Robot Tutoring: On the Feasibility of Using Cognitive Systems as Tutors in Introductory Programming Education

A Teaching Experiment

Sarah Müller
University of Applied Sciences
Neu-Ulm
Neu-Ulm, Bavaria
Sarah.c.Mueller@student.
hs-neu-ulm.de

Bianca Bergande
University of Applied Sciences
Neu-Ulm
Neu-Ulm, Bavaria
Bianca.Bergande@hs-neu-ulm.de

Philipp Brune
University of Applied Sciences
Neu-Ulm
Neu-Ulm, Bavaria
Philipp.Brune@hs-neu-ulm.de

ABSTRACT

Cognitive Systems and Artificial Intelligence (AI) are getting increasingly popular in many application domains. An upcoming field for AI are educational applications. Natural language processing (NLP) in combination with cognitive systems may provide many interesting possibilities regarding automatic advice and tutoring of students, e.g. by answering students' questions. Therefore, this paper investigates the possibility of using a cognitive software, namely IBM Watson, as a tool for developing a virtual tutor to answer common questions of students during an introductory Java programming course at a German University of Applied Sciences. The prototype software is evaluated using a qualitative study among participants. While the results indicate the future potential of such systems in education, in its current form it could only replace a tutor in very limited and specific scenarios.

KEYWORDS

Cognitive System, AI, Programming Education, IBM Watson

ACM Reference Format:

Sarah Müller, Bianca Bergande, and Philipp Brune. 2018. Robot Tutoring: On the Feasibility of Using Cognitive Systems as Tutors in Introductory Programming Education: A Teaching Experiment. In *ECSEE'18: European Conference of Software Engineering Education 2018, June 14–15, 2018, Seon/Bavaria, Germany*. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3209087.3209093>

1 INTRODUCTION

An upcoming field for artificial intelligence (AI) is the field of education [1, 3, 11, 12]. Due to the specific dynamics in the field of education, it is interesting to explore the possible role of AI as a teacher [11]. The advantages and opportunities of AI in education include 24/7 availability, neutrality towards minorities in the classroom and the capability to master a lot of subjects, while a human

teacher is naturally restricted to just a few subjects [11]. On the other hand, human teachers can interact with students in a flexible way since they have emotional intelligence, can create solutions to unforeseen events due to creativity and are able to create teaching materials on their own [11]. Can availability and neutrality beat emotional intelligence and creativity?

Therefore, in this paper a prototype solution for a virtual tutor in an introductory Java programming course is proposed and evaluated. The prototype is based on the IBM Watson platform¹ and is intended to be able to answer students' questions. It is evaluated by a qualitative study among real first-year students of an introductory programming course at a German University of Applied Sciences.

The rest of this paper is organized as follows: First, an overview of the related work on AI in education is provided in section 2. Section 3 describes the research design and section 5 presents the evaluation and its results. We conclude with a summary of our findings.

2 RELATED WORK

AI is not a new topic in computer sciences, since it has been discussed for at least 68 years, when Alan Turing created the Turing Test [14]. The Turing Test is simply put a scenario in which one participant holds a written, on screen conversation with two other participants, a human and an artificial intelligence and afterwards has to decide, which one of his or her discussion partners is human and which one is the AI. If it is impossible for the participant to distinguish both from one another the AI passed the Turing Test and can be viewed as equal to the “natural intelligence” of a human being [14]. AI is a general term for a variety of computing phenomena and there is still not one, final definition on what AI is. For this teaching experiment AI is defined as follows.

A common feature in AI definitions is the capability of an artificial artifact to perform acts of a cognitive nature, such as the autonomous interpretation of input or recognition of patterns in data [1, 3, 11, 12]. The definition of AI changes due to the focus of research [3, 5, 9]. While some focus on the ability of cognitive systems to adapt to input, or to “learn” so to say, other research is more bound to the emulation of the argumentation process of human beings [3, 9].

Both features of cognitive systems combined are already part of research on AI in the field of education [3, 5]. One form of AI is machine learning, which means the ability of computer programs

¹<https://www.ibm.com/watson/>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ECSEE'18, June 14–15, 2018, Seon/Bavaria, Germany

© 2018 Copyright held by the owner/author(s). Publication rights licensed to Association for Computing Machinery.

ACM ISBN 978-1-4503-6383-9/18/06...\$15.00

<https://doi.org/10.1145/3209087.3209093>

to simulate human learning, including the aforementioned features of AI [6].

For the purpose of this paper, AI is defined as a variety of autonomous acts of reasoning and creation of output performed by a machine. The differences between machine learning, cognitive computing and the emulation of human social behaviour and implications on a technical and ethical level are negligible due to the experimental stage of the investigated teaching scenario [1, 3, 11, 12]. Machine learning includes the recognition of patterns in data and predictive analysis of the occurrence of future patterns and variations of patterns based on complex algorithms, which makes this aspect of AI particularly interesting for educational purposes for the following reasons [2].

2.1 AI in Education

Ever since IBM Watson defeated the two best human players of the quiz show *Jeopardy!*, an open-domain question and answer game where participants must answer different questions about a range of subjects, cognitive systems' ability to create answers to questions has been widely discussed [13]. Answering questions is also a major task for teachers in the classroom, so that the idea of AI as a resource in education is standing to reason.

There are different opinions about the state of the art of AI in education. While big companies invest a lot of work into AI solutions in education and the usage of robots as teaching assistants is discussed, others argue that the usability of AI is limited to a small number of learning scenarios, since even intelligent machines only operate within the limits of their system [10–12]. The possible role of intelligent machines for teaching scenarios is yet undefined, but first user surveys about implementations of AI in education are promising to some extent [4, 10].

Recently, an adaptive free tool for elementary schools has been released, which provides support for teachers in organising and improving their work and gradually adapts to the individual users needs without additional adjustments, thus making this tool a cognitive system [4].

The ability of a machine to adjust to the individual needs of a user has major prerequisites: first, to be able to recognize predominant patterns in user behaviour, then make assumptions about future patterns based on analysing existing user data and afterwards to adapt according to the patterns, which are most likely to occur in the future. These challenging tasks are all parts of AI, mostly assigned to a domain called machine learning [1].

2.2 Machine Learning

The machine learning process can be explained by using an example of a pattern recognition. Pattern recognition includes the recognition of speech, handwriting or a face by a computer [6]. The machine learning process for the processing of handwriting is exemplified as follows. For the example, many different handwritten numbers are represented by a vector x . A target vector t , which is available for each handwritten number, represents the right digit. As a consequence, by running the learning algorithm, the vector x generates a vector y which is similar to the given vector t . The

accuracy of the algorithm depends on the accuracy of the training phase and the training data. The term, describing the correct assignment of new examples, is called generalization [1].

Mostly the input variables (in the example the vector x) are preprocessed to make the solving of the problem easier. Another expression for this preprocessing step is feature extraction. In the example, the images of the handwritten digits were brought to the same size (28 x 28 pixels) which leads to a consistent location and size of the handwritten number. It is important to preprocess new data the same way like the training data. This step might also fasten up the performance of the computer [1]. When the training data includes input and output vectors, they are called supervised problems. Classification problem is the name for the process of an input vector being tried to be categorized to a limited number of categories [1].

As mentioned in subsection 2.1, these processes of recognition, classification and generalisation are also at work in the free teachers tool [4].

If AI tools are already capable of organising educational processes, finding the right answers to questions and mimic human behaviours in written and spoken communication they should also be able to replace the teacher in a teaching scenario. Therefore, the research question derived from our research is:

RQ - How successful can AI take over the tasks of a teacher?

To investigate AI as a teacher in higher education, the following research design was conceived.

3 RESEARCH DESIGN

To answer the research question a teaching experiment using AI was conducted. The experiment design is based on design science research as illustrated in Fig. 1 [7]. The chosen environment in this teaching experiment is an university of applied sciences, which, as an institution in higher education (*organizational system*, Fig. 1), is also affected by current trends in education and could therefore benefit from AI (*problems and opportunities* Fig. 1). The university of applied sciences offers a number of courses with tutorials for its students (*People* Fig. 1), which need to be carried out with a tutor or lecturer and require a lot of time and staff as a consequence (*organizational and technical systems* Fig. 1). The *knowledge base* (Fig. 1), which leads to the research question is thoroughly elaborated in Section 2.

The course Programming technique was chosen to conduct the teaching experiment (*field testing* Fig. 1) since it is one of the most popular courses on campus with many tutorials, which means that the institution has to invest in lecturers and infrastructure to hold tutorials, that help many undergraduate students. While support is provided in the tutorials, it is not possible to perfectly adapt it to the personal needs of each and every student due to limited time in the lessons. Providing an intelligent solution to meet students' needs and save time and money could be beneficial for both students and the institution. To replace a human tutor the AI should be able to reply to questions and give examples. As a consequence, the *artifact* (Fig. 1) for the teaching experiment will have to be a cognitive system, capable of machine learning and fit to be used in an educational setting.

Figure 1: Design Science Research Process [7]

To build the *artifact* (Fig.1), an IBM Watson solution was chosen due to the following reasons. IBM offers the possibility to use different tools of Watson in a free lite version in the IBM Cloud [4]. IBM Watson as a cognitive system can reprogram itself with the help of machine learning and the knowledge it gains during use f.e. during the competition in *Jeopardy!* mentioned in subsection 2.1 [13]. It uses various innovations in question answering and natural language processing [15]. After adjusting Watson to the contents of the course, it was put to the test in a tutorial on programming technique with a randomized sample of undergraduate students, who could ask Watson questions and were asked to evaluate their experience with their AI tutor afterwards.

3.1 Communication with Watson - a brief description

To find out if it is possible that Watson answers students' questions during a programming technique tutoring session, it must fulfill different expectations. First off, it must find out whether a student is asking for a definition or wants to have an example. After that it has to be able to decide, which answer is the right one and give it to the student. As Watson offers lots of different tools, it was necessary to find the one that fits best. First, a natural language understanding tool was created to no avail, since it was not able to classify spoken language. It only searches for keywords, emotions and relations. As a consequence, it cannot figure out if the student asks for a definition or an example. As an alternative, a conversation tool was designed.

To understand language is not easy for a computer since language can use riddles, puns and irony which makes it hard to process, e.g. "Why do we fill in a form by filling it out?" [8]. Watson too cannot understand every single word, but recognizes the characteristics of the language instead. Within the conversation tool, Watson has given intents which stand for the purpose of the user's question. The use of entities enables the creator to define different actions for one intent and with the help of the dialog, the creator can formulate the fitting response. For the question answering game, it collated the text passage from a question with a text passage from the possible answer.

The first step is the question analysis. Here IBM Watson segments the question and finds the main characteristics. This is the input for the second step, the hypothesis generation. In this step, IBM Watson browses millions of documents. Using the keywords, IBM Watson tries to find a valuable answer. Hypothesis and Evidence Scoring is the third step. Here the different candidate answers are

Figure 2: the training process from top to bottom: First, there is an example for an intent, as entered by the creator followed by the the Dialog View of Watson, showing the predefined, correct answer for the intent.

rated by many different algorithms such as looking for matching terms and synonyms or type coercion. Every single algorithm in this step produces a different number of scores, which indicate how congruent the possible response is with the question in relation to the orientation of the algorithm.

In the final Confidence Merging and Ranking step, the final score of each candidate's answer is measured with a statistical model. This model is used to aggregate the amount of confidence IBM Watson has in the proof that this answer matches to the question. The model comes from the training period of Watson[8, 13].

4 DESIGNING THE ARTIFACT

The development of the prototype started by creating an account at the IBM Cloud Bluemix, which provides Watson. After creating the conversation service of Watson, it is possible to launch the tool and get started with its training.

At the beginning, it is necessary to define the goal of the users input, called intent like "#javaVirtualMachine" on the upper left side in Fig. 2. While defining the intent further keywords, like "class", "java" or "bytecode" in this case, can be added. If the conversation tool recognizes one of the keywords during a dialog, it will answer with the response belonging to this intent.

This response is added in the dialog view as shown in the lower half of Fig. 2. By labelling the associated intent, it is possible to define the answer and the following action. The conversation tool does not propose an answer automatically. Therefore, the creator has to formulate each response the tool should be able to give on its own later.

Finally, the tool must be trained. During testing Watson in a dialog wrong answers may occur. Watson starts training by the user clicking on the suggested intent and changing it to the correct one. Watson mostly adds some examples to the chosen intent.

It is important to work very accurate while training and to be as specific as possible to enable the tool to answer a specific question with a matching answer. Though it is possible to put a definition, an example and differences to other keywords in one intent, one intent for each class of reply was created. Putting all classes of answers into one intent can result in a long and unspecific answer, which may cause understanding problems later and corrupts the reason why it was created in the first place.

5 EVALUATION

5.1 Data Collection

To evaluate Watson and the conversation tool as AI in a tutorial a randomized sample of students “asked” Watson questions of their interest about the course programming technique, by typing them into the field of the “Dialog” interface. The sample consists of six undergraduates, who visit the lecture and one tutorial in this course. It is an introduction to computer science and programming, that focusses on basic concepts such as object oriented programming and basic coding skills in JAVA ² using eclipse ³ as integrated development environment. They mostly studied in their 1st semester by the time the experiment was conducted and had to answer the following questions right after their experience with Watson.

1. Which question are you asking?
2. What is the given answer?
3. Does this answer respond to the question?
- 3a. If no, what does not meet your expectations?
4. Does something need to be added/changed to the answer?
5. Further Comments?

Since the interviews were held in German, their answers were translated for this paper.

5.2 Results

When the first participant (student1 or S1) asked Watson “What is a class?” it replied: “a class is a blueprint for objects. In Java, a class starts with “class”, the name of the class and the braces, which enclose the content of the class. In there the structure (attributes) and the behavior (methods) are included. Example: class car /*attributes*/ private int leistung; private String farbe; /*methods*/ public void drive() system.out.print. ln“The car drives”)return;...”. The participant was stunned by the accuracy and felt, that her expectations were fully met. She added to like how fast and sufficient it gives answers was a real plus for her.

The second participant wanted to know “How is an object created?”, and Watson gave a long and very detailed answer that involved the adaptation of constructors to add attributes to an object, creating a new object in the process. The reply was so confusing, that the student had to read this reply a number of times, before coming to the conclusion that Watson was also right this time. As

a result the user gave mixed reviews, since the load of unnecessary details was utterly confusing to her.

The third participant also asked for a definition of what a class is and Watson gave exactly the same response as the first participant. This time the student had a hard time to follow Watsons explanation, since she was confused about the exact meaning of the terms “methods” and “attributes”. The next participant also asked for a definition of “class” in programming, but unlike S2 he was very satisfied with the reply.

The fifth participant asked a different variety of the topic “class” and wanted to know: “How can I create a class?”. Watson did not seem to notice any difference to the other class-question and replied word by word with the same response it had a number of times before. Nonetheless, S5 was very satisfied with Watsons answer and thoroughly surprised about “how well it works”.

The sixth and final participant also had a more specific question and asked how a constructor is created. Watson answered: “The method which is called to create an object, is called constructor. To add values to the attributes of the new created object, some parameters can be added to the constructor. Destructors are the methods which are called to dissolve an object. Destructors are normally free of arguments.” While S6 said, that it gave her a clear idea about what a constructor is, the explanation for eclipse creating the constructor itself should be added to give a sufficient answer in her opinion.

5.3 Interpretation

Overall, Watson always gave a correct answer to the question, but to use Watson as a real tutor it would be necessary to formulate the answers more detailed and more extensive to make different variations for similar questions possible. The problems detected match to the weaknesses of AI in educational settings as stated in [11] and can be divided into two major weaknesses:

- **1. Precision** Watson can cause more confusion then create answers, depending on the knowledge and underlying intention of the student asking. One example would be the superfluous second part of the answer for the second participant, which caused disarray. It should be able to interpret questions precisely and then give the matching reply. This points to another major weakness of Watson, which is linked to precision and investigated further in the following key-point.
- **2. Adabtability** Watson cannot detect different underlying intentions of a question, since a definition of “class” and an instruction how to create a class are two different motives. It also failed to provide the detailed information S3 and S6 hoped for to fully understand the concepts. The limited variation in its replies points to one of the limitations of AI: limited “creativity” and no empathy or intuition for underlying intentions or different interpretations of a question[11]. This shows that one of the desirable traits - individual adaptability - is strictly limited to the options explored in the initial training phase.

Watson lite does not meet the requirement to take over tasks of a teacher as described in 3.1 satisfactorily, since it lacks adaptability and precision. The answer to the research question therefore is:

²<https://www.java.com/de/>

³<https://www.eclipse.org/>

IBM Watson lite can only take over very simple tasks except you create many possible replies and keywords. This requires a lot of time and investment to this date, so that it can not successfully take over the tasks of a teacher in an effective way in its present version.

5.4 Limitations and Further Research

The evaluation of the described approach was only carried out once with a very small group of students, who were sampled randomly. A bigger number of interviewees on a bigger range of topics and questions would have had a higher significance. The obtained results must therefore be regarded as a qualitative prestudy, which has to be exemplified with a bigger sample once the tool is ready to work. Nonetheless, the survey provides valuable feedback for further research about AI in computer science tutorials for undergraduate students.

The evaluation showed, that it would be possible to use the conversation tool of Watson to help tutors answering the students' questions. But the conversation tool has the limitation to need all answers being written by the creator beforehand. In this case it can only be seen as a better FAQ tool.

In this paper, the Watsons lite Version was used. Watson offers different options: lite, standard and premium. Furthermore, two tools of Watson are not available while using the lite plan: the natural language classifier (only available with the standard plan) and the knowledge studio (only available for specific organizations). With the natural language classifier tool, Watson could be able to define the answer on its own, which could lead to more precise and up to date answers. The research in this paper also was limited by only using Watson. There are a few companies active in the field of AI in education that may offer some free solutions in the future. Moreover, it would be desirable to extend the prototype with putting a speech to text tool at first, to be able to talk with Watson. This in combination with a voice output of Watson would make a robot teacher more real.

6 CONCLUSION

In this paper the possibility of using a cognitive software, namely IBM Watson, as a tool for developing a virtual tutor to answer common questions of students was evaluated in the context of an introductory Java programming course. The prototype software was evaluated in a field test by using a qualitative study among selected participants of an introductory Java programming course at a German University of Applied Sciences. The results of the evaluation indicate that it would be possible to use the proposed tool in a course for the intended purpose.

However, the conversation tool of Watson lite needs all keywords and answers given in a rather detailed and comprehensive way. Therefore, it is very time-consuming to prepare it for this task. This effort is only acceptable if the tool can be used for several courses with the same content.

Another aspect is that students learn in different ways. One understands complicated derivations while the other needs a detailed example. But with the conversation tool of Watson lite, always one fixed answer is prescribed. It is advisable to use IBM Watson lite as a complementary tool to the regular interaction with a human teacher, but it cannot replace them at this point.

To conclude, an AI solution like IBM Watson lite can be a valuable tool in education, but is limited to very specific didactical settings and software limitations.

ACKNOWLEDGMENTS

The present work as part of the project EVELIN was funded by the German Federal Ministry of Education and Research under grand number FKZ 01PL17022E. The authors are responsible for the content of this publication.

REFERENCES

- [1] Christopher M Bishop. 2006. Machine learning and pattern recognition. *Information Science and Statistics*. Springer, Heidelberg (2006).
- [2] Erik Brynjolfsson and Andrew McAfee. 2014. *The Second Machine Age: Wie die nächste digitale Revolution unser aller Leben verändern wird*. Plassen Verlag.
- [3] Mauro Coccoli, Paolo Maresca, and Lidia Stanganelli. 2016. Cognitive computing in Education. *Journal of e-Learning and Knowledge Society* 12, 2 (2016).
- [4] J. Crozier. [n. d.]. *By Teachers For Teachers: Teacher Advisor With Watson*. version 1.6.0.
- [5] Seth Earley. 2015. Executive roundtable series: machine learning and cognitive computing. *IT Professional* 17, 4 (2015), 56–60.
- [6] Isabelle Guyon and André Elisseeff. 2006. An introduction to feature extraction. In *Feature extraction*. Springer, 1–25.
- [7] Alan R Hevner. 2007. A three cycle view of design science research. *Scandinavian journal of information systems* 19, 2 (2007), 4.
- [8] Rob High. 2012. The era of cognitive systems: An inside look at IBM Watson and how it works. *IBM Corporation, Redbooks* (2012).
- [9] Thomas R Hinrichs and Kenneth D Forbus. 2014. X goes first: Teaching simple games through multimodal interaction. *Advances in Cognitive Systems* 3 (2014), 31–46.
- [10] IBM. [n. d.]. *IBM Watson Education and Pearson to Drive Cognitive Learning Experiences for College Students*.
- [11] Robert Iles and Emre Erturk. [n. d.]. What is the vision for AI education resources? How well is it currently being met? ([n. d.]).
- [12] Stanislav Hristov Ivanov. 2016. Will Robots Substitute Teachers? *Browser Download This Paper* (2016).
- [13] Syed Shariyar Murtaza, Paris Lak, Ayse Bener, and Armen Pischdotchian. 2016. How to effectively train IBM Watson: Classroom experience. In *System Sciences (HICSS), 2016 49th Hawaii International Conference on*. IEEE, 1663–1670.
- [14] Alan M Turing. 2009. Computing machinery and intelligence. In *Parsing the Turing Test*. Springer, 23–65.
- [15] Wlodek W Zadrozny, Sean Gallagher, Walid Shalaby, and Adarsh Avadhani. 2015. Simulating ibm watson in the classroom. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education*. ACM, 72–77.