Leverage White-Collar Workers with AI

Stephan Jüngling, Angelin Hofer
FHNW University of Applied Sciences Northwestern Switzerland, School of Business
Peter Merian-Strasse 86, 4052 Basel, Switzerland
stephan.juengling@fhnw.ch, angelin.hofer@students.fhnw.ch

Abstract

While in the manufacturing industry robots do the majority of the assembly tasks, robotics process automation, where software robots are taking over repetitive tasks from humans have been introduced only recently. Many routine tasks continue to be executed without adequate assistance from tools that would be in reach of the current technical capabilities of AI. Using the example of taking meeting minutes, the paper presents some intermediate results of the capabilities and problems of currently available natural language processing systems to automatically record meeting minutes. It further highlights the potential of optimizing the allocation of tasks between humans and machines to take the particular strengths and weaknesses of both into account. In order to combine the functionality of supervised and unsupervised machine learning with rule-based AI or traditionally programmed software components, the capabilities of AI-based system actors need to be incorporated into the system design process as early as possible. Treating AI as actors enables a more effective allocation of tasks upfront, which makes it easier to come up with a hybrid workplace scenario where AI can support humans in doing their work more efficiently.

Introduction

Most physical goods such as cars are predominantly manufactured by industrial robots. The international federation of robotics states that the worldwide robot density is currently 74 robot units per 10'000 employees (IFR, 2018). The allocation of this workforce to blue-collar and white-collar workers is quite different. The automation of tasks from blue-collar workers in manufacturing sites is still much more common than the automation for document centric tasks of white-collar workers in enterprise back-offices. The potential of machine learning and knowledge engineering is huge and a broad variety of different products and services enabled by AI are on the market to support humans with office tasks. The solutions are mostly isolated silo-type of applications and far away from being seamlessly integrated into the white-collar business processes. However, in some business domains such as banking or insurance companies, new technologies such as robotics process automation or the usage of customer facing chat bots gained momentum.

Robotics process automation is increasingly used to automate tedious repetitive tasks where humans currently transfer information from one application front end to the other. Especially in cases, where the backend integration of two systems is too cumbersome or takes too much implementation effort, RPA can be setup to quickly take over. Although, the complexity and stability of the resulting IT software architectures might be questioned, rule-based AI and RPA is increasingly used for process automation. In consequence, there is a fortunate side effect from regulatory requirements (e.g. implementing the four eyes principle, clarify responsibilities in case of failures) resulting in an increasing pressure to establish explicit guidelines for the supervision and governance of AI-driven processes in enterprises and the collaboration between humans and software robots in general.

Human computer interaction patterns

Although the current potential of AI in general and NLP (natural language processing) in particular would allow for more human centricity, we still largely rely on constraints from the past (Jüngling et al., 2018). Most office work and human computer interaction (HCI) are still mainly done with the mouse and the keyboard, which is older than 50 or 150 years respectively. Beyond leveraging HCI with NLP as a more human-centric interface, the system designs in terms of input and output needs to be leveraged according to the fact that AI components learn and become active. AI-driven system design where expert systems, knowledge

 In A. Martin, K. Hinkelmann, A. Gerber, D. Lenat, F. van Harmelen, P. Clark (Eds.), Proceedings of the AAAI 2019 Spring Symposium on Combining Machine Learning with Knowledge Engineering (AAAI-MAKE 2019). Stanford University, Palo Alto, California, USA, March 25-27, 2019.

repositories or deep learning capabilities of AI components are taken into account goes much beyond the conventional type of thinking about HCI. Similar to cooperative robots, so called cobots (Fast-Berglund et. al, 2016), which are designed to collaboratively produce physical goods, supervised or unsupervised machine learning and rule-based system components should be seen as active components in a collaborative workplace and could contribute substantially to the creation of digital goods such as meeting minutes.

Example – Recording Meeting Minutes

Creating meeting minutes consumes time, money and resources. An average Fortune 500 company with around 50000 employees spends 5 Million USD annually on the creation of meeting minutes (IBM, 2018).

Meetings are very important planning and coordination vehicles for organizations in order to mutually communicate and track status information to meet the overall business goals. Meetings that bind *n* resources simultaneously are *n-times* more expensive than individual work. In many cases, at least some of the participants consider it as waste of time, especially if the meetings are too long or unstructured. On the other hand, structured meetings are real time-savers if the right people meet at the right time for the right outcome, and helps to reach a certain objective. In these cases, meetings are very valuable for organizations and need appropriate documentation for later recall and information retrieval. The sooner meeting minutes are available with the relevant information, decisions and action items, the better participants and absent co-workers can start working on the action items. Thus, early distribution of the meeting minutes can boost productivity.

However, recording meeting minutes is very challenging during the meetings as well as time-consuming for later rework and consolidation. While recording, it is difficult to follow the conversation flow because participants speak at a greater pace than it is possible to take notes. Moreover, the minute taker is absorbed and cannot actively participate in the meeting. All of these aspects tend to lead to inaccurate, incomplete and inconsistent meeting minutes. That said, automated recording and writing of meeting minutes would be of great value to enterprises and leads to the following research questions:

- **RQ1**: What problems can be identified in creating meeting minutes?
- **RQ2**: What are the requirements for an information system to create meeting minutes?
- **RQ3**: To what extent can a speech recognition system support the speech to text transcription so that it still contains the relevant information from the meeting?
- **RQ4**: To what extent does a speech recognizer separate multiple voices in a meeting?
- **RQ5**: To what extent does a particular NLP component extract information from the speech-to-text transcript?

Current state of work

As a preliminary result from the master thesis (Hofer, 2018), the following findings are distilled based on observations with the following experimental setup. Several informants, some of them regularly compile meeting minutes as part of their job, were asked to write meeting minutes of a "progress meeting" video (Lauris Beinerts, 2018). The resulting data collection was compared to each other as well as to the reference meeting minutes. Every participant got the same template and the task to capture the most important information such as decisions and action items. The process for automated creation of meeting minutes suggest to capture first and analyze later, the first step is to create a speech-to-text (STT) transcript of the meeting. Two different systems, Otter.AI and Watson STT Service were used to capture the speech (IBM Watson, 2018; Otter.ai, 2018). The automatically created transcripts were then compared to the reference transcript of the video using a plagiarism software (Copyleaks, 2018). Identical parts of each transcript were analyzed and compared to generate the desired result. The accuracy ranged from 87-95% compared to the reference transcript.

- **RQ1:** Main difficulties occur during and after the meeting. During the meeting, the main issue is the speed of the utterance and after the meeting the reconstruction of the completeness of decisions and action items as well as the timely distribution of the minutes. Preliminary results from the feedback of the informants show that over 70% report difficulties to capture the content of the meeting. A majority of 57.1% think that their minutes only partly represent the meeting, while 14.3% fully and 28.6% do not agree that their minutes reflect the content of the meeting.
- **RQ2**: An initial analysis and functional decomposition of the feedback forms led to the following most desired requirements. The particular system should be capable to recognize and separate speech from different participants, transcribe speech to text and extract information. As most important building blocks speech recognition, speech separation, and information extraction were identified. Less desired tasks were the tracking of action items, organizing the meeting minutes and the distribution to the participants.
- **RQ3**: The extent of accurate speech recognition depends on many aspects. In order to create meeting minutes, it is invaluable to have an accurate speech transcript in order to perform later processing steps (e.g. information extraction). In addition, accuracy depends on the

particular STT system, where different products show different performance.

- **RQ4**: Speech separation means to allocate the different text segments to their originating speakers in the transcript. Thus, all speakers need to be recognized. Overlapping speech and multiple sources of speech are critical and well known as the "Cocktail Party" problem (Settle et al., 2018; Yul et al., 2017). Further challenges, such as the loudness and distance of the voices affect their separation. For optimum results of speech separation, multiple microphones are desirable. However, in the test setup, only the mono-audio channel from the video was used and in the preferred application scenario, a single microphone of a mobile device (e.g. smart phone, tablet) would be used as well. Several speech recognizers already have speech separation integrated but usually are limited by recognizing only two to three speakers. Preliminary results prove that it still seems to be very difficult. None of the examined speech recognizers came close to the actual speech segments of the reference transcript.
- **RQ5**: The desired NLP component that accurately extracts action items is still missing. A possible approach to extract relevant information is with "named entity recognition" (NER), but no considerable results have been found so far that would allow automating this task entirely (Goyal, 2018). Reason 8, a smartphone app, claims to extract decisions and action items (Reason8.ai, 2018). Preliminary tests have shown that it seems to be difficult to extract the expected decisions and action items. Moreover, in order to capture a meeting with Reason 8 at least two devices are required.

Leverage human tasks with augmented AI

It not very surprising, that it is currently not possible to generate meeting minutes solely based on AI. However, the task could be decomposed into subtasks that can be allocated to humans and machines in a cobot-like scenario. Both parties could solve those parts of the problem where they are most capable. In the case of taking meeting minutes, humans that manually take notes struggle with the speed of typing the sentences while AI based STT conversion would outperform humans by far. On the other hand, speaker recognition is not a problem for humans while AI components are still inaccurate and fail in our preferred application scenario.

Why current applications are not designed using the potential of both? How could novel hybrid approaches be stimulated upfront? In many cases, requirements engineering and high-level specification of system design start with a UML use case diagrams where all actors and use cases are identified. Even in this early design phase, AI needs to be taken into account. Adding additional AI-system actors in UML use case diagrams would best represent the active role that AI can have and be comparable to the role of humans, such as shown in figure 1.

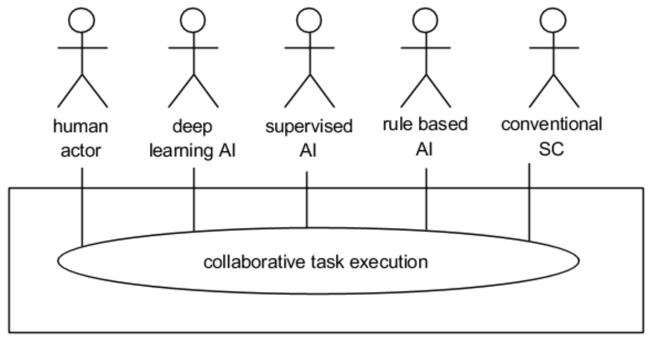

Figure 1 – AI-augmented application scenarios and HCI

In consequence, additional AI based system actors lead to additional swim lanes in UML activity diagrams, such as shown in figure 2.

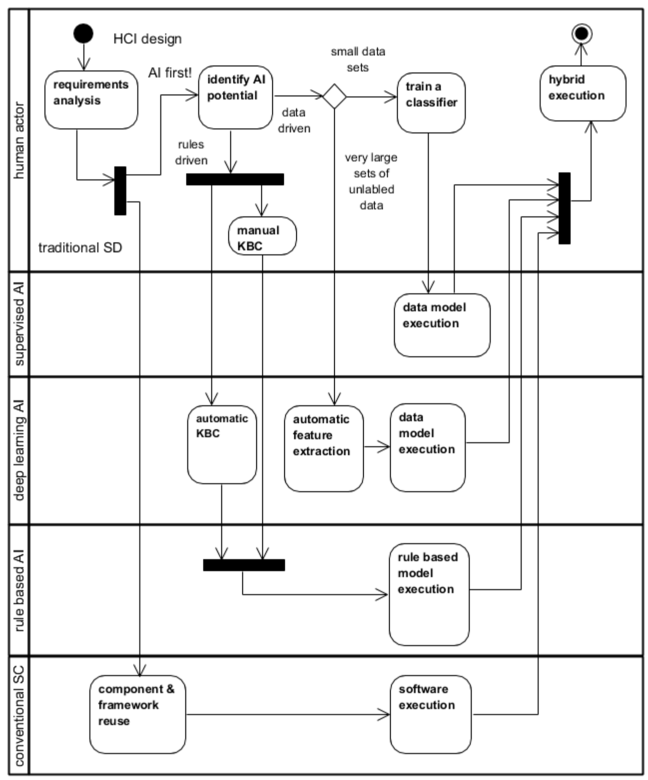

Figure 2 - AI system actors in UML activity diagrams

This allows allocating the different activities explicitly to the different swim lanes and afore-mentioned AI-system actors are responsible to execute their activities. By designing systems in such a way, the mutual collaboration is stimulated and the "AI-first" scenario, which is said to be

the successors of “cloud-first” and “mobile-first” application architectures, comes into play in a hybrid workplace with an AI-augmented system design upfront.
In such a hybrid system design, one can better focus on the different strengths and weaknesses of human and AI actors in order to improve the effectiveness of current IT applications.

In most scenarios, where AI tools are used today, the functionality is embedded in silo-type of applications. Data scientists are using the functionality provided by the graphical user interface (GUI) of their business intelligence tools. In the case of solving data classification problems, experts train the classifiers with the help of tools, where data sets are imported in order to train and apply the models. Knowledge engineers are manually building up ontologies and rule-based semantic models for specific business domains, which are executed in their workbenches. Such scenarios are time consuming and resource intensive and it would be a relief to delegate the knowledge base construction process (KBC) to a deep learning system (Ratner, 2018).

Although many AI services are at hand that can be called by appropriate APIs, they have to be considered as “black-box” logic and are not suitable to combine with company internal business logic, traditional software components (SC) and software design (SD) in cases where the data should stay on premises. It should be the goal to facilitate a seamless way of combining the different components more interactively, and to design more hybrid systems, where the strengths and weaknesses of humans and AI can be allocated more effectively.

In the case of creating meeting minutes, one should construct a GUI, which makes it possible to enable AI actors. For cobots in the manufacturing industry, different methods have been developed to plan the sharing of tasks (Michalos, 2018). Similar attempts should be done for software cobots. Although hybrid physical and AI augmented software co-workspaces might deal with similar problems of orchestration, some of the physical constraints are not present in software. Typical problems of using cobots, such as separating the working areas from robots and humans due to security and safety regulations are not relevant for software. Aspects of ergonomics are replaced by user-friendly GUI design that enables to delegate challenging tasks for humans such as STT transcription quickly and seamlessly to AI actors. If the transcript is generated automatically and visualized as live transcript, the minutes taker could highlight parts with pre-determined tags / buttons for the speaker recognition, action item allocation, decisions detection or automatic text-summarization.

Discussion

The number of RPAs, digital agents and software cobots is still way behind the number of physical robots and cobots in the manufacturing industry. Nevertheless, the role of AI in the digitalization process is increasing and the different business scenarios, where AI algorithms reach from cancer diagnostics in health care over fraud detection in banking to voice-based digital assistants in the consuming sector. However, AI algorithms are rarely seen as active participants in most use cases. The focus of the system design should be leveraged from specifying the requirements of the components at design-time and include the specification of the business scenario at run-time, and how humans can delegate tasks to AI actors. Furthermore, not only the ability to act but also the ability to learn are new features of software components which definitely have an impact on the design of the different run-time business scenarios. Treating AI as system actors can be visualized at the early design phase with UML use case diagrams and changes the way how HCI is perceived. An AI-augmented system design is also visible in UML activity diagrams, where the activities are allocated according to the strengths and weaknesses of the different actors.

In many cases, the current distribution of tasks between humans and AI actors can be optimized, as can be seen in the example of taking meeting minutes. Although an application that autonomously can take accurate meeting minutes is out of reach with the current technology at hand, a hybrid scenario with a GUI that supports the collaboration with a software cobot, tedious tasks such as STT transcription, which is one of the mature components of NLP, could easily be delegated to an AI actor. Even more, the interactions on the GUI could be used as supervised learning for speaker recognition and built into the application itself. Rule-based systems could be established that help to distinguish the different formats of meeting minutes, that could be necessary for different meeting types such as decision meetings, brainstorming sessions or even specialized meetings for a scrum team. In cases where the quality of the speaker recognition is currently insufficient, the system could learn it over time if appropriate AI actors are incorporated at design-time in order to learn during run-time.

Conclusion and Outlook

Compared to blue-collar workers, where the majority of repetitive manufacturing tasks are delegated to robots, many white-collar workers still lack appropriate tool support where at least some of the tasks can be delegated to machines and a more optimized cooperative workplace with humans and AI enabled system actors can be designed. As demonstrated with the example of taking meeting minutes, humans have difficulties with the speed of the utterance while currently available AI based STT transcription systems are much more accurate. On the other hand, speech separation is easy for humans, but none of the two speech recognizers came even close to the actual speech segments, which confirms the well-known “cocktail party” problem. In conclusion, more efficient systems could be designed that take the particular strengths and weaknesses of humans and AI-based software components into account. By starting the system design with use cases where the different AI based components are taken into account as system actors, the capabilities of rule-based AI as well as machine learning can be taken into account explicitly and up-front. Later during the systems design, the activities can be allocated to the most appropriate system actors and the lanes, and the activity diagram can be used as design methodology to optimize the interactive workplace of humans and cobots.

In some application areas such as autonomous driving cars, AI has already the role of an assistive technology. Although AI has the potential to replace human drivers in the long run, the current application scenario is hybrid and consists of a mutual collaboration between human and AI actors. Other than the Turing test, where the focus is towards building a machine that cannot be distinguished from a human being, the goal should be to build cobots with capabilities that can complement those of humans. Building systems that combine the capabilities of traditional software components, knowledge engineering and machine-learning components more seamlessly can help reshaping the traditional HCI in a way where humans can benefit the most. Humans can focus on valuable activities they can accomplish better than machines and benefit from delegating tedious tasks to machines where they are superior.

References

Fast-Berglunda, A.,Palmkvistb, F., Nyqvista, P., Ekereda, S. and Akermana, M. (2016). Evaluating Cobots for Final Assembly. In *6th CIRP Conference on Assembly Technologies and Systems*. p.175-180

Goyal, A. Gupta, V., Kumar and Recent, M. 2018. Named Entity Recognition and Classification techniques: A systematic review. PG Department of Information Technology, GGDSD College, Chandigarh, India

Hofer, A. (2018). Towards a Meeting Minutes Assistant. Master thesis. Master of Science in Business Information Systems, School of Business, University of Applied Sciences Northwestern Switzerland, Olten, Switzerland

IBM Watson. (2018). Speech to Text Demo. Retrieved June 11, 2018, from https://speech-to-text-demo.ng.bluemix.net/

IBM. (2018). Terminuter - automated cognitive meeting minutes. Retrieved May 3, 2018, from https://terminuter-demo.eu-gb.mybluemix.net/

International Federation of Robotics. (2018). Retrieved October 30, 2018, from https://ifr.org/ifr-press-releases/news/robot-density-rises-globally

Jüngling S., Lutz J., Korkut S. and Jäger J. (2018). Innovation Potential for Human Computer Interaction Domains in the Digital Enterprise. In *Dornberger R. (eds) Business Information Systems and Technology 4.0. Studies in Systems, Decision and Control,* vol 141. Springer, Cham

Beinerts L. (2018). The Expert: Progress Meeting (Short Comedy Sketch) - YouTube. Retrieved June 5, 2018, from https://www.youtube.com/watch?v=u8Kt7fRa2Wc

McCann, D. (2016). Robots, robots everywhere. In CFO Magazine (September 15). from http://ww2.cfo.com/applications/2016/09/robots-robots-everywhere/

Michalos G., Spiliotopoulos J., Makris S., and Chryssolourislem G. Solving. 2018. A method for planning human robot shared tasks, In *CIRP Journal of Manufacturing Science and Technology,* https://doi.org/10.1016/J.CIRPJ.2018.05.003

Mohan, M. and Bhat, A. (2018). Joint Goal Human Robot collaboration from Remembering to Inferring, *8th Annual International Conference on Biologically Inspired Cognitive Architectures,* BICA-2017

Otter.ai. (2018). Otter Voice Notes. Retrieved June 11, 2018, from https://otter.ai/login

Ratner, A. and Ré, C. (2018). Knowledge Base Construction in the Machine-learning Era. Queue – Machine Learning. https://doi.org/10.1145/3236386.3243045

Reason8.ai. (2018). reason8 with your meeting partners – for project managers, assistants, business analysts and everyone making notes & follow-ups. Retrieved October 29, 2018, from https://reason8.ai/

Villani, V., Mechatronics. (2018). https://doi.org/10.1016/j.mechatronics.2018.02.009