Spá: a web-based viewer for text mining in Evidence Based Medicine

Kuiper, J[1]., Marshall, I.J.[2], Wallace, B.C.[3], and Swertz, M.A.[1]

[1] University of Groningen P.O. Box 30001, 9700 RB Groningen
{joel.kuiper,m.a.swertz}@rug.nl
[2] King's College London, London SE1 3QD, UK
iain.marshall@kcl.ac.uk
[3] Brown University, Providence, RI 02906, USA
byron_wallace@brown.edu

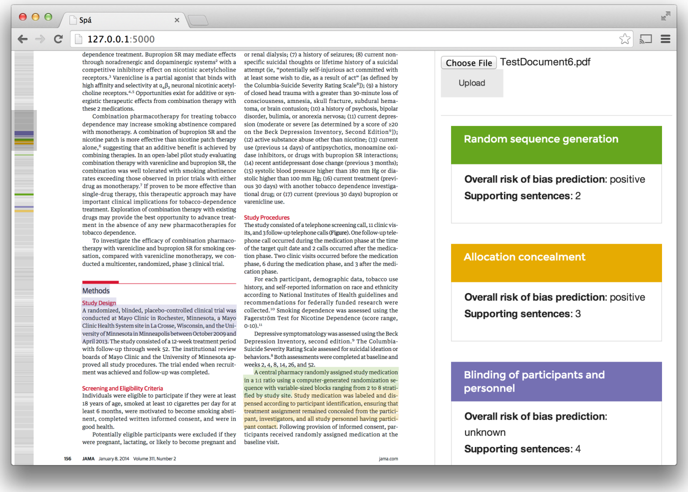
Abstract. Unstructured PDF documents remain the main vehicle for dissemination of scientific findings. Those interested in gathering and assimilating data must therefore manually peruse published articles and extract from these the elements of interest. Evidence-based Medicine provides a compelling illustration of this: many person-hours are spent each year extracting summary information from articles that describe clinical trials. This represents an enormous burden, especially in light of the expotentially increasing volume of published biomedical articles. Machine learning provides a potential means of mitigating this burden by automating extraction. But for automated approaches to be useful to end-users, we need tools that allow domain experts to interact with, and benefit from, model predictions. To this end, we present a web-based tool called Spá[4] that accepts as input an article and provides as output an automatically visually annotated rendering of this article. More generally, Spá provides a framework for visualizing predicted full-text PDF annotations, both at the document and sentence level.

1 Introduction

Imposing structure on full-text documents (e.g., identifying specific sentences of interest) is an important, challenging and practical task in natural language processing and machine learning. Consider *systematic reviews* – fundamental tools in Evidence-based Medicine (EBM) [7] – which aim to address specific clinical questions by identifying and synthesizing data from all relevant published articles. When doing this one often needs, for example, to assess the *risk of bias* for a particular study across different domains, such as bias due to improper blinding of participants and personnel. For this task, one wants to make a summary judgement (e.g., low risk of bias) and simultaneously extract the sentences supporting that judgement.

[4] Source code available under GPLv3 at https://github.com/joelkuiper/spa [3]; demo available at http://clinici.co/

Extracting such information from the unstructured text of clinical trial articles is a laborious process. Machine learning methods provide the machinery to automate such extractions; as they can effectively impose the desired structure onto PDFs. For example, we can train a classifier to automatically assess the risk of bias across different domains and simulteanously identify the sentences supporting these assessments. We may also train a model that extracts the sample size from a paper describing a clinical trial. But if such technologies are to be practically useful, we need tools that visualize these model predictions and annotations. Here we describe Spá, which aspires to realize this aim. Spá is an open-source, web-based tool that can incorporate state-of-the-art machine learning predictors to automatically annotate PDF articles. As a practical (and useful) demonstration of this technology, we have built a machine learning system that automatically annotates PDFs to facilitate EBM. Specifically, this system leverages a multi-task model that simulteanously assesses the risk of biases across several domains, and extracts sample sizes from articles. This tool is unique in that it leverages state-of-the-art ML models applied to full-text articles to assist practitioners of EBM (and other biomedical researchers).



dependence treatment. Bupropion SR may mediate effects through noradrenergic and dopaminergic systems[2] with a competitive inhibitory effect on nicotinic acetylcholine receptors.[3] Varenicline is a partial agonist that binds with high affinity and selectivity at $\alpha_4\beta_2$ neuronal nicotinic acetylcholine receptors.[4,5] Opportunities exist for additive or synergistic therapeutic effects from combination therapy with these 2 medications.

Combination pharmacotherapy for treating tobacco dependence may increase smoking abstinence compared with monotherapy. A combination of bupropion SR and the nicotine patch is more effective than nicotine patch therapy alone,[6] suggesting that an additive benefit is achieved by combining therapies. In an open-label pilot study evaluating combination therapy with varenicline and bupropion SR, the combination was well tolerated with smoking abstinence rates exceeding those observed in prior trials with either drug as monotherapy.[7] If proven to be more effective than single-drug therapy, this therapeutic approach may have important clinical implications for tobacco-dependence treatment. Exploration of combination therapy with existing drugs may provide the best opportunity to advance treatment in the absence of any new pharmacotherapies for tobacco dependence.

To investigate the efficacy of combination pharmacotherapy with varenicline and bupropion SR for smoking cessation, compared with varenicline monotherapy, we conducted a multicenter, randomized, phase 3 clinical trial.

Methods

Study Design

A randomized, blinded, placebo-controlled clinical trial was conducted at Mayo Clinic in Rochester, Minnesota, a Mayo Clinic Health System site in La Crosse, Wisconsin, and the University of Minnesota in Minneapolis between October 2009 and April 2013. The study consisted of a 12-week treatment period with follow-up through week 52. The institutional review boards of Mayo Clinic and the University of Minnesota approved all study procedures. The trial ended when recruitment was achieved and follow-up was completed.

Screening and Eligibility Criteria

Individuals were eligible to participate if they were at least 18 years of age, smoked at least 10 cigarettes per day for at least 6 months, were motivated to become smoking abstinent, completed written informed consent, and were in good health.

Potentially eligible participants were excluded if they were pregnant, lactating, or likely to become pregnant and or renal dialysis; (7) a history of seizures; (8) current nonspecific suicidal thoughts or lifetime history of a suicidal attempt (ie, "potentially self-injurious act committed with at least some wish to die, as a result of act" [as defined by the Columbia-Suicide Severity Rating Scale[8]]); (9) a history of closed head trauma with a greater than 30-minute loss of consciousness, amnesia, skull fracture, subdural hematoma, or brain contusion; (10) a history of psychosis, bipolar disorder, bulimia, or anorexia nervosa; (11) current depression (moderate or severe [as determined by a score of ≥20 on the Beck Depression Inventory, Second Edition[9]]); (12) active substance abuse other than nicotine; (13) current use (previous 14 days) of antipsychotics, monoamine oxidase inhibitors, or drugs with bupropion SR interactions; (14) recent antidepressant dose change (previous 3 months); (15) systolic blood pressure higher than 180 mm Hg or diastolic higher than 100 mm Hg; (16) current treatment (previous 30 days) with another tobacco dependence investigational drug; or (17) current (previous 30 days) bupropion or varenicline use.

Study Procedures

The study consisted of a telephone screening call, 11 clinic visits, and 3 follow-up telephone calls (**Figure**). One follow-up telephone call occurred during the medication phase at the time of the target quit date and 2 calls occurred after the medication phase. Two clinic visits occurred before the medication phase, 6 during the medication phase, and 3 after the medication phase.

For each participant, demographic data, tobacco use history, and self-reported information on race and ethnicity according to National Institutes of Health guidelines and recommendations for federally funded research were collected.[10] Smoking dependence was assessed using the Fagerström Test for Nicotine Dependence (score range, 0-10).[11]

Depressive symptomatology was assessed using the Beck Depression Inventory, second edition.[9] The Columbia-Suicide Severity Rating Scale assessed for suicidal ideation or behaviors.[8] Both assessments were completed at baseline and weeks 2, 4, 8, 14, 26, and 52.

A central pharmacy randomly assigned study medication in a 1:1 ratio using a computer-generated randomization sequence with variable-sized blocks ranging from 2 to 8 stratified by study site. Study medication was labeled and dispensed according to participant identification, ensuring that treatment assignment remained concealed from the participant, investigators, and all study personnel having participant contact. Following provision of informed consent, participants received randomly assigned medication at the baseline visit.

156 JAMA January 8, 2014 Volume 311, Number 2 jama.com

Random sequence generation

Overall risk of bias prediction: positive
Supporting sentences: 2

Allocation concealment

Overall risk of bias prediction: positive
Supporting sentences: 3

Blinding of participants and personnel

Overall risk of bias prediction: unknown
Supporting sentences: 4

Fig. 1. Screenshot of a PDF with highlighted risk of bias.

While our application of interest is EBM, we emphasize that the visualization tool can be used for any domain in which one wants to annotate PDFs. Thus the contribution of this work is two-fold, as we present: (1) a practical tool (Spá) that incorporates cutting edge machine learning to help researchers rapidly assess the risk of biases (and sample sizes) in published biomedical articles, and, (2) a general open-source web tool for visualizing the predictions of trained models from full-text articles. These contributions are described further in sections 2 and 3, respectively.

2 Automating Evidence-based Medicine

2.1 Systematic Reviews

The process of pooling and summarizing clinical trials is called *systematic reviewing*, and forms the corner-stone of current EBM. Systematic reviewing consists of specifying a precise clinical question and accompanying study inclusion criteria, and then exhaustively searching the literature to identify eligible studies. Identified eligible studies are subsequently analyzed (i.e., synthesized), resulting in a summary of all of the published evidence that pertains to the clinical question of interest. Systematic reviews thus represent a rigorous, data-driven approach to answering clinical questions. But conducting systematic reviews is complicated by the massive number of trials that are conducted: for example, the Cochrane Library alone indexes 286,418 trials that were conducted in the last decade [8]. While methods and publishing standards are improving, many legacy publications will remain available only as PDF documents.

Extracting information of interest from these PDFs is imperative but laborious. To mitigate this burden, we have constructed machine learning models that automatically extract some characteristics of interest. Specifically, our model predicts *risk of biases* across different domains as a function of the texts comprising articles. It simultaneously extracts sentences supporting these predictions. We briefly outline our approach to achieving this below. We also note that we have a a separate model that extracts the study sample size from articles. Spá – our visualization tool – provides a framework to upload PDFs and then render these predicted annotations. We foresee adding additional predictors to this tool that, for example, may identify different treatment mentions within the text (e.g., drugs).

2.2 Machine Learning Approaches

We briefly describe our model for assessing the study risk of bias (and supporting sentences) across the following domains: random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data and selective reporting. To train our model, we have leveraged the Cochrane Database of Systematic Reviews (CDSR) in lieu of manually annotated data, which would be expensive to collect. The CDSR contains descriptions and data about clinical trials reported in existing systematic reviews. Briefly, we match (the full-texts of) studies to entries in the CDSR, which contains risk of bias assessments; this provides document level labels. The CDSR also contains sentences that annotators indicated as supporting their assessments. We match these strings to substrings in the PDFs to provide sentence-level supervision. This can thus be viewed as a *distantly supervised* [5,6] approach.

Concerning the risk of bias predictions, we have two tasks for a given article: (1) predict the overall risk of bias for each of the domains of interest, and, (2) extract the sentences that support these assessments. In both cases we leverage standard bag-of-words text encoding and linear-kernel support vector machines

(SVMs). Because the risk of bias predictions are correlated (across domains), we take a *multi-task* [2] approach to classification and jointly learn a model for the domains. We accomplish this by way of a feature space construction that includes both shared and domain-specific terms (similar to the domain adaptation approach in [1]).

With respect to the second task of sentence extraction, we would like to *jointly* assess the risk of bias associated with a given article *and* extract the sentence that supports this judgement. To this end, we first make sentence level predictions (using one set of trained models), and then insert features representing the tokens (words) in the predicted sentences for exploitation by the document level classifier. Due to space constraints here, we omit further specifics of our approach, but we provide further details elsewhere [4]. Figure 1 shows example system output: the overall risk of bias with respect to random sequence generation is judged to be low, and supporting sentence

3 Spá Architecture Overview

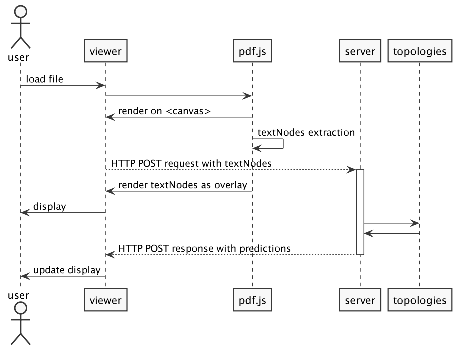


Fig. 2. Sequence diagram of a typical request-response in Spá.

Spá relies on Mozilla pdf.js[5] for visualization of the document and text extraction. The results of the text extraction are processed server-side by a variety of processing topologies, as outlined in figure 2. Results, which could come from complicated machine learning systems, are communicated back to the browser and displayed using React components[6].

For each of the annotations the relevant nodes in the document are highlighted and a custom scrollbar, inspired by substance.io[7], that acts as a mini-map is projected to show where it resides within the document. The user can interactively activate and inspect specific results.

4 Future work

We present a web-based tool for interactive visualization of annotations and meta-data on PDF documents. This allows users to see the results from machine learning predictions for a specific document. We have demonstrated the utility of this system within the context of EBM by automatically extracting potential risks of bias (and supporting sentences) via state-of-the-art ML methods.

[5] http://mozilla.github.io/pdf.js

[6] http://facebook.github.io/react

[7] http://substance.io/beta/

More generally, we believe the tool to be potentially useful for a much wider range of text mining and machine learning applications. To increase the generality of the tool we are developing a pluggable system for processing topologies, allowing developers to quickly plug in new systems for automated PDF annotation. Furthermore, work is being done on allowing users to save selected annotations, possibly embedded within the document itself, for sharing and off-line use. The ultimate vision is to have a extensible system for semi-automated (machine assisted) screening, data extraction and data summarization for EBM, and to allow rapid development of similar systems for other domains.

Acknowledgments Part of this research was funded by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 261433.

References

1. Daumé III, H.: Frustratingly easy domain adaptation. In: Association for Computatoinal Linguistics (ACL). vol. 1785 (2007)
2. Evgeniou, T., Pontil, M.: Regularized multi–task learning. In: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 109–117. ACM (2004)
3. Kuiper, J., Wallace, B.C., Marshall, I.J.: Spa. `http://figshare.com/articles/Spa/997707` (2014)
4. Marshall, I.J., Kuiper, J., Wallace, B.C.: Automatically assessing bias in clinical trials and extracting supporting sentences. In: under review at Knowledge Discovery in Databases (KDD) (2014)
5. Mintz, M., Bills, S., Snow, R., Jurafsky, D.: Distant supervision for relation extraction without labeled data. In: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2. pp. 1003–1011. Association for Computational Linguistics (2009)
6. Nguyen, T., Moschitti, A.: End-to-end relation extraction using distant supervision from external semantic repositories. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2. pp. 277–282. Association for Computational Linguistics (2011)
7. Sackett, D.L., Rosenberg, W.M., Gray, J., Haynes, R.B., Richardson, W.S.: Evidence based medicine: what it is and what it isn't. BMJ: British Medical Journal 312(7023), 71–72 (1996)
8. Valkenhoef, G., Tervonen, T., Brock, B., Hillege, H.: Deficiencies in the transfer and availability of clinical trials evidence: a review of existing systems and standards. BMC Medical Informatics and Decision Making 12(1), 95 (2012), `http://www.biomedcentral.com/1472-6947/12/95`