doi.org/10.22498/pages.26.2.72

CSciBox: Artificial intelligence for age-depth modeling

Elizabeth Bradley, T.H. Nelson and L. Rassbach de Vesine

Artificial intelligence (AI) provides major opportunities for scientific analysis. Automated reasoners can explore problem spaces quickly and alert practitioners to possibilities that they had not considered. As a case in point, we describe the CSciBox system. Working with data from a paleorecord, such as ^{14}C dates from a sediment core (Fig. 1a) or ^{18}O values from an ice core, CSciBox produces a set of age-depth models, plus a description of how each one was built and an assessment of its quality.

The AI field has two branches: symbolic methods capture human reasoning in closed form; statistical methods such as neural networks, aka "machine learning" (ML), fit sophisticated models to sets of labeled examples. Both have strengths and weaknesses. ML methods are powerful, but training them requires a large number of examples. This is problematic in the context of age-depth models, where there is rarely more than one published example for each core. The symbolic AI approach has its own challenges: human reasoning is remarkably difficult to capture in formalized, useful ways. However, an AI system seeded with that kind of knowledge can narrate its choices and explain its actions as it solves problems – an absolutely essential feature for a scientific assistant, and one that ML methods cannot provide.

CSciBox marries these two different types of approaches. Its toolbox includes a number of traditional data-analysis methods, along with a set of statistical methods that model the different underlying physical processes (e.g. sediment accumulation). A symbolic AI engine explores the search space of possible age-depth models: choosing among those methods, invoking them on the appropriate data fields with appropriate parameter values, analyzing the results, making appropriate modifications, and iterating until the results match the scientist's physical understanding of the world.

There can be evidence and reasoning both in favor of and against any given age model. CSciBox uses one of the few AI techniques that handle this situation, "argumentation" (Bench-Capon and Dunne 2007), which involves constructing all arguments for and against each candidate age model and then weighing them against one another (Rassbach et al. 2011). In the case of the data in Figure 1(a), CSciBox reasons from the latitude and longitude of the core to choose the IntCal marine 13 curve (Reimer et al. 2013) and the reservoir-age correction (calib.org/marine), then searches for an age-depth model to fit the calibrated, corrected age points. It first tries linear regression but discards the resulting model because the argument against it (large observed residuals) is stronger than those in favor (consistent slope, no reversals). It then tries piecewise-linear interpolation, producing the age model shown in panel C of the figure, but finds that that, too, is a bad solution (low residuals but inconsistent slope and presence of reversals). CSciBox then builds and evaluates an age-depth model using Bacon (Blaauw and Christen 2011), constructing and balancing arguments about the consistency of the slope (good) and the size of the residuals (small) against the fact that Bacon does not converge to a single distribution – as is clear from Figure 1d – and that some of the age points are outside the error bounds.

Like many powerful tools, Bacon's actions are guided by a number of free parameters. CSciBox encodes a number of rules that capture how experts tune those parameter values, which it uses to explore the parameter space and improve the Bacon model. This is a major advance; tools like Bacon are very powerful, but they can be difficult to use. At the end of the exploration process, CSciBox presents the strongest model to the user, together with a full narration of the process involved in building it. CSciBox uses LiPD (McKay and Emile-Geay 2016) to store all of this information (data and metadata), making the analyses completely documented and reproducible, as well as smoothly interoperable with any other LiPD-enabled software. Like LiPD, CSciBox is open-source; see Bradley et al. (2018) for code and documentation.

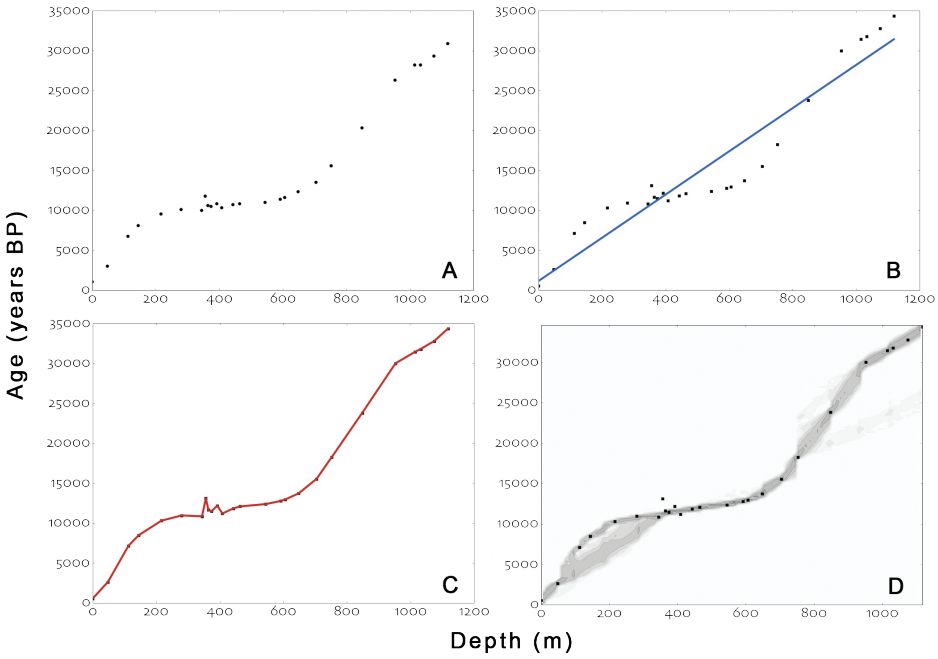

Figure 1: Screenshots of CSciBox building an age model for a marine-sediment core from the Gulf of Mexico (Xie et al. 2012). **(A)** raw ^{14}C ages (•), **(B)** linear regression, **(C)** piecewise-linear interpolation, and **(D)** Bacon model. All plots are age in years BP vs. depth in meters. (■) indicates an age point that has been corrected for reservoir age and undergone a CALIB-style calibration.

ACKNOWLEDGEMENTS

This material is based upon work sponsored by the National Science Foundation (#1245947).

AFFILIATIONS

Department of Computer Science, University of Colorado at Boulder, USA

CONTACT

Elizabeth Bradley: lizb@cs.colorado.edu

REFERENCES

Bench-Capon T, Dunne PE (2007) Artif Intell 171: 619-641

Blaauw M, Christen JA (2011) Bayesian Anal 6: 457-474

Bradley E et al. (2018) CSciBox - building age models of paleorecords. Zenodo

McKay NP, Emile-Geay J (2016) Clim Past 12: 1093-1100

Rassbach L et al. (2011) AI Mag 32: 69-78

Reimer P et al. (2013) Radiocarbon 55: 1869-1887

Xie R et al. (2012) Paleoceanography 27: PA002337

