Technology Assisted Review of Documents

Ashish Prasad, Esq.
Noah Miller, Esq.
Joshua C. Garbarino, Esq.
October 27, 2014

Table of Contents

INTRODUCTION

Recent judicial decisions have encouraged attorneys to consider utilizing the techniques of technology assisted review ("TAR"), and predictive coding in particular, to improve the efficiency and accuracy of the document review phase of discovery. The court in *Da Silva Moore*, speaking through noted e-discovery technology supporter Judge Andrew Peck, gave an opinion validating the defendants' use of predictive coding. *Da Silva Moore v. Publicis Groupe & MSL Group*, 287 F.R.D. 182, (S.D.N.Y 2012). The court stated that "computer-assisted review is an available tool and should be seriously considered for use in large-data-volume cases where it may save the producing party (or both parties) significant amounts of legal fees in document review." Relatedly, in *In re Actos Prods. Liab. Litig.*, the court issued a case management order which set down procedures, upon which both parties agreed, to use predictive coding. *In re Actos (Pioglitazone) Products Liability Litigation*, 2012 WL 7861249 (W.D. La. 2012). This included a "Search Methodology Proof of Concept" for the use of predictive coding. *Id.*, at 3.

This article discusses the workflows and roles that attorneys should utilize when implementing TAR in document review projects.

WHAT IS TAR?

TAR, at its base, utilizes the knowledge of a specialized group of reviewers to train a computer program to analyze data.[1] However, there is conflicting data on the effects of TAR. While according to some sources, TAR lowers costs, and may offer more accuracy of results, other observers have suggested that computers may be less accurate "reviewers" than humans, and may yield little savings in review costs in practice. *See*, for benefits, *Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review*, Grossman and Cormack, Richmond Journal of Law and Technology, Volume XVII, Issue 3 (Spring 2011). *See*, for pitfalls, *Law Technology News: EDI Oracle Study: Humans Are Still Essential in E Discovery*, Bay, Law Technology News (November 20, 2013) http://www.lawtechnologynews.com/id=1202628778400/EDIOracle-Study-Humans-Are-Still-Essential-in-EDiscovery (looking at the results of the EDI-Oracle Study on predic-

tive coding); E-Discovery Tool Relieves But Doesn't End Doc Review Burden, Strickler, Law 360 (April 16, 2014) http://www.law360.com/articles/527585/e-discovery-tool-relieves-but-doesn-t-end-doc-review-burden. As Judge Peck points out in *da Silva Moore*, "computer-assisted review is not a magic, Staples-Easy-Button, solution appropriate for all cases." *da Silva Moore*, at 189.

In general, TAR is most useful in cases with short discovery periods, high dollar values, and mostly textual electronic data. http://www.insidecounsel.com/2013/11/14/using-technology-assisted-review-in-the-right-case. This is because utilizing TAR allows for a quicker analysis of large data sets than typical manual review. Moreover, any upfront, software costs of TAR are defrayed by having less manual review in the later stages of the review project.

TAR WORKFLOWS AND ROLES

The TAR workflow differs significantly from the traditional linear review model. In a typical linear review, reviewers are given groups of documents, generally batched by custodian, and they review them one by one until all documents are coded. In contrast, in a standard TAR workflow, a group of specialized, highly knowledgeable attorneys will code groups of documents ("seed sets"), which will then be used to train the computer program to code the remaining documents. This will continue in successive iterations until confidence in the coding results reaches a certain threshold.

There are typically four main roles in a TAR workflow: lead attorney; subject matter expert (SME); contract attorneys; and project manager. *See* http://www.theediscoveryblog.com/tag/technology-assisted-review/. The lead attorney is ultimately responsible for guiding the review project and making executive decisions such as when to end the review and how to provide additional training to the machine. The SME is an attorney who has a deep knowledge of the case and the issues involved, and is capable of making high-level, substantive coding decisions. The contract attorneys are trained to have knowledge of a specific issue or issues, but do not have the expansive knowledge and decision making ability of the SME. Finally, the project manager with specialized knowledge of the tool and the TAR process must design the workflow, validate, and track the results.

---

[1] The Sedona Conference defines TAR as "[a] process for prioritizing or coding a collection of Electronically Stored Information using a computerized system that harnesses human judgments of subject matter expert(s) on a smaller set of documents and then extrapolates those judgments to the remaining documents in the collection." *The Sedona Conference Glossary E-Discovery & Digital Information Management, 4th Ed (July 2014), at 48.*

As an initial step, SMEs review and code a sample set of documents; this sample set is then used to train the computer to code the larger population of documents. The contract attorneys are used to review various data sets, which can include producible materials, privileged sets, or even documents likely not to be responsive. Generally, it is prudent to have contract attorneys review any data which will eventually be produced, but often there is less necessity of reviewing documents deemed non-responsive by the computer. Reviewers focus on doing high-quality review of a more limited universe of documents than would be reviewed in a liner review. The lead attorney oversees the process and verifies methodology with the project manager.

Predictive Coding Workflows

Predictive coding is perhaps the most recognized incarnation of TAR. It offers a method to systemically codify a large data set through the analysis by human reviewers of a more limited exemplar set of documents, which are then used to code the rest of the data set. While, as noted above, there may be uncertainty among observers on the accuracy of predictive coding versus manual review, it is indisputable that predictive coding can offer large time and cost savings over manual review in the appropriate circumstances.

The initial stages of a predictive coding review are of paramount importance. If a bad system and coding regime is put into place, the knowledge on which the computer bases its decisions will be flawed. This can spell disaster, as these coding decisions will be applied to potentially the entire data set. Having a well-defined methodology is essential to producing accurate coding results.

To ensure a solid foundation for the review, and to provide a point of comparison for the computer output, the SME should create a control set. The control set is kept independent of the seed sets in order to eliminate the influence of the control set on the classifier's model of relevance. It is also used to estimate the percentage of responsive documents ("richness") in the population at large. The SME tags these documents according to relevance standards, and should run analytics on the data to organize the documents conceptually. The control set is then used as the measuring post against which to test the computer results, which will validate the output of the predictive coding tool. In essence,

the control set will evidence how well the computer is doing.

The SME then determines the size of the seed set. The size of the seed set, as well as the method for assembling the seed set, will vary depending on the tool selected and the lead attorney. Higher level reviewers review the seed set normally, and tag based on the coding requirements determined by the lead attorney. The values are fed back into the review tool and a report is created and reviewed to determine the tagging differences between the human coded documents and computer categorized documents. The tool "learns" what an accurately categorized document should look like. This process is repeated until the disagreements between human code and machine code (the "overturn" rate) has an uncategorized rate of 10% or less.

When utilizing contract attorneys, the client and the review provider should ensure that the review team is updated on any changes of protocol. The workflow will vary depending on the technology used, but will generally involve iterative review by the SMEs which is then submitted to the computer. The computer will use the review to classify and identify where it needs more input, and iterations will continue until the computer is sufficiently in agreement with the SMEs. At this point, documents can be categorized by the computer and/or prioritized for a tiered review. Documents that the computer is unable to review (*e.g.*, those that are not electronic) are addressed in a separate workflow, which is frequently linear or sampling-based review.

After processing, the computer expresses a likelihood that documents are responsive or non-responsive, and that information is then used to structure further review. Based on the information provided by the computer, documents are considered for further review, such as to confirm responsiveness, for privilege, for issue coding, or a combination thereof. The computer's output will be used to plan further review in order to maximize efficiency.

Conclusion

It is clear that, in appropriate cases, using predictive coding allows an attorney to review large volumes of data with fewer reviewers than would be required in a linear review process. Predictive coding augurs for increased coding accuracy, increased efficiency, reduced

review personnel, and reduced costs. As the sophistication of both the software and practitioners increases, these gains will become more widespread. Attorneys should be prepared when appropriate to utilize the predictive coding workflows described in this article to achieve these benefits for their clients.

ABOUT THE AUTHORS

ASHISH PRASAD, VICE PRESEIDENT, GENERAL COUNSEL

Ashish Prasad is widely regarded as among the leading experts on discovery in the United States. He has served, among other things, as Litigation Partner, Founder, and Chair of the Mayer Brown LLP Electronic Discovery and Records Management Group, Executive Editor of The Sedona Principles: Best Practices Recommendations & Principles for Addressing Electronic Document Production (2004), Co-Editor in Chief of the Practising Law Institute treatise Electronic Discovery Deskbook: Law and Practice (2009), Adjunct Professor of Law at Northwestern University Law School, Chair of the Defense Research Institute Electronic Discovery Committee, and Chair of the Advisory Council of the National South Asian Bar Association. Ashish has authored over two dozen articles and given over a hundred continuing legal education seminars on topics of electronic discovery before judges, practicing lawyers, and industry groups in the United States, Europe, and Asia. Ashish is a graduate of the University of Chicago Law School, where he was a member of the Law Review, and the University of Michigan, where he graduated with High Honors and High Distinction.

NOAH MILLER, ESQ., STAFFING SPECIALIST

Noah Miller is an attorney who is a Staffing Specialist with Discovery Services LLC. With experience in both large and small firms as well as in document review leadership, Noah offers a unique perspective on document review and electronic discovery in general. Noah holds a B.A. in Political Science and English from the University of Michigan, and a J.D. from the University of Michigan Law School.

JOSHUA C. GARBARINO, ESQ., DIRECTOR OF MANAGED REVIEW

Josh Garbarino manages reviews in the Washington, D.C. office. He has experience in various aspects of e-

discovery, including review, case management, strategic planning, and government investigations. Josh received his J.D. from Syracuse University College of Law, his MA from the University of Illinois, and his BA from Hobart College. He is a member of the National Italian American Bar Association. Josh's interests include anthropology, Siberian husky rescue, and the outdoors.

ABOUT eTERA CONSULTING

eTERA Consulting, Built by the Clients, for the Clients™ is an internationally recognized and award winning leader in data and technology management providing innovative solutions to help corporate, law firm and government clients overcome the high costs of managing large volumes of data, electronic discovery, content searching and operational challenges. As a consultancy, eTERA offers five key services across the Electronic Discovery Reference Model including Early Information Assessment®, Forens1cs One℠, Intell1gent One™, Opt1mum One® and Rev1ew One™. These solutions help clients to proactively identify and analyze data early in the process allowing for significant data reduction, enhanced decision-making abilities, increased efficiencies, compliance, and lower risk resulting in significant savings on eDiscovery costs. eTERA Consulting provides clients with the necessary team of data management, technology, eDiscovery and contract staffing experts needed to effectively manage litigation, government investigations, regulatory oversight and compliance matters. Headquartered in Washington, DC, eTERA Consulting has served the legal vertical since 2004 was selected by the Legal Times in 2013 as the Best Data and Technology Management eDiscovery Provider and recognized by the National Law Journal for three consecutive years as the nation's top end-to-end eDiscovery provider and litigation consulting firm.