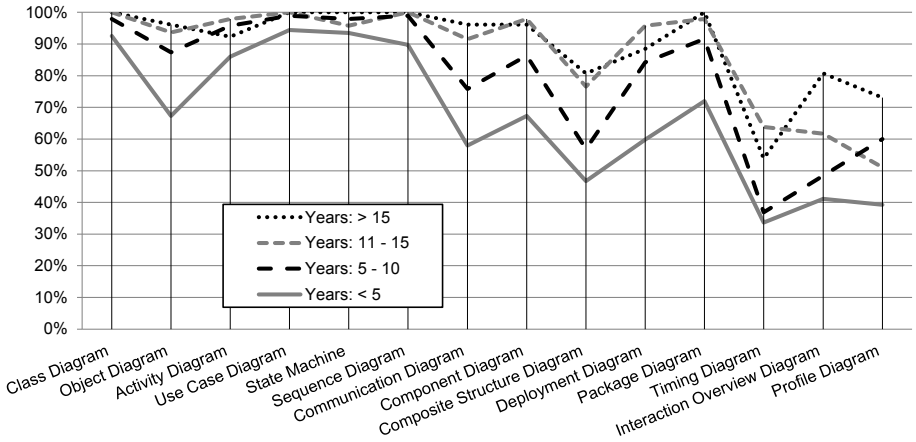

Fig. 8. UML diagrams knowledge: How many years have you been using UML?

The way the UML has been learned seems to influence the knowledge level, see Fig. 9. Indeed, while participants that studied UML on their own are only slightly better that who learned UML in university courses, we can observe that participants that studied UML in dedicated courses have a better knowledge in particular for communication, component, composite structure, deployment and package diagrams.

Finally, there is almost no difference in the UML knowledge if we consider the gender of the survey respondents (see Fig. 10).

Thus, we can answer to question **RQ4** saying that the factors we have considered (working role, seniority, education, and gender) do not have an influence on the UML knowledge.

3.6 Threats to Validity

In our opinion the main threats to validity of this study are the following: (1) non-probabilistic sampling method, (2) possible self-exclusion from participants not inter-

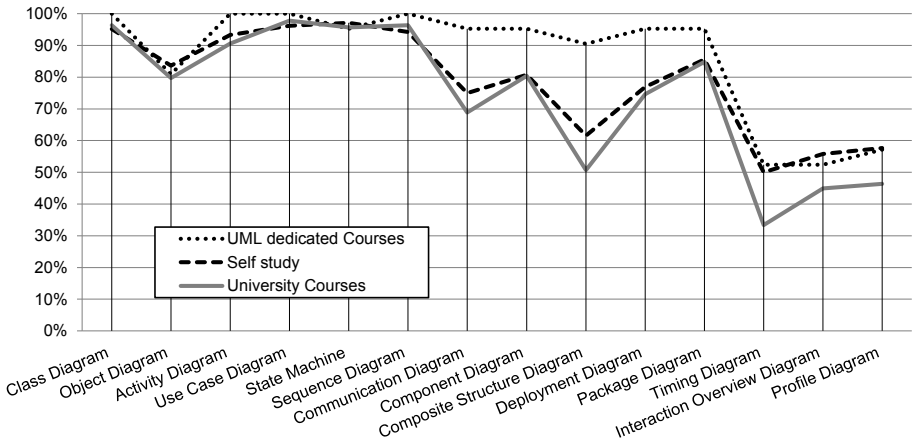

Fig. 9. UML diagrams knowledge: How did you learned UML?

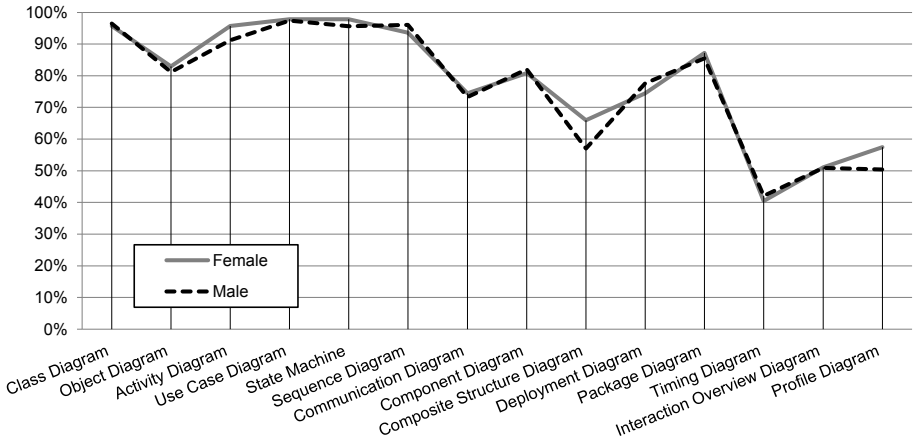

Fig. 10. UML diagrams knowledge: Male versus Female

ested in our survey, (3) possible non representativeness of the sample, and (4) sample size and not uniform geographic distribution of the data points.

We opted for non-probabilistic sampling methods, even if we know very well all the problems of this sampling, for two reasons. First, this survey is exploratory. Second, we thought that the target population was hard to identify and reach (especially non-Italian people), and of limited availability (this is often true in software engineering surveys). This should be carefully considered when interpreting the results we obtained: with this kind of sampling it is difficult to generalize the results to the entire population.

We cannot exclude that some participants could have avoided to answer because they have already a well-defined opinion about the UML (e.g. "it is not useful" and "there is nothing to discuss about it"). Self-exclusion is a well-known problem in Internet surveys, in particular when advertised by means of mailing lists and groups as we did.

We were expecting the questionnaire to be filled by people with at least minimal knowledge of the UML. Examining the collected data, we discovered that eight respondents answered "I have never seen it" at all the questions about UML diagrams and thus we decided to delete them from the sample. Moreover, we received 22 incomplete questionnaires. We contacted some of these respondents to understand the drop out reason. In several cases the motivation was little knowledge of UML. So we can conjecture that those that completed the questionnaire are people with a non-negligible knowledge of the UML, and thus perhaps the survey results may present a higher level of knowledge of the UML than the real one.

Finally, we are aware that the size of our sample is greater than the one of previously performed software engineering surveys [2, 9, 20, 21]. Of course, further geographically distributed data points are highly desirable to generalize our findings.

4 Related Work

UML is currently on of the most widely used modelling language [11] and it is often employed by companies in the software analysis and design phases [18, 21]. However, it

is also perceived as a very complex notation. For this reason, in the last decade, several works [3, 4, 8, 25] have been presented with the aim of identifying a relevant subset of the UML. In the tentative to find the "essential UML", Erickson and Siau [3] have conducted a Delphi study[9] with the goal of identifying a UML kernel for three well-known UML application areas: Real-Time, Web-based, and Enterprise systems. The participants to the study were asked to rate the relative importance of the various UML diagrams in building systems. UML overall results (i.e. non-domain specific) were: 100% for class and state machine diagrams, 95.5% for sequence diagrams, 90.9% for use case diagrams. All the others diagrams received a percentage lower than 50%, e.g. 27.3% for activity diagrams. This last finding is in contrast with our results where the activity diagrams are well-known (92%) and used (67%). Another personal opinion survey by Grossman et al. [4] about UML confirms the results of Erickson and Siau. Results indicate that the three most important diagrams are use case diagram, class diagrams and sequence diagrams. Wrycza and Marcinkowski [25], in another personal opinion survey, have tried to downsize the UML finding the most useful diagrams. The participants perceived use case, class, activity, and sequence diagrams as the most useful.

In comparison to these works, our survey is quite different for several reasons: (1) we invited very different categories of UML users to participate, from academia and industry, while, for instance, the sample used by [25] is composed only by students (180 in total) with a very homogeneous background (e.g. they have all studied UML on the same book), and Grossman et al. focused their study only on developers; (2) we tried to get a remarkable number of participants (i.e. 275) compared for instance to the 131 of Grossman et al. [4] and the 44 of [3].

Another personal opinion survey about UML (171 professionals in total), by Dobing and Parsons [2], points out another strong statement: "regular usage of UML components were lower than expected". The authors of [2] suggest that the difficulty of understanding many of the notations "support the argument that the UML may be too complex". The same claim, in more or less different forms, is present in several blogs, where many proposals of UML simplification are arising[10]. Maybe, the most authoritative is the one of Ivar Jacobson entitled "Taking the temperature of UML" [5], where he wrote: "*Still, UML has become complex and clumsy. For 80% of all software only 20% of UML is needed. However, it is not easy to find the subset of UML which we would call the 'Essential' UML. We must make UML smarter to use*". The need to simplify the UML is also shown by the recently released OMG draft proposal about this topic [19]. Moreover, the complexity of the UML seems to be one of the factors that limit its diffusion and usage in the industry [13].

The work in [8] shares our opinion on the importance to determine " which parts of UML are extensively used, which are scarcely or never used" (using their own words); however, differently from us they decided to analyse models publicly available on the web. Moreover, in their paper the frequency of use refers to the *concrete metaclasses* appearing in the metamodel (that should correspond to modelling concepts), and to *language units* (groups of tight-coupled modelling concepts), whereas we consider UML

[9] It attempts to form a reliable consensus of a group of experts in specialized areas.

[10] e.g. http://www.devx.com/architect/Article/45694 and http://blogs.msdn.com/b/sonuarora/archive/2009/11/02/simplify-uml.aspx

diagrams and visual constructs, that it is a user-view of the UML. The result of this study is that class, use case, and interaction are the most used language units, which roughly correspond to class, use case, sequence and interaction overview diagrams.

The main conclusion from a systematic literature review by Budgen et al. [1] is: while there are many studies that use the UML in some way, there are relatively few for which the UML is itself the object of study; there is a need to study the UML and its elements much more rigorously and to identify which features are valuable, and which could be discarded. Our work can be considered a first attempt in this direction.

5 Conclusions and Future Work

In this paper we have presented some results from a personal opinion survey, with respondents being both members of the academia and professionals, performed to investigate the level of knowledge and usage of the various UML diagrams, and to understand whether the working context, the working role, the education, the seniority (both obviously referring to the UML), and the gender influence the answer. This survey is part of a wider study aimed to gain information on the knowledge and usage of the UML diagrams and of the constructs of use case and activity diagrams, that includes a survey of static informative sources on the UML [14, 15] (an extended version of [15], including also the constructs of use case diagram will appear in [16]).

The found results show that the level of knowledge and of usage of the various types of diagrams is quite different, and that these differences are quite stable also when considering different categories of peoples, for example the timing diagram is almost always the less known/used. Furthermore, these differences among the diffusion of the various diagram types are consistent with those found by the survey on the static sources.

The UML diagram types may be placed in three different groups. In the first group there are the widely known and used diagrams. They cover the basic modelling functionalities: structural (class diagram) and behavioural (state machine, sequence diagram and activity diagram), while use case diagram has a specific role for the requirement specifications based on use cases (and it seems to be used also without any other UML diagram). All of them were already present in the UML version 1. In the second group there are diagrams useful for representing more specific aspect of systems/software (e.g. communication and object diagrams) or useful in specific cases (e.g. component and deployment diagram) or needed only to structure a model in packages (package diagram). They are known and used but at a less extent than those in the first group. The third group contains diagrams covering very specific aspects of a system (e.g. interaction overview and timing diagrams) or having a very specific role (profile diagram), while composite structure diagram has two different forms, one for modelling structured classes and one for collaborations (the latter are quite useful for example for modelling SOA based systems, see e.g. SoaML [12]). All these diagrams have been introduced in UML version 2, and the profile diagram only in UML 2.1.

As future work we would like: (1) to extend our personal opinion survey in several directions (e.g. adding more data points, extending the survey to other nations and extending the goal of our survey, for example for understanding why professionals choose to use some diagrams and ignore others), (2) to extend the static sources considering also web resources and model repositories, e.g. ReMoDD.

References

1. D. Budgen, A. J. Burn, O. P. Brereton, B. A. Kitchenham, and R. Pretorius. Empirical evidence about the UML: a systematic literature review. *Software Practice and Experience*, 41(4):363–392, Apr. 2011.
2. B. Dobing and J. Parsons. How UML is used. *Communications of the ACM*, 49(5):109–113, May 2006.
3. J. Erickson and K. Siau. Can UML be simplified? Practitioner use of UML in separate domains. In *Proceedings of 12th International Workshop on Exploring Modeling Methods for Systems Analysis and Design*, volume 365 of *EMMSAD 2007*, pages 81–90. CEUR Workshop Proceedings, 2007.
4. M. Grossman, J. E. Aronson, and R. V. McCarthy. Does UML make the grade? Insights from the software development community. *Information and Software Technology*, 47(6):383–397, 2005.
5. I. Jacobson. Taking the temperature of UML. Web site: http://blog.ivarjacobson.com/taking-the-temperature-of-uml/, 2009.
6. A. Jedlitschka, M. Ciolkowski, C. Denger, B. Freimut, and A. Schlichting. Relevant information sources for successful technology transfer: A survey using inspections as an example. In *Proceedings of 1st International Symposium on Empirical Software Engineering and Measurement*, ESEM 2007, pages 31–40. IEEE Computer Society, 2007.
7. B. A. Kitchenham and S. L. Pfleeger. Personal opinion surveys. In F. Shull, J. Singer, and D. I. K. Sjoberg, editors, *Guide to Advanced Empirical Software Engineering*, pages 63–92. Springer London, 2008.
8. P. Langer, T. Mayerhofer, M. Wimmer, and G. Kappel. On the usage of UML: Initial results of analyzing open UML models. In *Modellierung 2014, 19.-21. März 2014, Wien, Österreich*, volume 225 of *LNI*, pages 289–304. GI, 2014.
9. M. Leotta, F. Ricca, M. Ribaudo, G. Reggio, E. Astesiano, and T. Vernazza. An exploratory survey on SOA knowledge, adoption and trend in the Italian industry. In *Proceedings of 14th International Symposium on Web Systems Evolution*, WSE 2012, pages 21–30. IEEE, 2012.
10. T. C. Lethbridge. A Survey of the Relevance of Computer Science and Software Engineering Education. In *Proceedings of the 11th Conference on Software Engineering Education and Training*, CSEET 1998, pages 56–66. IEEE, 1998.
11. P. Mohagheghi, V. Dehlen, and T. Neple. Definitions and approaches to model quality in model-based software development - a review of literature. *Information and Software Technology*, 51(12):1646–1669, Dec. 2009.
12. OMG. *Service oriented architecture Modeling Language (SoaML) Specification Version 1.0.1*, 2012. Available at: http://www.omg.org/spec/SoaML/1.0.1/PDF.
13. M. Petre. UML in practice. In *Proceedings of 35th International Conference on Software Engineering*, ICSE 2013, pages 722–731. IEEE, 2013.
14. G. Reggio, M. Leotta, F. Ricca, and D. Clerissi. What are the used UML diagrams? A preliminary survey. In *Proceedings of 3rd International Workshop on Experiences and Empirical Studies in Software Modeling (EESSMod 2013)*, volume 1078, pages 3–12. CEUR Workshop Proceedings, 2013.
15. G. Reggio, M. Leotta, F. Ricca, and D. Clerissi. What are the used activity diagram constructs? A survey. In *Proceedings of 2nd International Conference on Model-Driven Engineering and Software Development*, MODELSWARD 2014, pages 87–98. SciTePress Digital Library, 2014.
16. G. Reggio, M. Leotta, F. Ricca, and D. Clerissi. What are the used UML diagram constructs? A document and tool analysis study covering activity and use case diagrams. In S. Hammoudi, L. F. Pires, J. Filipe, and R. C. das Neves, editors, *Model-Driven Engineering and Software*

Development, Communications in Computer and Information Science (CCIS). Springer, (in press). (Revised Selected Papers of 2nd International Conference on Model-Driven Engineering and Software Development).
17. G. Reggio, F. Ricca, and M. Leotta. Improving the quality and the comprehension of requirements: Disciplined use cases and mockups. In *Proceedings of 40th Euromicro Conference on Software Engineering and Advanced Applications*, SEAA 2014. IEEE, (in press).
18. G. Scanniello, C. Gravino, and G. Tortora. Investigating the role of UML in the software modeling and maintenance - A preliminary industrial survey. In *Proceedings of 12th International Conference on Enterprise Information Systems*, ICEIS 2010, pages 141–148. SciTePress, 2010.
19. E. Seidewitz. UML 2.5: Specification simplification. Presented at "Third Biannual Workshop on Eclipse Open Source Software and OMG Open Specifications", May 2012.
20. M. Torchiano, M. Di Penta, F. Ricca, A. De Lucia, and F. Lanubile. Migration of information systems in the Italian industry: A state of the practice survey. *Information and Software Technology*, 53:71–86, 2011.
21. M. Torchiano, F. Tomassetti, F. Ricca, A. Tiso, and G. Reggio. Relevance, benefits, and problems of software modelling and model driven techniques: A survey in the Italian industry. *Journal of Systems and Software*, 86(8):2110–2126, 2013.
22. UML Revision Task Force. *OMG Unified Modeling Language (OMG UML), Superstructure, V2.4.1*, 2011.
23. M. Völter. MD* best practices. *Journal of Object Technology*, 8(6):79–102, 2009.
24. D. S. Walonick. *Survival Statistics*. StatPac, Inc., 1997.
25. S. Wrycza and B. Marcinkowski. A light version of UML 2: Survey and outcomes. In *Proceedings of the Computer Science and IT Education Conference*, CSITEd 2007, 2007.