



LEXBE ASSISTED REVIEW+

Transparent & Scalable Predictive Coding

Overview

Technology assisted review (TAR) - also called predictive coding, computer assisted review (CAR), and other names - has in recent years been used to allow computer based review of large document sets for production, with human reviewers only reviewing and coding relatively small seed and quality control sets to train and check the computer. As noted by the ABA Section on Litigation: "Technology has created a problem [by building] an overwhelming volume of data that is exponentially more expensive to deal with in litigation" and potentially "technology [can] solve the problem that technology created".¹ That's the promise of Technology Assisted Review. With TAR, a skilled reviewer can train a computer to code an entire set of documents from a relatively small training set.

Lexbe's technology assisted review (Assisted Review+) allows cases with millions of documents and tens of millions of pages to be reviewed for production with expert human reviewers coding only about 5,000 documents, split between a seed set and a control set. Lexbe's Assisted Review+ is based on a 'Bayesian Classifier'², a well-known and accepted classification algorithm that can be effectively applied to large predictive coding tasks.³ Lexbe's transparent approach is in stark contrast to the 'black box' approach of many other eDiscovery service providers that don't disclose the exact algorithm used to generate computer assisted review results. With 'black box' methodologies, the production of responsive documents, and the withholding of non-responsive documents, is done by a secret algorithm that cannot be tested or reviewed for accuracy or completeness. We prefer instead and implement a transparent approach to this important application of computer power to the discovery process.

Assisted Review+ and other assisted algorithms operate by coding a small sampling of a litigation document collection - a 'seed set'. The computer algorithm is given a base set of data it uses to determine text connections and patterns between document content and how the reviewer has coded the document. The assisted review algorithm then uses what it has learned to automatically review the rest of the document collection. Then several control sets of documents are released to a human reviewer or reviewers to confirm or overturn the coding

¹ Predictive Coding, ABA Section of Litigation 2012 Section Annual Conference (April 18-20, 2012).

² Wikipedia, *Bayes' Theorem*, http://en.wikipedia.org/wiki/Bayes%27_theorem (last visited Jan. 22, 2015)

³ Rajiv Maheshwari, *Predictive Coding Guru's Guide* 49-50 (2013)



decisions made by the assisted review algorithm. As the reviewer works through the control sets, an 'F-score' indicating the responsiveness and recall measures of the computer assisted review is generated. Reviewers continue to review the control set until the F-score - shown in ex. 1 of the Assisted Review Report - stabilizes. This will usually be about 2,500 documents.

Lexbe's Assisted Review+ service is available as an Analytics feature within the Lexbe eDiscovery Platform and is alternatively available via HighCapacity Processing+ - Lexbe's stand-alone processing services for cases or projects in other review platforms. Within the Lexbe eDiscovery Platform, Assisted Review+ is applied to documents that have been loaded into a case within your account. Through HighCapacity Processing+, case documents and a coded seed set and control set (5,000 documents total) are transferred to Lexbe by secure FTP (or sent on a portable drive), Assisted Review+ is applied, and an overlay load file containing the assisted review coding designations for the entire document collection is returned.

Technology Assisted Review v. Manual Review

Although manual review (also known as 'linear review') has long been considered an optimal review methodology, *The Sedona Conference Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-Discovery* states as follows:

"[T]here appears to be a myth that manual review by humans of large amounts of information is as accurate and complete as possible – perhaps even perfect – and constitutes the gold standard by which all searches should be measured. Even assuming that the profession had the time and resources to continue to conduct manual review of massive sets of electronic data sets (which it does not), the relative efficacy of that approach versus utilizing newly developed automated methods of review remains very much open to debate."⁴

In fact, the effectiveness of manual review done by attorneys has been studied and found to

⁴ The Sedona Conference, *The Sedona Conference Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-Discovery*, 8 SEDONA CONF. J. 189, 199 (2007) [hereinafter Sedona Search Commentary].

be anything but a 'gold standard'.⁵ This question of whether the outcomes of assisted review or manual review are more accurate or complete -- critical to judicial and professional approval of predictive coding -- has been extensively investigated by the TREC legal track⁶. Even after considering the cost implications of manual review, the TREC study concluded the following:

“Overall, the myth that exhaustive manual review is the most effective – and therefore, the most defensible – approach to document review is strongly refuted.

Technology-assisted review can (and does) yield more accurate results than exhaustive manual review, with much lower effort. Of course, not all technology-assisted reviews (and not all manual reviews) are created equal. The particular processes found to be superior in this study are both interactive, employing a combination of computer and human input. While these processes require the review of orders of magnitude fewer documents than exhaustive manual review, neither entails the naïve application of technology absent human judgment. Future work may address which technology-assisted review process(es) will improve most on manual review, not whether technology assisted review can improve on manual review.”⁷

Key Definitions Used in Technology Assisted Review

Bayes' Law

Bayes Law (or Bayes' Theorem) is based on a Bayesian interpretation of probability⁸ -- the same concept used in the development of artificial intelligence in machines that are able to learn from data inputs they receive. In assisted review, Bayesian Probability models the likelihood of something being true about a document, i.e. responsive, based on the millions of data connections created while coding the seed set. A Naive Bayesian Classifier⁹ is a probability model with assumptions that allow for pattern recognition among multiple

⁵ Maura R. Grossman & Gordon V. Cormack, *Technology-Assisted Review in E-Discovery Can Be More Effective And More Efficient Than Exhaustive Manual Review* Richmond Journal of Law & Technology, 2011

⁶ Bruce Hedin *Overview of TREC 2009 Legal Track*, Text Retrieval Conference, 2009

⁷ Maura R. Grossman & Gordon V. Cormack, *Technology-Assisted Review in E-Discovery Can Be More Effective And More Efficient Than Exhaustive Manual Review*, Richmond Journal of Law & Technology, 2011

⁸ University of Utah College of Engineering, *Computational Statistics* (2013)

⁹ Wikipedia, *Naive Bayes Classifier* http://en.wikipedia.org/wiki/Naive_Bayes_classifier (last visited Jan. 22, 2015).

independent variables.¹⁰

Seed Set

Assisted review algorithms are not able to perform their function without being properly trained. The coding performed in seed set training and subsequent control set coding is critical because it identifies the patterns and connections between responsive documents. All assisted review algorithms are highly sensitive to coding errors and mistakes. Assisted review relies on all words in the documents, not just keyword hits. Coding the seed set follows the adage of 'garbage in garbage out'. That is, rushing the coding of the seed set and giving poor input data will inevitably result in a less than optimal assisted review performance. On the other hand, a high quality coding of the seed set by experienced reviewers will produce the most accurate assisted review results.

The seed set should be comprised of around 2,500 documents. There are two schools of thought as to how the seed set should be generated. The first approach is that the seed set should be randomly generated to minimize the risk of creating a seed set that is not a representative sample of the entire document collection. The 'random generation' approach, if used, works best when there is a fairly even distribution of responsive and non-responsive documents in the seed set to provide the algorithm with enough information to code both types of documents in the entire set. A second approach is to use keyword searches or document groupings or classifications to generate an equal sampling of responsive and non-responsive documents in the seed set. This evenly distributed seed set is then used to generate responsive coding for the entire document set. This second assisted review approach will generate a different result than merely keyword searches against the entire document collection. Lexbe Assisted Review+ allows a flexible workflow and integrates with either approach.

Control Set

The control sets serve as a quality control function of the assisted review process.¹¹ When setting up the assisted review job within the Lexbe eDiscovery Platform, the reviewer

¹⁰ Dupré, Tipler, *New Axioms For Bayesian Probability, Bayesian Analysis* 599-606 (2009).

¹¹ Victor Stanley, Inc. v. Creative Pipe, Inc., No. MJG-06-2662 (D. Md. Sept. 9, 2010).

determines the size of each control set. The reviewer also either confirms or overturns how the assisted review algorithm has coded the documents in the control sets. (i.e. the reviewer changes a document identified as responsive to non-responsive and vice-versa). Overturning a document that has been coded as 'non-responsive' negatively affects the recall element of the F-score.

Reviewers should continue to review control sets until there is a stabilization in the displayed F-curve (ex. 1 in the assisted review report). Usually this will require review of about 2,500 documents in the control set. Stabilization in the F-curve is an indication that the metrics used to evaluate assisted review have also stabilized and will likely be unaffected by continued review of control sets. Continuing to review control sets will serve only to reduce the margin of error associated with the F-score.

F-Score

F-scores¹² are determined by considering the precision and recall of the assisted review algorithm.¹³ Precision¹⁴ is a measure of how often the algorithm accurately predicts a document to be responsive, that is the percentage of produced documents that are actually responsive. Recall¹⁵ is a measure of what percentage of the responsive documents in a data set have been classified correctly by the algorithm. A low precision score indicates an abundance of false positive identifications, or over-delivery. But, a high precision score does not mean that all the responsive documents have been found. A low recall score is an indication of under-delivery and a high recall score shows the percentage of responsive documents that have been accurately identified. Precision can also be seen as a measure of relevance (a high precision score indicates a low rate of false positive coding), whereas recall is a measure of what percentage of all of the actually existing responsive documents in the set were accurately coded (a high recall score is an indication of a low rate of false negative coding).

¹² Wikipedia *F1 Score* http://en.wikipedia.org/wiki/F1_score (last visited Jan. 22, 2015)

¹³ C.J. Van Rijsbergen *Information Retrieval* (2nd ed. 1979).

¹⁴ Wikipedia *Precision and Recall* http://en.wikipedia.org/wiki/Precision_and_recall (last visited Jan. 22, 2015)

¹⁵ id.

Neither the American Bar Association¹⁶ nor the influential Sedona Conference¹⁷ have suggested that F-scores should be used in determining defensibility of technology assisted review methods. The Association of Certified E-Discovery Specialists has stated that the F1 score is not to be interpreted as a measure of review quality¹⁸ but rather is an indication of a) how well the case lends itself to TAR and b) the quality of the seed set training. More important than the nominal F score is the stability (slope) of the F-curve.

Margin of Error

The margin of error¹⁹ is a statistical measure of uncertainty based on the possibility that the data sampled was not an accurate representation of the entire data set, assuming a normal distribution of documents. As the amount of data sampled increases, the margin of error is reduced. In assisted review, the margin of error decreases as more control sets are reviewed to verify that the algorithm correctly coded the documents. The margin of error should be interpreted along with the final F-score. For example, if there is a final F-score of 0.75 and a margin of error of $\pm 5\%$, then there is 95% certainty that the harmonic mean²⁰ of the recall and precision in this instance of assisted review is between 0.7 and 0.8.

Judicial Approval of Assisted Review

Court cases have increasingly considered and approved use of various forms of computer assisted review. The following is an illustrative sampling.

Judge Paul Grimm's hallmark opinion in Victor Stanley, Inc. v. Creative Pipe Inc.²¹ was examined by the ABA Section of Litigation in 2012 with regard to the defensibility of TAR

¹⁶American Bar Association *Predictive Coding Written Materials* http://www.americanbar.org/content/dam/aba/administrative/litigation/materials/sac_2012/14-1_predictive_coding_writen_materials_authcheckdam.pdf (last visited Jan. 22, 2015)

¹⁷ ("Defensible" By What Standard?", Hon. Craig B. Shaffer U.S. Magistrate Judge, District of Colorado Denver, CO. Sedona Conference, 2012)

¹⁸ Association of Certified eDiscovery Specialists *Measuring Predictive Coding Performance and Why the F1 Score is Virtually Worthless* <http://www.aceds.org/measuring-predictive-coding-performance-and-why-the-f1-score-is-virtually-worthless> (last visited Jan. 22, 2015)

¹⁹ Wikipedia *Margin of Error* http://en.wikipedia.org/wiki/Margin_of_error (last visited Jan. 22, 2015)

²⁰ Wikipedia *Harmonic Mean* http://en.wikipedia.org/wiki/Harmonic_mean (last visited Jan. 22, 2015)

²¹ *Victor Stanley, Inc. v. Creative Pipe, Inc.*, No. MJG-06-2662 (D. Md. Sept. 9, 2010).

methods to affirm Judge Grimm’s conclusion that “either (a) the parties collaborate and agree to the process and technology or (b) a single party creates a defensible process with thoughtful input on the front end, iterative improvement in the middle, and a test for over-inclusiveness and under-inclusiveness at the end.”²²

In The New Mexico State Investment Council v Bland²³, Judge Sarah M. Singleton found that plaintiff firm Day Pitney’s use of “machine learning tools such as predictive coding, concept grouping, near-duplication detection and e-mail threading . . . enabled the reviewers on the document analysis teams to work more efficiently with the documents and identify potentially relevant information with greater accuracy than standard linear review.”²⁴

In Federal Housing Finance Agency v. HSBC North America Holdings, Inc., Judge Denise Cote allowed defendants “to produce its documents through the use of predictive coding. The literature that the Court reviewed at that time indicated that predictive coding had a better track record in the production of responsive documents than human review”²⁵

In Dynamo Holdings Ltd. Partnership v. Commissioner of Internal Revenue, Judge Ronald Buch stated that “our discovery Rules, are to ‘be construed to secure the just, speedy, and inexpensive determination of every case.’ Rule 1(d). Petitioners may use predictive coding in responding to respondent’s discovery request.”²⁶

Lexbe Assisted Review+ Outcomes Compared

The following exhibits compare the results of Lexbe’s Assisted Review+ with other predictive coding or computer assisted review services or algorithms in the market using the results of a 2009 TREC study.²⁷ The results include a precision, recall, and overall F-score. Both the TREC study figures and Lexbe results were found by running assisted

²² ABA Section of Litigation *Predictive Coding* 7 (2012).

²³ The New Mexico State Investment Council v. Bland, No. D-101-cv-2011-01534, 2014 WL 772860 (N.M. Dist. Feb. 12, 2014)

²⁴ The New Mexico State Investment Council v. Bland, No. D-101-cv-2011-01534, 2014 WL 772860 (N.M. Dist. Feb. 12, 2014) ¶ 32

²⁵ Federal Housing Finance Agency v HSBC North America Holdings, Inc., No. 11 Civ. 6189 (DLC) p. 2 (S.D.N.Y. 2014).


²⁶ Dynamo Holdings Ltd. Partnership v. Commissioner of Internal Revenue, Docket No. 2685-11 p.19 T.C. 149 (2014)

²⁷ Bruce Hedin *Overview of TREC 2009 Legal Track*, Text Retrieval Conference, 2009

review/predictive coding methodologies on the EDRM Enron data set using “fantasy football” as the responsive/non-responsive criteria.²⁸ The EDRM Enron Data Set is an industry-standard collection of email data that is used frequently for electronic discovery performance testing and training. The EDRM Enron data set is maintained by the EDRM Data Set Project²⁹ and “provides industry-standard, reference data sets of electronically stored information (ESI) and software files that can be used to test various aspects of e-discovery software and services.”

Lexbe’s predictive coding results were favorable to others tested in the TREC Report, with Lexbe’s methodology producing precision and recall scores of .822 and .902, respectively and an overall F1 score of (0.860).

Ex. 1

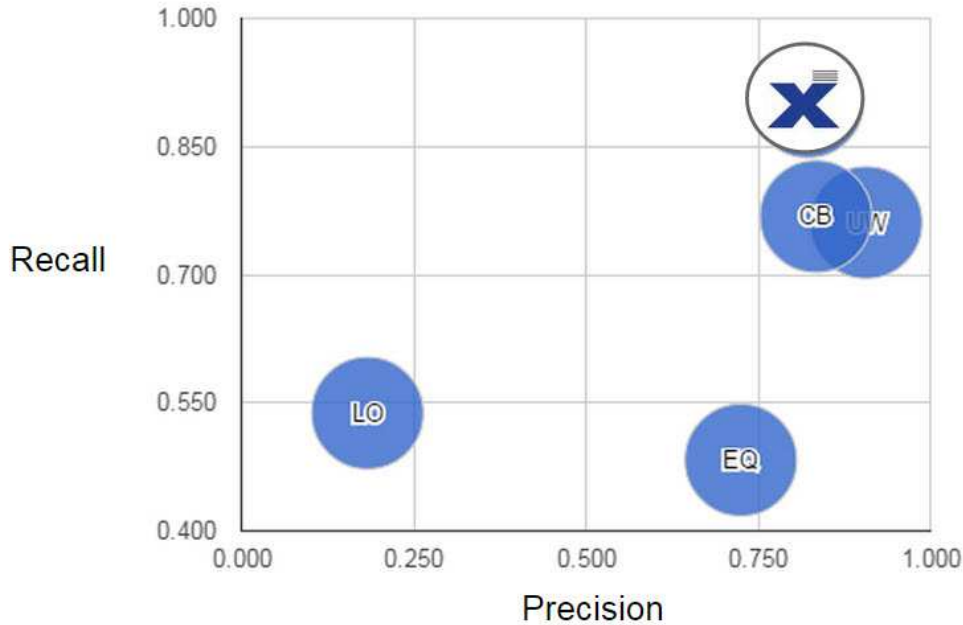
		Precision	Recall	F1
Lexbe		0.902	0.822	0.860
University of Waterloo	UW	0.761	0.907	0.828
Cleary Gottlieb Steen & Hamilton + Backstop LLP	CB	0.768	0.834	0.799
Equivio	EQ	0.483	0.725	0.550
Logik	LO	0.538	0.183	0.273

²⁸ id. (Topic 207)

²⁹ Electronic Discovery Reference Model *Enron Dataset* <http://www.edrm.net/projects/dataset> (last visited Jan. 22, 2015)

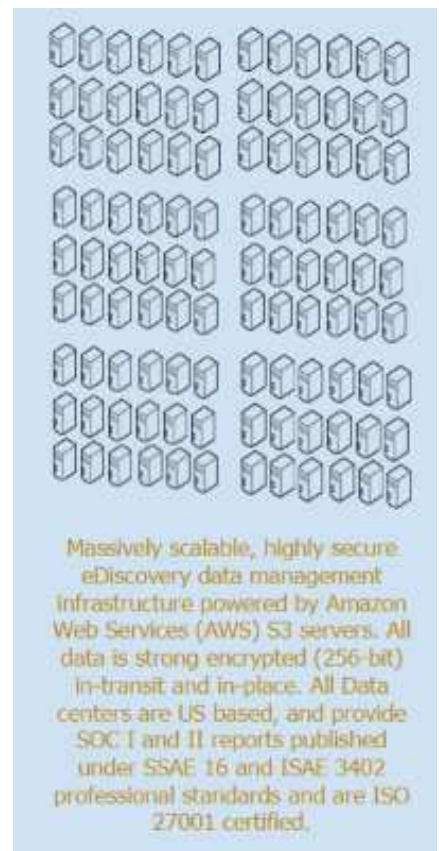
Ex. 2

Lexbe vs. Competitors



The Lexbe Engine

The speed and performance of Assisted Review+ is driven by our massively scalable server infrastructure back-end, The Lexbe Engine, hosted by Amazon Web Services (AWS). All data is strong encrypted (256-bit) in-transit and in-place. All Data centers used are US-based, and provide SOC I and II reports published under SSAE 16 and ISAE 3402 professional standards and are ISO 27001 certified. Whether accessing Assisted Review+ via Lexbe eDiscovery Platform or HighCapacity Processing+, this scalable back-end configuration allows rapid deployment of as many servers as is needed to complete even the largest predictive coding projects within short discovery deadlines.





eDiscovery Consulting Services

Lexbe offers eDiscovery Consulting Services to help plan and execute an assisted review project, including educating parties on the benefits of computer assisted review, preparing language permitting assisted review for ESI Orders, assisting in Rule 26/Meet and Confer activities, and expert testimony on assisted review in general or our procedures used in particular. Please contact our sales department if we can be helpful on a project or case.

Summary

Technology Assisted Review increasingly has become an accepted tool that can be helpful in coding large document collections for production. When used properly it can do as well as -- or better -- than human reviewers. As document collections involved in litigation continue to explode in size, TAR can be an invaluable asset in the litigator's arsenal to enable speedy and cost-effective document review and production. Lexbe's Assisted Review+ is available for use in cases hosted in the Lexbe eDiscovery Platform, and can also be applied to cases being hosted in other review platforms through HighCapacity Processing+. In either approach, Assisted Review+ offers a transparent, defensible, and fast predictive coding workflow, powered by the massively scalable Lexbe Engine. Please contact sales at 800-433-7809 x22 or at sales@lexbe.com if we can provide a quote to use Lexbe's Assisted Review+ to a project.