

Maximizing the Potential of Artificial Intelligence to Perform Evaluations in Ungauged Washbowls

Sandesh Achar

Director of Cloud Engineering, Workday Inc., Pleasanton, California, USA

*Corresponding Contact:

Email: sandeshachar26@gmail.com

ABSTRACT

Long short-term memory networks (LSTM) offer precision in the prediction that has never been seen before in ungauged basins. Using k-fold validation, we trained and evaluated several LSTMs in this study on 531 basins from the CAMELS data set. This allowed us to make predictions in basins for which we did not have any training data. The implication is that there is usually sufficient information in available catchment attributes data about similarities and differences between catchment-level rainfall-runoff behaviors to generate out-of-sample simulations that are generally more accurate than current models when operating under ideal (i.e., calibrated) conditions, i.e., when using under idealized conditions. In other words, existing models are generally less accurate when working under idealized conditions than out-of-sample simulations. We found evidence that including physical limits in LSTM models improves simulations, which we believe should be the primary focus of future research on physics-guided artificial intelligence. Putting in place additional physical constraints on the LSTM models.

Key words:

Artificial Intelligence, LSTM, Machine learning, Ungauged Basins

12/31/2020

Source of Support: None, No Conflict of Interest: Declared

This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Attribution-NonCommercial (CC BY-NC) license lets others remix, tweak, and build upon work non-commercially, and although the new works must also acknowledge & be non-commercial.



INTRODUCTION

This technical note presents an ML strategy for PUB, which was developed by this study. In our study, we discovered that LSTMs outperform a conceptual model (SAC-SMA) that was calibrated independently for each catchment, as well as a distributed, process-based model in terms of performance out of sample (NWM). This particular illustration serves two purposes. First and foremost, it is necessary to establish that the current hydrological data record contains sufficient information to allow for reasonable projections in ungauged basins—at the very least some of the time To demonstrate, second, that artificial intelligence

can be a realistic path forward for obtaining this data and, more broadly, for public sector data retrieval. Working with an existing model in mind that, on average, performs as well as the LSTMs we describe here is a good starting point. This study will provide some philosophical and practical considerations for future work that could be undertaken to increase the utility of artificial intelligence in a complex systems science such as hydrology. For the purpose of summarizing our key findings, we may say that, on average, ML outperforms both a lumped conceptual model calibrated in gauged basins and a state of the art distributed process-based model calibrated in ungauged basins (i.e., in more catchments than not). LSTMs and deep learning in general are not intended to be comprehensive analyses of the application of these techniques to PUB; rather, it is intended to highlight early discoveries that may spur further development of these and other techniques.

LITERATURE REVIEW

The vast majority of them use it to evaluate the performance of scam detection, process optimization, and opinion mining algorithms and techniques. Artificial intelligence is a subfield of artificial intelligence that learns by observing and analyzing data and circumstances. By learning from data and identifying significant models with minimal human participation, it allows technology to improve at a given profession with capabilities by learning from data (Achar, 2016). In artificial intelligence, the art of allowing computers to learn from data without being explicitly taught is known as artificial intelligence. Individuals and corporations have become more interested in this data-analytics technique in recent years because it lets them to examine their datasets in a broader and more detailed perspective. Based on an analysis conducted by Forbes, artificial intelligence is utilized by one out of every 10 firms. The vast majority of them use it to evaluate the performance of scam detection, process optimization, and opinion mining algorithms and techniques. Artificial intelligence is a subfield of artificial intelligence that learns by observing and analyzing data and circumstances (Rahman et al., 2019). By learning from data and identifying significant models with minimal human participation, it allows technology to improve at a given profession with capabilities by learning from data. The vast majority of them use it to assess the success of scam detection, process optimization, and sentiment analysis, among other applications. Artificial intelligence is a branch of artificial intelligence that derives knowledge through reasoning and past experiences. Using artificial intelligence, it is possible for technology to improve at a certain profession by learning from data and recognizing essential models with little or no human involvement. The unstructured data includes images, text, video, and audio. Unstructured data can include any data that does not have the operational arrangement required by the procedures to be applied for breakdown and variety, and it can account for up to 80 percent or more of all company data (Achar, 2017). Because it is intended to avoid rigorous rules, the semi-organized data format, which straddles the line between entirely organized and shapeless data, was developed. Extended markup language and other text communication software, for example, are examples of semi-organized data that takes up approximately 5-10 percent of all data on the internet (Addor et al., 2017). Exclusive and acquired data, as well as data generated by the Internet of Things, can all be gathered together. The data generated by social media may be unstructured, making it difficult to use for big data analytics purposes and thus more expensive to use. Data that includes visual and auditory information can also be discovered in organized data, and this information may be important in the event of a disaster. Even though big data and artificial intelligence are closely related, the two concepts are not the same. Artificial intelligence is a machine learning strategy that

uses data to actualize the algorithms' learning approach, which is also known as artificial intelligence (Achar, 2019).

Using a regionally trained long short-term memory (LSTM) network, Pasupuleti & Adusumalli (2018) demonstrated that LSTM-type models can extract information from observable catchment characteristics in order to differentiate between different rainfall-runoff behaviors in hydrologically drained catchments, outperforming basin-specific calibrations of several traditional hydrology models. Ultimately, the purpose of this research is to illustrate how we may use this expertise to forecast in basins that have not been gauged previously. The relative merits of data-driven models vs process-driven models has long been a source of debate in the field of hydrology, and it continues today (Pasupuleti & Adusumalli, 2018). In any circumstance in which an ML model beats a process-based model (although we are unaware of any study that has directly examined this notion), we can conclude that the process-based model is not making full advantage of the input/output data's whole information richness. Those kinds of examples, at the very least, demonstrate that the process-based paradigm has opportunity to be improved (s). When making predictions in ungauged basins, accuracy of out-of-sample forecasts is critical, as is the case in many other situations (PUB). During the period 2003 to 2012, the International Association of Hydrological Sciences (IAHS) designated PUB as its decadal challenge. Regionalization, parameter transfer, catchment similarity, and surrogate basin procedures that are state-of-the-art result in less accurate streamflow predictions than models that are calibrated separately in gauged catchments (Fadziso et al., 2018). For example, current community best practices for PUB are focused on obtaining detailed local knowledge of a specific basin, which is expensive for individual catchments and impossible for large-scale simulations (e.g., continental) like those from the United States National Water Model (NWM) or the stream flow component of the Northern American Land Data Assimilation System. Pasupuleti et al. (2019) also suggested that the calibration of lumped catchment models necessitates the use of at least 2 to 3 years of gauge data (even this is likely an underestimate of the amount of data necessary for reliable model calibration). PUB is still in operation today. Because the vast majority of streams around the world are either ungauged or poorly gauged, and the number of gauged catchments is dwindling, even in the United States, PUB continues to be a significant source of concern for environmentalists (Chen et al., 2018).

RESULTS AND DISCUSSION

The out-of-sample PUB LSTM ensemble surpassed both in-sample benchmarks in more than half of the catchments, with the exception that the basin-calibrated SAC-SMA has a slightly lower average difference between the 95th percentile flows (both SAC-SMA and the PUB-LSTM underestimated peak flows to some extent). In 307 of 531 catchments (58%) the PUB LSTM outperformed the SAC-SMA, and in 347 of 531 (66%) the PUB LSTM outperformed the NWM. The PUB LSTM ensemble had higher mean and maximum NSE scores than the benchmark models, however SAC-SMA outperformed the PUB LSTM in low NSE catchments. The weight optimization method and the random weight initialization of the LSTMs inject some randomness into the training process (we used an ADAM optimizer, Achar, 2018). As a result, LSTM-type models produce better ensemble forecasts. Not every LSTM model trained on the same data would perform poorly in the same catchment. In our instance, N=10 was used (the same size as the SAC-SMA ensemble developed by Ganapathy, 2018, that was used here for benchmarking). In these LSTM models,

randomness rather than systematic model structural error accounts for a large proportion of the uncertainty. The global LSTM model with static catchment features outperforms all other models in our tests. The Global LSTM's performance is compared to other models (SAC-SMA and the Global LSTM without static catchment attributes). It outperforms the Global LSTM without catchment characteristics in most cases. Two things are clear. While static catchment features provide useful information, having them in some catchments really harms us. We examined this association briefly but found no patterns.

The second aspect to highlight from the Global LSTM vs SAC-SMA comparison is that SAC-SMA still has room for improvement. The LSTM finds rainfall-runoff correlations in individual catchments that the SAC-SMA cannot. Its superior performance in some catchments emphasizes the necessity of integrating physical constraints in a hydrological model. In many cases, LSTMs are either overfit or unable to mimic similar catchment behaviors in the training data set. The information supplied in the preceding section can be used to deduce three important points:

- We could make improvements to the process-driven hydrological models that we utilized as benchmarks in our research. When used in conjunction with either the SAC-SMA or the NWM, the LSTM produces a more accurate functional representation of rainfall-runoff dynamics in most catchments.
- It is possible that the notion that process-driven models are favoured is not valid in out-of-sample scenarios. When it comes to retrieving information from large, diverse data sets in a range of hydrological settings, modern artificial intelligence algorithms are highly capable.
- Results of a comparison of models with and without static catchment attributes as inputs reveal that catchment attribute data contains sufficient information to distinguish between unique rainfall-runoff correlations in at least half of the catchments we analyzed in the United States.

CONCLUSION

The purpose of this study is to evaluate the effectiveness of various methods that can forecast low flows as well as flood runoff in catchments that have not been gauged. The objective is to gain insight from the parallels as well as the differences that exist between catchments located in various locations and to understand the differences in performance in terms of the climatic and geographical factors that are at play. Deep learning and artificial intelligence technologies have the ability to synthesize information from a variety of locations and scenarios into a single model, which is why they are so effective for solving problems such as this one. When simulating catchment behavior under nonstationary conditions (such as changing climate), a single LSTM trained to recognize and distinguish different types of hydrological behavior (as shown here) will be able to predict a wider range of conditions under which it will be expected to remain realistic than a model calibrated to past conditions in only one basin. Our prediction is that theory-guided data science will be the most effective technique in the coming years, if not already. The current state of scientific methodology allows for considerable fusions of domain knowledge with artificial intelligence and other methods for learning and predicting directly from data across a wide range of scientific disciplines and applications. Adoption of solutions such as these will be critical in the coming years.

REFERENCES

- Achar, S. (2016). Software as a Service (SaaS) as Cloud Computing: Security and Risk vs. Technological Complexity. *Engineering International*, 4(2), 79–88. <https://doi.org/10.18034/ei.v4i2.633>
- Achar, S. (2017). Asthma Patients' Cloud-Based Health Tracking and Monitoring System in Designed Flashpoint. *Malaysian Journal of Medical and Biological Research*, 4(2), 159–166. <https://doi.org/10.18034/mjmbr.v4i2.648>
- Achar, S. (2018). Security of Accounting Data in Cloud Computing: A Conceptual Review. *Asian Accounting and Auditing Advancement*, 9(1), 60–72. <https://4ajournal.com/article/view/70>
- Achar, S. (2019). Early Consequences Regarding the Impact of Artificial Intelligence on International Trade. *American Journal of Trade and Policy*, 6(3), 119–126. <https://doi.org/10.18034/ajtp.v6i3.634>
- Addor, N., Newman, A., Mizukami, N., & Clark, M. P. (2017). Catchment attributes for large-sample studies. <https://doi.org/10.5065/D6G73C3Q>
- Chen, S., Deming, C., & Adusumalli, H. P. (2018). Safety Assessment of IoT: Warning Scan for Security. *技术与管理回顾*, 1(1), 1–6. Retrieved from <https://xn--jhqs8sh4jbvevnt0xk4h3c.xn--6frz82g/index.php/tmr/article/view/1>
- Fadziso, T., Adusumalli, H. P., & Pasupuleti, M. B. (2018). Cloud of Things and Interworking IoT Platform: Strategy and Execution Overviews. *Asian Journal of Applied Science and Engineering*, 7, 85–92. Retrieved from <https://upright.pub/index.php/ajase/article/view/63>
- Ganapathy, A. (2018). Cascading Cache Layer in Content Management System. *Asian Business Review*, 8(3), 177–182. <https://doi.org/10.18034/abr.v8i3.542>
- Pasupuleti, M. B., & Adusumalli, H. P. (2018). Digital Transformation of the High-Technology Manufacturing: An Overview of Main Blockades. *American Journal of Trade and Policy*, 5(3), 139–142. <https://doi.org/10.18034/ajtp.v5i3.599>
- Pasupuleti, M. B., Miah, M. S., & Adusumalli, H. P. (2019). IoT for Future Technology Augmentation: A Radical Approach. *Engineering International*, 7(2), 105–116. <https://doi.org/10.18034/ei.v7i2.601>
- Rahman, M. M., Pasupuleti, M. B., & Adusumalli, H. P. (2019). Advanced Metering Infrastructure Data: Overviews for the Big Data Framework. *ABC Research Alert*, 7(3), 159–168. <https://doi.org/10.18034/abcra.v7i3.602>

--0--

Issue Link: <https://abc.us.org/ojs/index.php/ei/issue/view/89>