

A Simple Machine Learning Framework to Aid Citation Screening in Systematic Reviews and Meta-Analyses of Aging and Longevity Research Studies

Marko Lalović

March 10, 2021

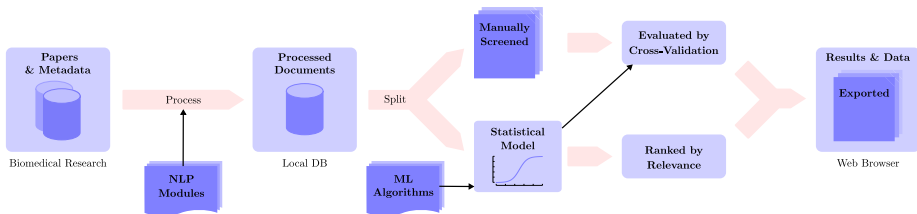


Figure 1: General overview of proposed framework.

Background

- A systematic review typically addresses a specific clinical question by collecting and analyzing data from all the relevant and unbiased set of studies.
- Citation screening is the first yet tedious task of narrowing down the large set of citations retrieved via a broad database query to those relevant for the review.

Idea

Develop a machine learning framework to semi-automate citation screening in systematic reviews and meta-analyses, thereby reducing reviewers workload (and screening errors).

Related Work

- Wallace et al. developed a semi-automated citation screening algorithm for systematic reviews of biomedical literature.
- Bannach-Brown et al. described their approaches to aid citation screening for a systematic review of pre-clinical animal studies.
- Howard et al. deployed a general software system that automate the required methodologies called “SWIFT-Review”.
- Przybyła et.al introduced a web-based software system called “RobotAnalyst”.
- O'Mara-Eves et al. performed a systematic review of current approaches.
- To date, no use of any tools related to automating (or semi-automating) the screening process of systematic reviews or meta-analyses of aging and longevity research was reported.

Our Contribution

- Present a simple machine learning framework that can be used in the screening stage of systematic reviews or meta-analyses of aging and longevity research studies.
- Present our experimental results on dataset related to [Dasatinib and Quercetin Senolytic Therapy Risk-Benefit Analysis](#)¹(D&Q Analysis).

¹The analysis is part of “Rejuvenation Now” non-profit initiative that: “seeks to continuously identify potential rejuvenation therapies and systematically evaluate their risks, benefits, and associated therapeutic protocols to create transparency” published by Forever Healthy Foundation.

Outline

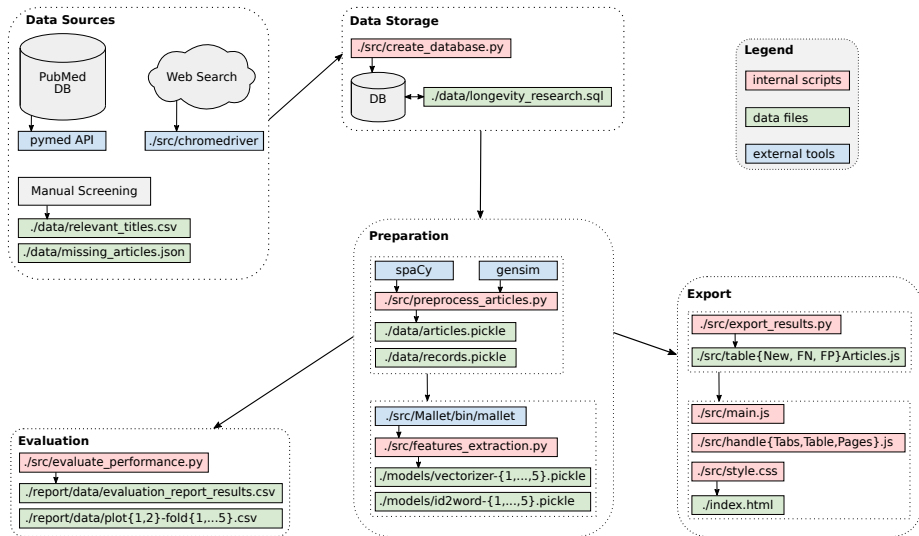


Figure 2: Technical Overview.

Features

Extracted features are based on:

- Provided list of search terms: dasatinib, senolytic, senescent, ...
- Possible publication types: case report, clinical trial, review, ...
- TF-IDF scores of terms: chronic myeloid, adverse event, tyrosine kinase, ...
- LDA model probabilities of belonging to topics:

Topic Terms

- | Topic | Terms |
|-------|--|
| 1. | trial, clinic, efficaci, report, safeti, assess, evalu, show, advers, ... |
| 2. | week, placebo, treatment, group, extract, symptom, score, patient, hypericum, ... |
| 3. | dasatinib, patient, case, treatment, report, chronic, leukemia, therapi, myeloid, ... |
| 4. | cvd, risk, cardiovascular, prevent, factor, profil, diseas, lipid, import, ... |
| 5. | flavonoid, anthocyanin, individu, adult, genistein, dietari, isoflavon, flavanon, intak, ... |
| 6. | sunitinib, sorafenib, target, imatinib, includ, cancer, erlotinib, anticanc, malign, ... |
| 7. | bone, marrow, chromosom, abnorm, deriv, prognost, signific, aberr, delet, ... |
| 8. | lifespan, elegan, longev, stress, effect, life, span, extend, increas, ... |
| 9. | muscl, smooth, vsmc, skelet, aortic, havsmc, resist, accordingli, folfiri, ... |
| 10. | quercetin, effect, cell, concentr, human, dose, depend, studi, increas, ... |

...

Model

- Used L1 regularized logistic regression model from scikit-learn.
- Fitted using Liblinear solver with balanced class weights.
- Placed more emphasis on recall by using recall scorer.

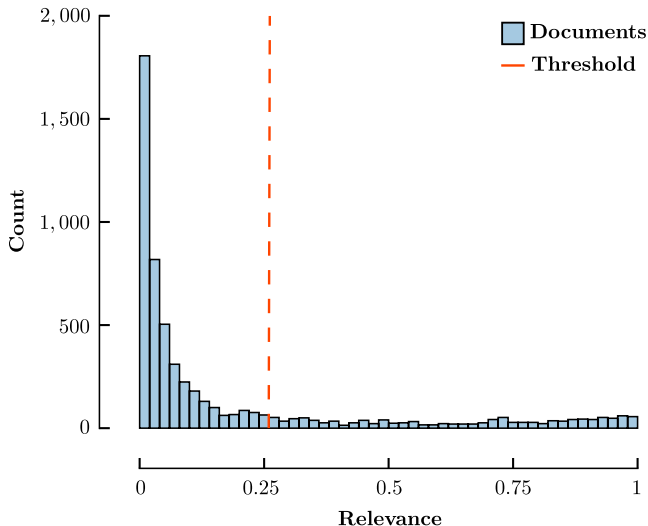
Definition (Relevance)

The model estimates the conditional probability, called *relevance score*, that a given document d is relevant given feature vector X^d :

$$\Pr(d \text{ is relevant} | X^d)$$

Threshold Selection

The selected cut-off threshold for D&Q Analysis was 0.26 where the binary classifier achieved 95% recall:



Errors

There are four possible outcomes of comparing the classification result with human assigned label:

		Classified as relevant	
		Yes	No
Labeled as relevant	Yes	True Positive	False Negative
	No	False Positive	True Negative

Performance Evaluation

Performance was assessed by:

Definition (Statistical measures)

Precision P is the fraction of documents labeled as relevant among documents classified as relevant:

$$P = \frac{TP}{TP + FP}$$

Recall R (also known as *sensitivity*) is the fraction of documents labeled as relevant that were also classified as relevant:

$$R = \frac{TP}{TP + FN}$$

Work Saved over Sampling $WSS@R$ is the reduction of documents that need to be screened compared to a random ordering of the documents to achieve a level of recall R :

$$WSS@R = \frac{TN + FN}{N} - (1 - R)$$

Evaluation Results

Fold	Recall	Precision	PR-AUC	WSS@R
1	0.94	0.13	0.54	0.53
2	0.90	0.16	0.33	0.61
3	1.00	0.14	0.48	0.63
4	0.94	0.19	0.54	0.67
5	0.97	0.20	0.43	0.71
Mean	0.95 (0.03)	0.17 (0.03)	0.46 (0.08)	0.63 (0.06)

Table 1: Summarized results of 5-fold cross-validation for D&Q Analysis

Evaluation of Results

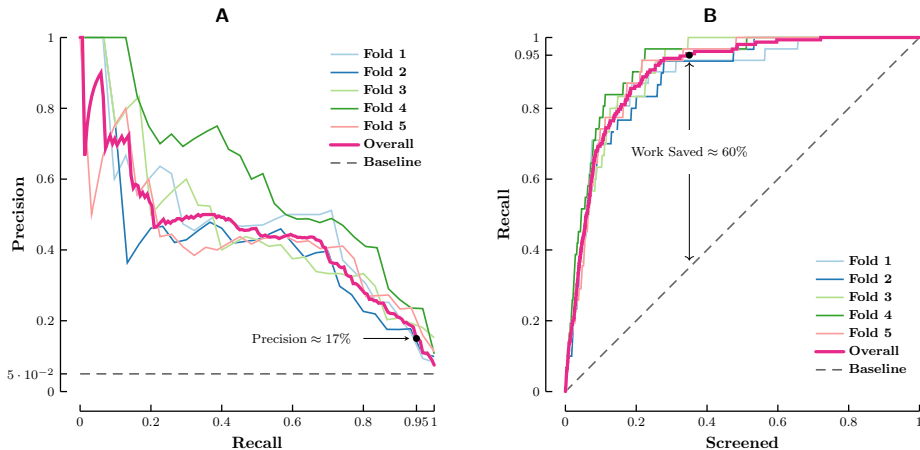


Figure 3: Visualized results of 5-fold cross-validation for D&Q Analysis

Results for Dasatinib and Quercetin Senolytic Therapy Risk-Benefit Analysis

Table 2: Estimated Relevance for False Negatives until 17. April, 2020 included in the Risk-Benefit Analysis but not classified as relevant.					
<div><div>New Articles</div><div>False Negatives</div><div>False Positives</div></div>					
Relevance	Date	Title	Abstract	Expand	URL
0.3	22. März, 2013	Bioavailability of quercetin: problems and promises	Quercetin (QC) is a typical plant flavonoid, possesses d...	<div>+</div>	pubmed
0.29	1. Januar, 2005	Cytotoxicity of flavonoids toward cultured normal hum...	The cytotoxicity of flavonoids, including apigenin, erio...	<div>+</div>	pubmed
0.25	6. Februar, 2016	Association Between BCR-ABL Tyrosine Kinase Inhib...	Importance: A phase 3 trial with ponatinib in patients ...	<div>+</div>	pubmed
0.24	22. August, 2016	Targeting Pro-Inflammatory Cells in Idiopathic Pulmonary Fibrosis: a Human Trial (IPF) Abstract The study team hypothesizes that intermittent (3 doses administered over 3 consecutive days in 3 consecutive weeks) oral administration of combination Dasatinib (100 mg/d) + Quercetin (1250 mg/d) will be safe and well tolerated in patients with IPF. Treatment with D+Q will result in reduced abundance of pro-inflammatory cells within subjects over baseline. Finally, the reduction in biomarkers of cellular pro-inflammatory state will be related to no change in functional and patient reported outcomes.		<div>-</div>	clinicaltrials
0.22	19. Januar, 2017	Identification of cellular targets involved in cardiac fail...	Aims: The aims of the present study were to evaluate t...	<div>+</div>	pubmed
0.22	3. Oktober, 2017	Short-term High Dose of Quercetin and Resveratrol Alt...	Background: Hyperglycemia-mediated oxidative stress ...	<div>+</div>	pubmed
0.22	14. Februar, 2018	BCR-ABL Tyrosine Kinase Inhibitors: Which Mechani...	Imatinib, the first-in-class BCR-ABL tyrosine kinase in...	<div>+</div>	pubmed
0.19	1. Januar, 2008	Quercetin pharmacokinetics in humans	The purpose of this study was to examine the pharmac...	<div>+</div>	pubmed
0.18	6. Mai, 2009	Tyrosine kinase inhibitor-induced platelet dysfunction i...	Dasatinib is associated with increased risk of bleeding ...	<div>+</div>	pubmed
0.18	17. März, 2004	Quercetin, an over-the-counter supplement, causes neur...	A 22-month-old boy, who regularly consumed the oral ...	<div>+</div>	pubmed

Page: 1 2






Last updated: Do 25. Feb 11:38:01 CET 2021

Figure 4: Interactive tables of exported documents for D&Q Analysis.

Links

- Interactive tables of exported documents for D&Q Analysis:
<https://markolalovic.com/longevity-research-screening/>
- Technical report:
<https://zenodo.org/record/4593957/files/zenodo.4593957.pdf>
- Source code:
<https://github.com/markolalovic/longevity-research-screening/>

References

-  Bannach-Brown, A., Przybyła, P., Thomas, J. et al. "Machine learning algorithms for systematic review: reducing workload in a preclinical review of animal studies and reducing human screening error.", Syst Rev 8, 23 (2019).
<https://doi.org/10.1186/s13643-019-0942-7>
-  Howard BE, Phillips J, Miller K, et al. "SWIFT-Review: a text-mining workbench for systematic review.", Syst Rev. 2016;5:87. Published 2016 May 23.
[doi:10.1186/s13643-016-0263-z](https://doi.org/10.1186/s13643-016-0263-z)
-  O'Mara-Eves, A., Thomas, J., McNaught, J. et al. "Using text mining for study identification in systematic reviews: a systematic review of current approaches.", Syst Rev 4, 5 (2015).
<https://doi.org/10.1186/2046-4053-4-5>
-  Przybyła P, Brockmeier AJ, Kontonatsios G, et al. "Prioritising references for systematic reviews with RobotAnalyst: A user study.", Res Synth Methods. 2018;9(3):470-488.
<https://doi.org/10.1002/jrsm.1311>
-  Wallace, B.C., Trikalinos, T.A., Lau, J. et al. "Semi-automated screening of biomedical citations for systematic reviews.", BMC Bioinformatics 11, 55 (2010).
<https://doi.org/10.1186/1471-2105-11-55>