(De)centralized AI Solutions

Technological and commercial design parameters

Patrick Schmid
ISM Munich, Germany
Patrick.Schmid@ism.de

Matthias Lederer
ISM Munich, Germany
Matthias.Lederer@ism.de

ABSTRACT

Developments concerning artificial intelligence and machine learning have gained a lot of traction recently. Originally thought of as originating from a centralized device like a mainframe computer, a major modification has been introduced in 2017 with the "federated learning" concept performed by decentralized agents. As this constitutes a relatively new development, this research-in-progress contribution addresses the topic of centralized vs. decentralized artificial intelligence in a broader context and in more detail. After a brief introduction of the new concept with a mentioning of accompanying developments, a selection of most relevant technological aspects as well as commercial design parameters to decide on (in particular the smart factory setting) is outlined.

KEYWORDS

Artificial Intelligence, Decentralized Autonomous Organization, Disappearing Internet of Things, Embedded AI Solutions, Federated Learning, Machine Learning, Smart Factory, Softwarization.

1 Introduction and Background

As artificial intelligence (AI) applications typically involve machine learning (ML) routines relying on the processing of big data, the time needed for computational tasks is seen as a major constraint [1]. This applies even more if the learning procedures are performed on a central device like a mainframe computer [2], [3]. One of the main ideas behind the "federated learning" concept that has been introduced in 2017 (the corresponding algorithm already in 2016) is to parallelize computing for machine learning in order to decrease the latency [4].

This development can be put into the context of decentralized autonomous organizations [5] since it is comparable to the fundamental idea of cross-functional working in a BPM (Business Process Management) context. The often mentioned prime example is the parallelized machine learning for image recognition on smart phone agents [4]. This short paper discusses major topics of centralized vs. decentralized AI in more detail and displays further application areas.

There are several accompanying developments that encourage decentralization: for example "softwarization" and "disappearing internet of things" [2]. The idea is roughly that devices are more becoming like terminals as end-nodes of a network, thus spanning a seamless continuum from the devices to the cloud [2]. Here in particular a lot of computation already takes place at the devices, so that computation is moving "toward the edge", hence enforcing Edge and Fog Computing (here Fog Computing covers the gap between Cloud and Edge Computing) [2]. As simple computations can typically evolve into machine learning routines, there seems to be a similar development for decentralized AI solutions in a broader context (AI at the edge instead of simple computations at the edge) and not only in the aforementioned image recognition example. So the decentralized devices can not only be seen as contributors to a consolidated learning procedure, but in fact learning for their own purpose as well as exchanging information and/or synchronizing when it seems useful. In short, AI is no longer purely centralized and also no longer purely a result of one single participant in learning. So, there are two distinct conceptual setups: either several agents contribute to a consolidated learning in a hierarchical structure without substantially learning for themselves (for example when they only perform a limited number of iterations [6]) or several agents synchronize on the same hierarchy level meaning that they produce an aggregated model that is shared between all of them.

2 Relevant technological aspects

There are several relevant (technological) aspects with accompanying challenges that have to be understood and mastered in order to decide on the parameters of central/decentral AI approaches in a meaningful manner:

1. Speed: As already mentioned, multiple agents can be regarded as a kind of parallelized computing mechanism in order to use available time windows more efficiently. This objective is in particular achieved via delegating subtasks in a hierarchical setup to various agents hence improving "virtually" the computing power. Furthermore, as the computational procedures are performed at the local clients at the edge (in the sense of Edge Computing), several data transmissions and information exchanges with the cloud become obsolete [6]. In summary, speed is not only increased by higher computational power due to the number of participating devices, but also by reduction of now unnecessary back-forth transmissions of the training data which is stored locally [6]. Additionally, a reduction of (where applicable) communication costs can occur [6]. So under business process considerations, federated learning can constitute one method to avoid bottlenecks and speed up critical process steps.
2. Quality: Apart from "just" parallelizing tasks for ML procedures, where the volume of the data that has to be processed is very large, the learning of several agents can be used to improve the overall quality of the respective ML routine [6]. Typically, this is done via ensemble methods of machine learning (for instance via "bagging" or "boosting" to name a few examples) [7]. One might question whether quality is strictly improved by consolidating the inputs of various learning agents (as they might provide misleading adjustments). The quality improvement cannot be strictly guaranteed and is dependent on the method used for aggregation, but typically, more input from agents should generate the potential for superior quality in learnings [7]. Bagging and boosting may be the most widely known procedures for this, but of course there exist many different variants with their own nuances / strengths and this topic is very much an area of active research [7]. Detailed analysis is needed what specific algorithmic procedure in what setup best leverages the already gained insights of the single distributed learners (to illustrate the nuances of such a procedure, it shall be mentioned that the newly developed federated learning concept also considers which subparts of the training set are sent to the distributed clients) [4].
3. Compatibility: The agents that participate in hierarchical decentralized learning or same level synchronization must adhere to the same machine learning method / model as a necessary condition in order to facilitate consolidation and exchange of information (for example weight functions). Of course, as there exists a huge variety of machine learning procedures which are all performed differently, these methods need to be aligned. The same applies to questions of implementation, since, for example, in a smart factory various proprietary devices with

underlying architectures and software solutions have to work together [8].

4. Security/Privacy: The security challenges arise when a lot of not fully trustworthy agents contribute to the decentralized learning (either hierarchical or same level) [6]. One can refer here to discussions about security in service-oriented process environments - there are similar challenges (for example, asking for a trustful registry). It could be that information supplied for consolidation is "manipulated" intentionally to distort the quality of learning [6]. Here lies the corresponding challenge of suppressing such occurrences. Apart from that challenge, decentralized learning is commonly regarded as a way to increase privacy since confidential data can be kept locally and only processed (anonymized) information, like weightings of ML routines, is then shared [6].

Considering these technological aspects, it can be discerned that there are several potential advantages which can be gained from the application of decentralized learning scenarios. However, also respective challenges need to be addressed in order to fully realize these potential advantages.

3 Commercial application areas and use cases

After understanding technological aspects, business questions and typical use cases that build on the decentralized learning idea shall be addressed. Generally, decentralized artificial intelligence can be applied to any field where currently artificial intelligence solutions are deployed since there are typically multiple devices in use which are up to now not connected regularly in a hierarchical or same level relationship. In the following, two selected areas are outlined since they play a major role in the ongoing discussion [9] in science and business practice: the “embedded AI solutions” and applications in the smart factory setting.

3.1 Embedded AI Solutions

There is one related thinking approach by C. E. Bouee [10] that deserves special attention as he was the first to address the decentralization issue of AI in full force (in particular with respect to its commercial dimension). He envisions an enforced development and usage of “portable AI” solutions which are decentralized AI solutions that can be integrated (embedded) into all kinds of devices, typically resembling a highly individualized smart assistant [10]. Furthermore, he calls into doubt that the large digital platforms that dominate the tech sector today will dominate the market for portable AI solutions in the future and even believes that portable AI could end their monopoly [10].

He provides several examples for such emerging companies that are active in the segment of embedded AI solutions, thereby illustrating the parameter decision of (de)centralization [10]:

1. *Snips* (from France) attempts to put “AI in every device”, hence creating “a unified voice strategy with a complete range of interface offerings - from simple voice commands to comprehensive natural language voice recognition” [11]. So they are exactly addressing the softwarization topic already described.

2. *Arago* (from Germany) focuses on the B2B sector and its solutions are meant “to automate enterprise IT and business operations” via machines “with human problem-solving skills” that underwent supervised training [12]. Hence, they essentially support the automation of business processes and show the potential of decentralized AI solutions therein.

3. *SenseTime* (from China) is active in many deep learning and supercomputing areas, in particular with its independently developed large-scale training system it also employs decentralized machine learning [13]. Therefore, they act as a kind of orchestrator for decentralized learning.

From this exemplary listing one can already get an idea in what fields entrant companies try to capture market share from the incumbent large tech companies. Foremost, they are supposed to have a competitive advantage over the incumbent companies since they are not burdened with having negative publicity concerning data privacy and security issues. Furthermore, it is conceivable that they gain advantages from more appealing product offerings with higher individualization and “tailor-made” fit [10]. One further argument of Bouee for the competitive edge of these emerging firms is that data becomes obsolete faster and that therefore the advantages which incumbent companies possess from their already amassed data consequently diminishes very quickly [10]. However, on the contrary it can be argued that the large incumbent companies with their already existing wide outreach and computational power do still retain an advantage in this respect since bigger amounts of data are created faster and hence require much more computational power for assessing and evaluating these large volumes. This is certainly the case with respect to potent cloud offerings which heavily rely on the magnitude of computational strength.

The aspect of highly individualized smart assistants is also very interesting, as this is somehow the opposite of a smart assistant provided by a large technology company that could be sometimes prone “to lose its voice” (as featured in a famous commercial) [14].

Highly individualized smart assistants are then AI solutions in their most decentralized form as these solutions do not just complement technical devices, but also every single end user [10]. Bouee elaborates that basically new business models will arise from this as these individualized smart assistants can act as much more strict gatekeepers for personal decisions than common smart assistants currently do [10]. An interesting immediately arising question is, whether companies can and will deliberately set up barriers such that personalized smart assistants cannot be used to full extent (much like the already existing bot-barriers for travel websites).

The topic of decentralization is of particular relevance in a likely “second wave” of digitization. While in the first (B2C marketing-oriented) wave large (U.S.) companies use centrally stored personalized data, manufacturing processes and B2B transactions will then be digitized in a second wave. Here, the tendencies to centralize AI will likely not take place due to already established physical equipment (with various therein embedded software standards) which is difficult to merge. Furthermore, the possible fear of process participants of losing control over their data encourages decentralization.

3.2 Smart Factory

The other large application field that should be discussed in more detail is the smart factory setting which was originally envisioned by the German Ministry of Education and Research (BMBF) [15]. “Smart factory” refers to manufacturing processes (including resources) in which relevant involved machines and devices are all equipped with sensors and perform typical smart tasks, like Condition Monitoring, Diagnosis for Maintenance and Optimization of Processes, that can be improved via machine learning [16]. It is certainly conceivable that an enterprise has multiple smart factories in use that either share the full infrastructure or at least key components. This is then the “factory network” scenario that strives to optimize network performance [17]. This typically addresses the four basic layers of the “Life Cycle Value Stream” of the RAMI 4.0 framework: Asset, Integration, Communication and Information [18].

As data basically is the “lifeblood” of smart factories, an evident approach for maximizing network performance would be to harness the available data of the whole network in a more efficient way by centralizing the decentralized learnings of specific similar tasks.

Here one has to carefully distinguish according to the degree of similarity. To give two extreme examples: Concerning autonomous driving it makes a lot of sense to align and share learnings, since increasing the security of autonomous driving has highest priority and conditions of streets are highly standardized. However, if natural language processing is concerned, it seems recommendable not to mix and interchange training data or learning weights of different dialects within a language (although they do represent the same language) as this rather tends to create confusion instead of resolving it.

In the same manner one has to proceed in the smart factory setting, since there might be tasks and routines that are more similar to each other than to the rest (depending on the type of smart factory). Certainly, such training and synchronization procedures can be applied to the same type of task within the same type of factory. But generalization to similar tasks is, as always, accompanied by certain challenges in deciding whether the similarity is sufficient enough. Here it is conceivable that differences for Condition Monitoring and Diagnosis for Maintenance concern certain types of machines, whereas Optimization of Processes might depend on the level in a multi-level production process. However, if a sufficient similarity can be guaranteed, such (de)centralized learning concepts make a lot of sense. One could argue that different tasks are never similar enough, however, considering that a factory is a highly standardized setting, it is conceivable that there might be a smart transformation/translation of corresponding use cases.

But not only the question “whether” it does make sense, also the question “how often or regular” such a synchronization shall take place has to be answered. In a mimimum scenario, the synchronization can take place in typical inactive periods like production pauses or overnight as not to interfere with the regular business operations (thereby reducing process speed). This is in particular apt for devices and components used in process steps that require fast decisions and actions. However, if devices and components are operating on a much slower time scale (for example registering inbound commodity flows or outbound logistics of finished products) the synchronization could take place continuously. Similar questions are relevant when integrating partners in SCM (Supply Chain Management) or PLM (Product Lifecycle Management) processes. However, in modern value networks which rely on data exchange for smart/reliable decisions, well-known process standards (e.g., EDI, flat XML) may not be sufficient.

According to experts of “Internet of Business” the 10 smart factory trends to watch in 2019 are the following [19]:

1. “Collaborative robots will augment workforces”
2. “Cloud robotics & APIs will give manufacturers greater control”
3. “Robotics-as-a-service will make robotics viable for smaller manufacturers”
4. “5G & Multi-access Edge Computing (MEC) will help keep factory workers informed”
5. “Edge computing will see new use cases”
6. “Cybersecurity will be given greater priority”
7. “AI & advanced analytics will become near-ubiquitous”
8. “Digital twins will be employed more widely across manufacturing and the supply chain”
9. “Additive manufacturing will be used to create final products”
10. “Wearables will become commonplace on the factory floor”

In this list there are two trends that have a particular high relevance for our discussed topic, namely “Edge computing will see new use cases” and “AI & advanced analytics will become near-ubiquitous”. By design, the components and devices of a smart factory are meant to make autonomous decisions. So in time-critical processes it is wanted that more computation is moving towards the edge in order to increase the speed and decrease the latency. Also crucial is the rising importance and spread of AI & advanced analytics, since this first ensures that local AI solutions are present and, furthermore, that there exists potential which can be realized via their alignment. It is certainly highly dependent on the particular use case whether the cooperation of the distributed AI solutions takes place on the same level or in a “master-slave” setup.

But it has to be noted that the aforementioned centralization procedures for AI solutions are not yet to be anticipated in full force for the current year, hence pointing out potentials for realizations that still have to find their place on the respective agendas. The strength of such shared learnings should in particular become apparent in cases where there is a lack of suitable training material, for example when whole processes in smart factory are redesigned according to machine learning insights. Although this would be a desirable feature of the smart factory, this redesign of processes is certainly regarded as a much more complex task than just checking the need for maintenance of a machine since much more factors (like implications for following processes and business partners) have to be considered and evaluated.

4 Conclusion

In this short paper relevant design aspects and implications of connecting formerly unconnected AI solutions with each other were discussed as well as the topic of creating much more agents that are amenable to such setups. The connection of AI solutions can either be understood as delegating tasks in a hierarchical setup or as a kind of synchronization between agents on an equal footing. The potential of such setups that exemplify a much greater extent of interconnection has yet to be realized, but it is expected to be adopted and enforced much more in upcoming developments, for example on the way to the fully integrated smart factory.

REFERENCES

[1] E. Alpaydin. 2014. Introduction to Machine Learning (3rd. ed.). The MIT Press, Cambridge, Chapter 1.

[2] A. Manzalini. 2015. Softwarization and the Disappearing Internet of Things, Retrieved from https://iot.ieee.org/images/files/pdf/softwarization_and_iot _webinar_24022015.pdf.

[3] C. Wurster. 2002. Computers: An Illustrated History. Taschen, Cologne.

[4] Company announcement. 2017. Retrieved from https://ai.googleblog.com/ 2017/04/federated-learning-collaborative.html.

[5] A. Kiulian. 2017. Why Decentralized Artificial Intelligence Will Reinvent The Industry As We Know It. Retrieved from https://www.forbes.com/sites/ forbestechcouncil/2017/11/16/why-decentralized-artificial-intelligence-will-reinvent-the-industry-as-we-know-it/#74f6f300511a.

[6] Mike (@mikepqr). 2018. Federated learning: distributed machine learning with data locality and privacy. Retrieved from https://blog.fastforwardlabs.com/ 2018/11/14/federated-learning.html.

[7] Z. Zhou. 2012. Ensemble Methods: Foundations and Algorithms (1st. ed.). Chapman and Hall/CRC, New York, Chapters 2,3.

[8] A. Abdi, M.R. Abdi, F.D Edalat and A.W. Labib. 2018. Integrated Reconfigurable Manufacturing Systems and Smart Value Chain. Springer, Cham, Chapter 4.

[9] S. Betz, M. Kurz, M. Lederer and W. Schmidt. 2017. Some say Digitalization - others say IT-enabled Process Management thought through to the End. Proceedings of the S-BPM ONE '17 (C. Zehbold and M. Mühlhäuser (Ed.)). ACM Press, New York, NY, USA. DOI: https://doi.org/10.1145 /3040565.3040574

[10] C.-E.Bouee. 2018. Smarter than Man Friday (Think:Act Magazine #24). Retrieved from https://www.rolandberger.com/en/Insights/Global-Topics/ Artificial-Intelligence/

[11] Snips company website. 2019. https://snips.ai/.

[12] Arago company website. 2019. https://arago.co/arago/.

[13] SenseTime company website. 2019. https://www.sensetime.com/.

[14] C. Gartenberg. 2018. Amazon has a clever trick to make sure your Echo doesn’t activate during its Alexa Super Bowl ad. Retrieved from https://www.theverge.com/2018/2/2/16965484/amazon-alexa-super-bowl-ad-activate-frequency-commercial-echo.

[15] German Ministry of Education and Research (BMBF). 2013. Zukunftsprojekt Industrie 4.0. Retrieved from https://www.bmbf.de/de/zukunftsprojekt-industrie-4-0-848.html.

[16] A. Maier, O. Niggemann and S. Schriegel. 2017. Big Data and Machine Learning for the Smart Factory—Solutions for Condition Monitoring, Diagnosis and Optimization. Industrial Internet of Things, 473–485. DOI: https://doi.org/ 10.1007/978-3-319-42559-7

[17] R. Burke, M. Hartigan, S. Laaper, A. Mussomeli and B. Sniderman. 2017. The smart factory (Deloitte Insights). Retrieved from https://www2.deloitte.com/ insights/us/en/focus/ industry-4-0/smart-factory-connected-manufacturing.html.

[18] ZVEI (Zentralverband Elektrotechnik- und Elektronikindustrie e.V.). 2015. Das Referenzarchitekturmodell RAMI 4.0 und die Industrie 4.0-Komponente. Retrieved from https://www.zvei.org/themen/industrie-40/das-referenzarchitekturmodell-rami-40-und-die-industrie-40-komponente/.

[19] A. Hobbs. 2018. Complete guide: 10 smart factory trends to watch in 2019 (Internet of Business), Retrieved from: https://internetofbusiness.com/complete-guide-10-smart-factory-trends- to-watch-in-2019/.