THE USE OF INDUCTIVE LEARNING IN INFORMATION SYSTEMS

Ilze Birzniece
Riga Technical University, Department of Systems Theory and Design, Meza 1/3, Riga, LV – 1048, Latvia, ilze.birzniece@rtu.lv

Abstract. Machine learning attempts to build computer programs that improve their performance by automating the acquisition of knowledge from experience. Inductive learning, one of machine learning paradigms, draws inductive inference from a teacher or environment-provided facts. Inductive learning enables the program to identify regularities and patterns in the prior knowledge or training data, and then to extract them as generalized rules. In literature there are proposed two ways of machine learning usage in information systems: (1) for building tools for software development and maintenance tasks and (2) for incorporation into software products to make them adaptive and self-configuring. However, considering information systems in more detail, division in three situations of inductive learning use in the context of information systems can be proposed, namely, first, in the information system development project management, second, to collect the information that is to be built in information system, third, to help the information system to adapt to the changing environment. The analysis of inductive learning role in information system development and usage is given. The future directions point to the e-commerce and similar domains on the Web for the role of inductive learning in information systems.

Keywords: inductive learning, machine learning, information systems, knowledge acquisition.

1 Introduction

Nowadays, the significance of information and information systems (IS) is increasing. Researchers and developers strive for sustainability and adaptability of IS [22]. Machine learning could be considered as one of the tools for ensuring the adaptation of IS. However, this topic has not been largely discussed in the 21st century IS conferences. The paper aims at investigating use of machine learning in general and inductive learning in particular in IS development to promote discussion on this issue. This survey could help to identify opportunities of embedding the learning capability in contemporary [1, 4, 23] IS solutions.

Research on Artificial Intelligence (AI) became active in sixties’ of the previous century leading to the emergence of a new type of software systems - knowledge based systems [9]. Most typical systems of this type are expert systems. Expert systems (ES) are software programs that accumulate and describe expert knowledge in some specific field with the aim to spread that knowledge and solve other tasks of this field. After creation of first expert systems knowledge acquisition has become a subject of computer science. Various methods have been used for knowledge acquisition. They can be divided in manual, semiautomatic and automatic methods [9, 14]. There are manual methods that are used by knowledge engineer in process of knowledge takeover, analysis and structuring. They include brainstorming, interviewing, and protocol analysis techniques. In case where a part of this manual process is carried out with the computer program, methods are called semiautomatic, e.g. knowledge acquisition interfaces. Automatic methods may perform their tasks without knowledge engineer assistance. As the knowledge takeover from human experts with manual methods is both time-consuming and expensive, in many domains the machine learning methods are more preferable because they are more effective and efficient [15]. Inductive learning is one of machine learning subfields highly estimated in data mining, rule inducing for autonomous systems and expert systems due to its ability to create transparent classifier [3]. Inductive learning methods have also been used in developing software products [21]. The purpose of this paper is (1) to show how inductive learning can be used in IS development and (2) to build basis for broadening existing application areas of inductive learning. The rest of the paper is organized as follows. Section 2 presents the definition of IS used for the purpose of this paper. Section 3 provides an overview of machine learning and inductive learning approach. In section 4 existing inductive learning applications in IS are surveyed. Section 5 points to the newest application areas of machine/inductive learning in IS development. Finally the conclusions are drawn.

2 What is information system

It is not an easy task to define what an IS is. Many different definitions have been proposed over the years by researchers and textbook authors. One simple and easy to understand definition is given by Davis: “Information system is a system in the organization that delivers information and communication services needed by the organization” [2]. Another proposed definition that is relevant to discuss in the context of this paper comes from Alter [2]. He defines IS as a special case of work system. The motivation for such definition is founded on fact that businesses operate through work systems. “An information system is a special case of work system in which human participants and / or machines perform work (processes and activities) using
information, technology, and other resources to produce informational products and / or services for internal or external customers” [2]. Defining IS as a work system implies that most of the concepts and knowledge that apply to work systems in general also apply to information systems in particular. This approach involves treatment of IS as a system, rather than a tool. System development often refers to developing software tools that meet requirements and satisfy the needs of users, rather than developing or modifying a work system in an organization. The tool view focuses on the form, affordances, and limitations of the interface. The system view includes those factors, but goes further in showing why the limiting factor is sometimes the person rather than the technology. This definition of IS is consistent with learning approach discussed in this paper because it includes both humans and technologies.

Before investigating the use of machine learning and inductive learning in IS, machine learning principles are outlined in next section.

3 Machine learning

Cios and Kurgan [5] define machine learning as the ability of a computer program to improve its own performance, based on the past experience, by generation of a new data structure that is different from an old one, like production rules from input data. The demand of machine learning applications, in particular in the areas of data, image and text mining, has created an urgent need for systems that can efficiently search for regularities or data descriptions in very large information sources [8]. Most typical applications of machine learning can be divided in three groups [21, 22]:

- Data mining in large data bases containing valuable implicit regularities that can be discovered automatically.
- Poorly understood problem domains where little knowledge exists to develop effective algorithm.
- Domains where programs must dynamically adopt to changing environment.

There is a wide range of methods to be used for machine learning [19, 17], e.g., artificial neural networks, Bayes classifier, K-Nearest Neighbours classifier, genetic algorithms, inductive learning, etc. Machine learning can be broadly classified into three categories: supervised learning, unsupervised learning and reinforcement learning [22]. The inductive learning is a special class of the supervised learning techniques. There are several more specific classifications of machine learning methods; classification varies depending on author and time period (e.g., different classifications are given in [22, 10]). Although machine learning algorithms are domain independent, in many domains generated descriptions or patterns need to have not only a high predictive accuracy, but also are required to be easy interpretable and comprehensible for user. Different applications may demand different description forms, i.e. reasoning system should be able to transform its results from one form to another. Inductive learning algorithms are preferable over other machine learning methods in systems where understanding of decision-making steps and further processing of results is needed. For instance, expert systems are systems where the rules induced by learning algorithms can be used [5]. Basic concepts of inductive learning, namely, the notion of inductive learning, overview of inductive learning methods, and general inductive learning system are discussed further in this section.

3.1 Inductive learning

Induction is a process of conversion of particular facts into general regularities. In computer science inductive learning is learning by example where a system tries to induce a general rule from a set of observed instances [12]. This involves classification - assigning the name of a class to every particular input. Classification is important to many problem solving tasks. Inductive learning methods are considered attractive for many real-life applications (e.g., medical diagnostic [3]), most due to their interpretability.

Inductive learning constructs a description of a function from a set of input/output examples. An example is a pair \((x, f(x))\), where \(x\) is the input and \(f(x)\) is the output of the function applied to \(x\). The task of induction is to return the hypothesis \(h\) that approximates \(f\) from a given set of examples [18]. Then, the generated hypothesis is applied to the new examples to predict their class membership [6].

Classification of machine learning methods isn’t strict. There is also no agreement among authors concerning the scope of methods belonging to inductive learning. In this paper with this name are denoted methods which can provide a classification for unseen instance not only on the basis of a given set of examples (as, e.g., K-Nearest Neighbours classifier does), but which also offer generalized model for classifying new instances. Such knowledge compression can be done by decision trees and related approaches (e.g., rule induction).

There are several dimensions along which learning algorithms can be classified. Depending on the way of learning, inductive learning methods can be divided in incremental and nonincremental (or static) ones. Other option to divide inductive methods is to consider the way the classifier is obtained and described – whether it forms decision tree, generates rules or combines both. For instance, the most popular algorithms in each category
are as follows [5, 17]: ID3, C4.5, and CART for decision tree, AQ for rules and CN2 for hybrid methods (quite often added to rules generating methods).

3.2 General inductive learning system

Generally classification task with inductive learning is organized as follows. First, the classifier for particular domain is formed; afterwards it is used for automatic or semi-automatic classification of new instance. Classifier formation consists of two parts, classifier training and testing, which is followed by applying it. General schema of inductive learning is proposed in Figure 1.

![Figure 1. Steps of classifier forming and using](image)

In the training phase an inductive learning method is used to infer description (either in form of decision tree or classification rules) from a given set of examples, where the class for every single record is known. Example set can be accumulated form observations, generated by expert or from both. The evaluation of description accuracy for unseen examples from the same domain follows. The class is assigned to every test example in accordance with description gained in training step. As the test example’s true class is known, one can rate the accuracy of predictions and the overall accuracy of the classifier.

4 The role of inductive learning in IS context

A software system is a part of IS. Machine learning and inductive learning methods are used especially in software systems. Zhang [21] points to two ways of machine learning usage in software systems. He states that inductive learning and other machine learning algorithms can be used for both (1) building tools for software development and maintenance tasks and (2) incorporation into software products to make them adaptive and self-configuring [21]. Considering IS development in more detailed way, „building tools for software development and maintenance tasks” can be divided in two more specific parts. Learning can refer to the IS development project management and to the development of particular IS that needs the results gained by inductive learning. So the division in three (instead of two) IS issues and situations of inductive learning use in the context of these issues is proposed.

1. **Inductive learning could be used in the IS development project management.**
   
   With the help of inductive learning models for software development process can be built. For example, software engineering data can be analyzed to predict software costs [22].

2. **Inductive learning could be used in particular IS development.**
   
   Some parts of IS may not be created without knowledge acquisition of problem domain, e.g. the initial rules base should be inferred to build it in the autonomous IS.

3. **Inductive learning could be applied in IS usage.**
   
   Inductive learning system could be implemented in the IS in order to let it learn from experience while it is working. An example of such a system is a practical diagnostic supporting workbench, which incrementally incorporates new competences with an existing knowledge base [3].

   There are four fields ("learning issues" in the context of this paper) identified which are tightly connected with IS development or IS usage and involve inductive learning. They are knowledge acquisition, software development, expert systems, and complex system modeling. These learning issues are discussed in more detail in sections 4.1. – 4.4. They aren’t isolated as knowledge acquisition is not just an independent subject but also the base for inductive learning use in software development and expert systems. The overlapping between three IS issues and four learning issues is reflected in Figure 2. The references point to the sources where the use of inductive learning in particular learning issue in connection with the IS issue is mentioned.
Knowledge acquisition emerges in all IS issues showing the need for this field in all IS stages. The need of learning in expert systems can show up both in IS development stage (creation of rules to be embedded into the system) and in IS use (generation of new rules from experience).

### 4.1 Knowledge acquisition by inductive learning

Knowledge acquisition in IS context is most often connected with expert systems. The main reason for research on knowledge acquisition is to develop techniques for making expert systems easier to build and to make them more explainable, robust, and intelligent [14]. Knowledge engineering captures human knowledge and places it into a computer system where it is used to solve complex problems normally requiring a high level of human expertise.

Knowledge acquisition is recognized as one of the major problems in an expert system development. Knowledge elicitation from domain experts and machine learning are two distinct approaches to knowledge acquisition [15]. Knowledge takeover from human experts are both time consuming and expensive. In many domains the machine learning methods should be preferred because they are automated and more effective and efficient [15]. Comparative review of knowledge engineering and inductive learning is done also by Jackson [7]. Both knowledge elicitation from domain experts and machine learning are employed in medical domain – tonsillectomy/adenoidectomy patients are classified into normal and abnormal diagnostic groups with respect to their predispositions to bleeding.

While Oprea [15] and Jackson [7] separate manual knowledge engineering and automated machine learning methods, Langley and Simon [10] argue that machine learning may never entirely replace knowledge engineering as a framework for knowledge based systems. It only increases the levels of automation in the knowledge engineering process. Taki [14] proposes the idea that knowledge acquisition should be done by integrating manual and inductive (automated) methods, where human expert evaluates and corrects automatically obtained knowledge, if necessary.

In 1995 Langley and Simon [10] stated that inductive learning and other learning methods will become increasingly prevalent in progress toward automation in knowledge acquisition. Nowadays, the use of software tools to aid the acquisition process has increased, but none dares to state that knowledge acquisition has become an automated process.

### 4.2 Inductive learning in software development

Many software engineering tasks could be formulated as learning and classification problems [21]. Machine learning and data mining methods (including inductive learning) can be and are used in developing software products [20, 21]. Classification that can be carried out with the help of inductive learning methods is one of mining tasks in software development [20]. For machine learning applications the following areas of software development and maintenance are identified:

- requirement engineering (knowledge elicitation, prototyping), which overlaps with knowledge acquisition;
- software reuse (application generators);
- testing and validation,
• maintenance (software understanding);
• project management (cost, effort, or defect prediction or estimation).

The strength of ML methods in software engineering lies in the fact that they have sound mathematical and logical justifications and they can be used to create and compile verifiable knowledge about the design and development of software artifacts [22].

Considering the use of inductive learning in software system development, particular applications from [22] report should be mentioned. In software quality prediction a decision tree based method is used to generate measurement based models of high-risk components. Software engineering data are analyzed to predict software costs. In maintenance task effort prediction, software resource analysis, correction cost estimation, and reusability prediction predictive models are built through decision tree induction. Inductive learning finds its application also in organization of reusable software engineering components by predicting the cost of rework. Since requirement acquisition is related to knowledge acquisition, where inductive learning methods take of great importance, inductive methods are used likewise in requirement elicitation. The observed situation proves that inductive learning approach has a stable position in software engineering process, most frequently dealing with the issue of how to build models or estimate certain property of software development process.

The use of data mining in software development has been described in a comprehensive review [20]. It shows up a practical classification tasks that are carried out for dynamic call graphs to determine the bug location in software and for text classification to assign bug reports to specific developer.

4.3 Inductive learning and expert systems

Expert systems are widely discussed in the literature. They are closely connected with knowledge acquisition as mentioned before. During knowledge-based system development the knowledge acquisition process is crucial and time consuming. Inductive learning methods can help in automating this process. Automation increases the speed and reduces the cost of development by reducing the amount of time needed from experts and knowledge engineers [16]. It could also be argued that, in some cases, automation may uncover knowledge that might otherwise be left unnoticed.

Diagnostic supporting tool (that is an expert system actually) described in [3] shows an example of expert system which use inductive learning facilities not only in development level of this system, but also in usage phase. Incremental decision tree is used to enrich systems knowledge-base with new data merging them into the existing tree model.

4.4 Complex system modeling

Inductive learning algorithms are also used for complex system modeling. The fundamentals of this approach are given by Madala and Ivakhenko in [11]. This monograph surveys new types of learning algorithms for modeling complex scientific systems in science and engineering.

One of complex tasks modelled by inductive learning is academic timetabling [13]. This task is complex due to dependencies on various rules and constraints [13]. Three inductive learning algorithms has been implemented and compared on timetabling for the purpose to see their ability towards solving complex problems and adapting to the new environment by changing rules. In building an optimal timetable system, a lot of rules, dependencies, and constraints are needed. Failure to identify that knowledge may cause the system to have an ineffective solution. The ability of inductive learning to learn and generalize rules is fully used in timetabling so that the system has its own competence to analyze the data and gather the needed information for proper functioning. The conclusions driven from real experiments with inductive learning algorithms ID3, AQ and ILA show that inductive learning algorithms can be successfully implemented into timetabling and complex problem domain overall [13].

4.5 Inductive learning methods used in practice

In Table 1 there are gathered particular inductive learning methods used in learning issues discussed previously in this section. This summary is not meant as complete listing of inductive learning algorithms that are used in these tasks. It just gives an insight into variety of inductive learning methods used in IS context.

Table 1. Inductive learning methods in IS fields

<table>
<thead>
<tr>
<th>Knowledge acquisition</th>
<th>Software development</th>
<th>Expert systems</th>
<th>Complex system modeling</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID3 [15], 7, 14, C4.5 [15], ILA [15], AQ [14], CN2 [14]</td>
<td>Decision trees (not specified) [22], ID3 [21], C4.5[21], Assistant [21]</td>
<td>*Learning [3], ITI (Incremental Tree Induction) [3], CLIP [5]</td>
<td>ID3 [13], AQ [13], ILA [13]</td>
</tr>
</tbody>
</table>
5 The new frontiers

It is obvious that most optimistic papers about automation of IS development with the help of machine learning methods are tree to eighteen years old. In IS conferences in 2009 (AMCIS, PACIS, CAiSE, ISD) such an issue almost can’t be found.

In recent years machine learning applications find their place in e-commerce and similar Web related areas. One of such an area is interactivity between Web content producers and consumers [4, 23]. Customer reviews posted on the Web have grown significantly. Because customers represent the primary stakeholder group of a company, understanding customers’ concerns expressed in reviews could help marketers and business analysts to identify market trends and to provide better products and services [4]. The approach consists of the steps of sentence extraction, aspect identification, sentiment classification, and review summarization. Likewise Zhang [23] describes analysis relative to e-commerce involving mining the product reviews that thrive on the Web. Typical tasks include polarity prediction – distinguishing positive, negative, and neutral reviews – and opinion extraction. This new task in text sentiment analysis improves product review ranking services, helping shoppers and vendors leverage information from multiple sources. Different but closely related field to e-commerce is Question Answering Communities (such as Naver, Baidu Knows, and Yahoo!) [1]. By posting questions for other participants to answer, information seekers can obtain specific answers to their questions. Modeling information-seeker satisfaction uses machine learning approaches.

Only some of new challenges for automated learning tasks in information systems were mentioned above. It seems promising for inductive learning to find its place in dealing with such problems. This possibility has to be studied in detail.

To extend concepts shown in Figure 2, learning processes within e-commerce could be placed in IS usage issue (see Figure 3). This field is initiated with knowledge acquisition because knowledge acquisition is used to accumulate product reviews.

![Figure 3. Extended view to inductive learning applications in IS context](image)

Figure 3 shows only areas of inductive learning application. To understand better the potential of inductive learning in these IS issues in the context of contemporary technological and economical environment, it is necessary to reveal and define workable characteristics of these IS issues relevant to inductive learning.

6 Conclusions

This paper summarizes the existing applications of inductive learning in IS development and usage. A new classification of the use of inductive learning in IS context is proposed. Inductive learning applicability in IS is divided in three issues: (1) inductive learning could be used in the IS development project management; (2) inductive learning could be used in development of particular IS; (3) inductive learning could be applied in the usage of IS.

There are four fields identified which are tightly connected with information system development or information system usage and involve inductive learning. They are knowledge acquisition, software development, expert systems, and complex system modeling. These fields aren’t isolated as knowledge acquisition is not just an independent subject but also the base for inductive learning use in software development and expert systems.
Future work includes revealing and specifying characteristics of IS issues defined in this paper to improve IS development and usage in the spheres related to machine learning and inductive learning.

7 Acknowledgement

This work has been supported by the European Social Fund within the project „Support for the implementation of doctoral studies at Riga Technical University”.

References