Revista Tecnologia e Sociedade
ISSN: 1984-3526

https://periodicos.utfpr.edu.br/rts

Machine learning e revisão sistemática de literatura automatizada: uma revisão sistemática

Denise Fukumi Tsunoda
dtsunoda@ufpr.br
Programa de Pós-Graduação em Gestão da Informação (PPGGI), Universidade Federal do Paraná

Paulo Sergio da Conceição Moreira
psxm54@gmail.com
Programa de Pós-Graduação em Gestão da Informação (PPGGI), Universidade Federal do Paraná

André José Ribeiro Guimarães
andrejrg@gmail.com
Programa de Pós-Graduação em Gestão da Informação (PPGGI), Universidade Federal do Paraná

RESUMO

Nos últimos tempos, diversos trabalhos de revisão sistemática da literatura têm sido publicados. No entanto, não foram identificados trabalhos que sumarizem os estudos até então publicados. Por este motivo, esta pesquisa busca identificar ferramentas e métodos de *machine learning,* artigos mais citados, países, autores de referência, áreas e *datasets* são empregados na automatização de revisões sistemáticas. Para isto, analisa, baseado na recomendação *Preferred Reporting Items for Systematic* Reviews, 35 de 328 documentos científicos recuperados das bases *Web of Science* (116) e Scopus (212), sem restrição de idioma ou recorte temporal. Os termos pesquisados foram: (a) *systematic review*, *machine learning* e *tool* e (b) revisão sistemática, aprendizagem de máquina e ferramenta. Nas pesquisas em língua portuguesa, não houve retorno nas duas bases consultadas. Como resultado, esta pesquisa demonstra a tendência de crescimento da produção relacionada ao tema, com 74,29% dos registros publicados após 2016 e identifica o interesse dos pesquisadores pela utilização de técnicas de *text mining*, sendo a palavra-chave mais utilizada pelos autores. Em relação aos métodos encontrados, evidencia o algoritmo *Support Vector Machine* como o mais frequente, sendo utilizado em nove trabalhos, seguido pelas heurísticas Redes Neurais Artificiais e Naïve Bayes, com três e duas aplicações cada. Uma das constatações da pesquisa é que a revisão sistemática de literatura tem aplicação majoritária à área médica, representando 51,43% dos documentos recuperados. Quanto às bases de dados da área médica, 34,29% (12/35) dos estudos explicitaram o uso da base PubMed como fonte; 5,71% (2/35) a Medline; e outros 5,71% (2/35) o O´Mara-Eves *dataset*. Os demais 5,71% (2/35) mencionaram o uso de *datasets* sem especificar a origem destes. Finalmente, o estudo realizado não identifica nenhuma ferramenta que possa ser utilizada de forma genérica e autônoma para revisão sistemática em áreas diversas do conhecimento. Os melhores resultados foram obtidos com experimentos semiautomatizados, onde parte da análise é executada pela ferramenta e outra parte, realizada manualmente.

PALAVRAS-CHAVE: Aprendizado de máquina. Revisão de literatura. Ferramentas de automatização. Mineração de textos.

INTRODUÇÃO

O processo de revisão de literatura é essencial para a criação de uma sólida fundamentação de pesquisas que buscam avançar o conhecimento de uma determinada área (WEBSTER; WATSON, 2002). Este processo pode facilitar o desenvolvimento de novas teorias, o reconhecimento de áreas onde já existe uma infinidade de outras pesquisas e a identificação de espaços ainda pouco explorados (WEBSTER; WATSON, 2002).

As revisões de literatura podem ser divididas em dois grupos. Conforme Webster e Watson (2002) e Ramdhani, Ramdhani e Amin (2014), o primeiro grupo avalia tópicos já consolidados, que possuem um corpo de pesquisas que necessita ser analisado e resumido. Este tipo de revisão identifica, descreve e propõe modelos conceituais que sintetizam as informações publicadas em determinada área específica, podendo apresentar ou não um recorte temporal. O segundo grupo de revisão de literatura, conforme Ramdhani, Ramdhani e Amin (2014), aborda tópicos novos ou emergentes que, pela sua natureza recente, ainda não passaram por uma revisão abrangente de literatura. Tais estudos usualmente resultam em análises iniciais ou preliminares, que conceituam um novo modelo ou delimitam uma nova área do conhecimento.

Segundo Ferenhof e Fernandes (2016) as revisões de literatura são, ainda, divididas em três tipos principais: narrativa, integrativa e sistemática. A narrativa é a considerada revisão tradicional ou exploratória na qual não são apresentados critérios explícitos para a seleção dos documentos e o autor pode incluir novos artigos conforme seu viés sem a preocupação em esgotar as fontes de informação.

A revisão integrativa é um método que tem como finalidade reunir e sintetizar resultados de pesquisas sobre um delimitado tema ou questão, de maneira sistemática e ordenada e abrangente, contribuindo para o aprofundamento do conhecimento do tema investigado (ERCOLE; MELO; ALCOFORDA, 2014).

A revisão sistemática é um método de investigação científica com um processo rigoroso e explícito para identificar, selecionar, coletar dados, analisar e descrever as contribuições relevantes a pesquisa. Trata-se de uma síntese rigorosa de todas as pesquisas relacionadas a uma questão específica com procedimentos transparentes metódicos e reproduzíveis (BORREGO; FOSTER; FROYD, 2014). Os autores complementam que este tipo de revisão visa extrair tendências, padrões, relações e oferecer um panorama geral dos estudos coletados.

As revisões sistemáticas são comuns em estudos da área da saúde e Melnyk e Fineout-Overholt (2011) afirmam que diversos resultados de estudos categorizados como de maior qualidade (nível 1, em uma escala de 1 a 7) são os cujos achados são oriundos deste tipo de revisão.

Diante do aumento massivo na produção de informação científica e das limitações inerentes às análises realizadas manualmente, há o surgimento de demandas por métodos e ferramentas que possam automatizar e agilizar o processo de revisão sistemática (TSAFNAT *et al.*, 2013; XIONG *et al.*, 2018). Diante desse cenário, iniciativas de processamento que utilizam abordagens do tipo *machine learning* (aprendizagem de máquina) têm se destacado. Em termos

gerais, *machine learning* define os métodos computacionais que utilizam a experiência acumulada para melhorar o desempenho ou para realizar previsões com alto grau de acurácia, que pode variar conforme a qualidade e o tamanho dos dados avaliados (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012).

Embora o emprego de *machine learning* para automatizar etapas do processo de revisão sistemática seja, até certo pronto, "óbvio" (WALLACE *et al.*, 2012b, p. 819), não foi possível identificar pesquisas que revelassem tendências, quais métodos de aprendizagem de máquina são predominantemente adotados para este fim, quais ferramentas e linguagens são frequentemente utilizadas e quais os resultados obtidos. Neste contexto, o objetivo desta pesquisa é identificar, por meio de uma revisão sistemática em publicações científicas, possíveis ferramentas que empreguem métodos de *machine learning* para automatização de revisões sistemáticas em qualquer área do conhecimento.

O artigo está organizado em quatro seções além desta introdução. Estas seções apresentam o tipo da pesquisa, materiais e métodos e as formas de avaliação (metodologia), apresentação e discussão dos resultados separados em duas seções: das análises quantitativas e das análises de conteúdo. Finalmente são apresentadas as principais conclusões e contribuições do estudo.

METODOLOGIA

Configurando-se como um estudo exploratório, a pesquisa é enquadrada como revisão sistemática baseada em pesquisa bibliográfica, com abordagem quali-quantitativa. Os procedimentos adotados se baseiam na recomendação PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*), proposta por Moher *et al.* (2009). Esta metodologia apresenta um fluxograma de quatro etapas (identificação, seleção, elegibilidade e inclusão) que foram adaptadas às necessidades da presente pesquisa.

Para a etapa de identificação, no dia 28/08/2020, foi realizada uma busca nas bases científicas *Web of Science* (WoS) e Scopus, ambas compatíveis com a ferramenta de análise bibliométrica Bibliometrix/Biblioshiny (ARIA; CUCCURULLO, 2017). Sem qualquer recorte temporal ou filtro em relação ao tipo dos documentos, as estratégias de busca realizadas foram: (*TS*=("*systematic review*") *AND TS*=("*machine learning*") *AND TS*=(*tool*)) e (TS=("revisão sistemática") AND TS=("aprendizagem de máquina") AND TS=(ferramenta)) para a WoS e (*TITLE-ABS-KEY* ("*systematic review*") *AND TITLE-ABS-KEY* ("*machine learning*") *AND TITLE-ABS-KEY* (*tool*)) e (*TITLE-ABS-KEY* ("*revisão sistemática*") *AND TITLE-ABS-KEY* ("*aprendizagem de máquina*") *AND TITLE-ABS-KEY* (ferramenta)) , para a Scopus. Nas pesquisas em língua portuguesa, não houve retorno nas duas bases consultadas.

Nas pesquisas com os termos em inglês, na base WoS, foram retornados 116 documentos, dos quais, após aplicação dos critérios de exclusão na seguinte ordem: artigos não disponíveis em acesso aberto (46), artigos não publicados (2), material editorial (1) restaram 67 artigos para a leitura parcial (*screening*).

Na etapa de *screening* de título, resumo e palavras-chave, artigos que não apresentaram análise de ferramentas foram eliminados da pesquisa, a exemplo dos cinco retornos mais recentes (2020): "*Diagnostic performance of artificial intelligence to detect genetic diseases with facial phenotypes: A protocol for*

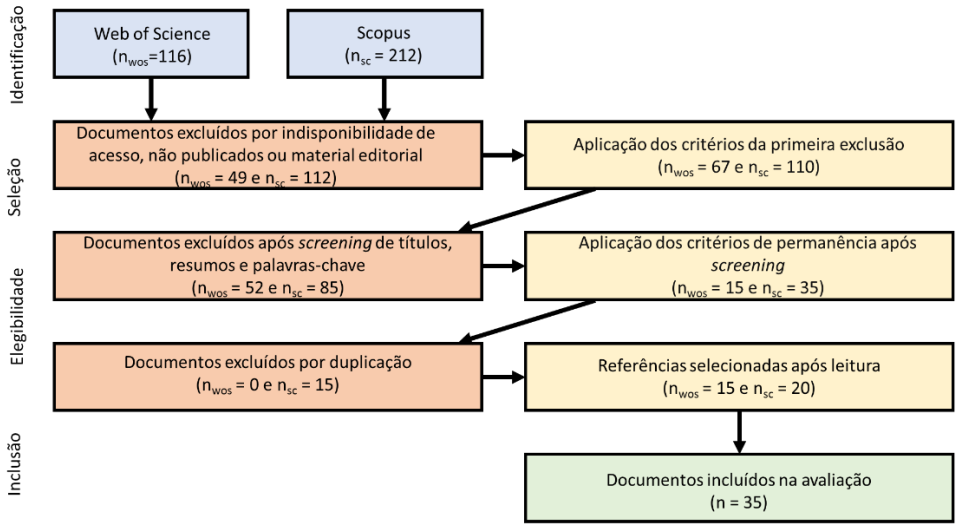
systematic review and meta analysis", "*Application of Support Vector Machine on fMRI data as biomarkers in Schizophrenia diagnosis: A systematic review*", "*Early warning score validation methodologies and performance metrics: A systematic review*", "*A systematic review of machine learning models for predicting outcomes of stroke with structured data*" e "*Smartphone applications targeting precision agriculture practices-A systematic review*". Pelos títulos dos trabalhos, apesar da relevância nas respectivas áreas, não tratam do uso de ferramentas de machine learning para revisões sistemáticas. Os trabalhos são revisões sistemáticas sobre um assunto que utiliza *machine learning*. Com este critério, 52 trabalhos foram eliminados e restaram 15 artigos para a leitura completa.

Na Scopus, a busca resultou em 212 documentos, dos quais, após a aplicação dos critérios de exclusão na seguinte ordem: artigos não disponíveis em acesso aberto (99), artigos não publicados (3), material editorial (0), restaram 110 artigos para a leitura parcial (*screening*).

Na etapa de *screening* de título, resumo e palavras-chave, artigos que não apresentaram análise de ferramentas foram eliminados da pesquisa, a exemplo dos cinco retornos mais recentes (2020): "*Translating research findings into clinical practice: A systematic and critical review of neuroimaging-based clinical tools for brain disorders*", "*Artificial intelligence to improve back pain outcomes and lessons learnt from clinical classification approaches: Three systematic reviews*", "*Application of systematic evidence mapping to assess the impact of new research when updating health reference values: A case example using acrolein*", "*Social, behavioral, and cultural factors of HIV in Malawi: Semi-automated systematic review*" e "*Artificial intelligence-based tools to control healthcare associated infections: A systematic review of the literature*". De forma análoga ao retorno da WoS, apesar da relevância nas respectivas áreas, os trabalhos não tratam do uso de ferramentas de *machine learning* para revisões sistemáticas. Com este critério, 75 trabalhos foram eliminados e restaram 35 artigos para a leitura completa.

A Figura 1 ilustra as etapas adotadas na criação da base de documentos analisados no estudo.

Figura 1 - Metodologia PRISMA adaptada para este estudo



Fonte: Elaborado pelos autores (2020).

Aplicando o critério de unicidade, verificou-se que 15 artigos da WoS estavam nos 35 da SCOPUS. Então, foram lidos, na íntegra, 35 artigos e nenhum foi eliminado.

Na sequência, os resultados foram unificados e os 35 documentos, exportados para o formato BibTex para análises (quantitativas e qualitativas) por meio das ferramentas Bibliometrix/Biblioshiny - um pacote desenvolvido na linguagem de programação R (ARIA; CUCCURULLO, 2017) - e Microsoft Excel®.

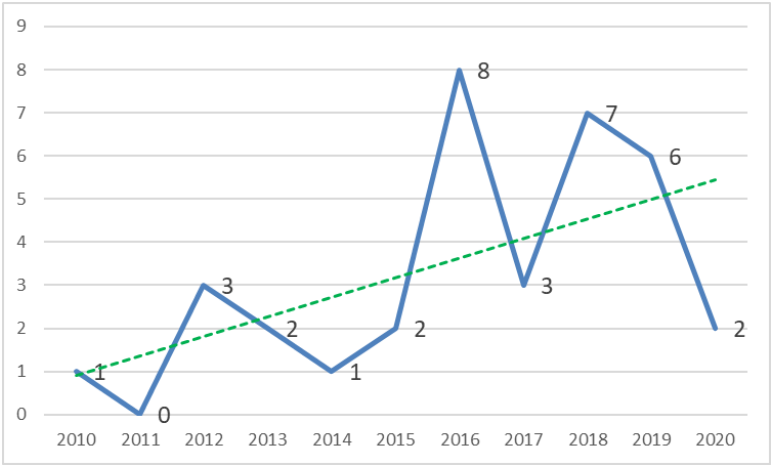
Na sequência, são apresentados e discutidos os resultados das análises quantitativas e de conteúdo dos 35 documentos.

RESULTADOS DAS ANÁLISES QUANTITATIVAS

Para averiguar a tendência na produção do tema pesquisado, verificou-se o número de publicações por ano. Essa análise revelou que 74,29% dos registros (26/35) foram publicados após o ano de 2016, confirmando o interesse das pesquisas relacionadas aos temas *machine learning* e revisão sistemática. A confirmação deste interesse é representada pela linha de tendência (pontilhada) apresentada no Gráfico 1.

Uma vez que a busca foi realizada em agosto de 2020, a ocorrência de dois artigos não representa a totalidade de registros do ano de 2020 e, de acordo com a linha tendência (tracejada no gráfico), este número deve ser maior até o fechamento do ano.

Gráfico 1 - Relação de número de publicações por ano com linha de tendência tracejada (verde)



Fonte: Elaborado pelos autores (2020).

Em relação à tipologia dos 35 registros recuperados, 65,70% dos registros consistem em artigos de periódicos (23/35); 14,29% correspondem a trabalhos publicados em eventos (5/35); 14,29% representam revisões (5/35); e, por fim, um registro consiste em uma nota (2,86%) e um editorial (2,86%). Todos os registros recuperados foram elaborados em inglês, consistindo no único idioma retornado.

Quanto às fontes de publicação, 18 fontes foram identificadas, das quais, destacam-se: a) *Systematic Reviews*, com doze publicações, representando 34,29% das publicações recuperadas; b) ACM *International Conference Proceeding Series*

e *Journal of Biomedical Informatics* com três registros (8,57%) cada; c) *BMC Medical Informatics and Decision Making* e *Journal of the American Medical Informatics Association*, com dois registros (5,71%) cada. As demais fontes apresentaram apenas um registro cada.

Ainda em relação às fontes de publicação, constatou-se ao aplicar a lei de Bradford (LEIMKUHLER, 1967; VENABLE *et al.*, 2016) que, apesar da baixa quantidade de registros, a distribuição das fontes em termos de "zonas" corresponde à suposição de que a primeira zona é composta por um pequeno grupo de periódicos "mais devotos ao tema"; enquanto que a segunda e a terceira zonas exigem um número maior de periódicos (LEIMKUHLER, 1967; VENABLE *et al.*, 2016). A discussão exposta acima é representada pela Tabela 1.

Tabela 1 - Distribuição de periódicos conforme a lei de Bradford

Zonas	Total de periódicos	Total da produção	Total da produção (%)
Zona 1	1	12	34,29%
Zona 2	6	12	34,29%
Zona 3	11	11	31,42%

Fonte: Elaborado pelos autores (2020).

Assim, é possível constatar que, em termos de produção, cada uma das três zonas é responsável por aproximadamente um terço da publicação registrada para o tema pesquisado.

Acerca do impacto das fontes identificadas, utilizou-se os índices H (HIESCH, 2005) e G (EGGUE, 2006) como métricas para a identificação das fontes com o maior impacto. As dez fontes com maior impacto são representadas na Tabela 2.

Tabela 2 – As fontes com o maior impacto

Fonte	Índice H	Índice G	Citações	Registros	Ano
Systematic reviews	8	12	154	12	2015
ACM international conference proceeding series	2	3	22	3	2016
Journal of biomedical informatics	3	3	68	3	2016
BMC medical informatics and decision making	2	2	36	2	2012
Journal of the american medical informatics association	2	2	75	2	2015
Advances in bioinformatics	1	1	29	1	2012
BMJ (online)	1	1	67	1	2013
Computacion y sistemas	1	1	2	1	2016
Conservation biology	1	1	14	1	2018
Expert systems with applications	1	1	56	1	2014

Fonte: Elaborado pelos autores (2020).

Além dos índices H e G, considera-se o número de citações recebidas por cada fonte, o total de registros associados e o ano do primeiro registro de cada uma das fontes elencadas. Adicionalmente, considerando ambos os índices, nota-se que as fontes com maior impacto são os periódicos *Systematic Reviews* - Qualis B1 para a área Interdisciplinar (quadriênio 2013-2016) - e *Journal of Biomedical Informatics*, - Qualis A2 na área Interdisciplinar (quadriênio 2013-2016). A ACM *international conference proceeding series* não aparece na referida avaliação.

Contudo, apesar de se mostrarem as fontes com os maiores índices H e G, a fonte com o maior número de citações é o periódico *Expert Systems with Applications* - Qualis A1 na área Interdisciplinar (segundo quadriênio 2013-2016) - com 56 citações.

Quanto aos termos mais frequentes dos trabalhos analisados, na Figura 2, à esquerda, apresenta-se os termos com maior ocorrência nos títulos, as palavras-chave dos autores mais empregadas (centro); enquanto, à direita, ilustra-se os termos gerados pela maior ocorrência nos documentos (palavras-chave extendidas).

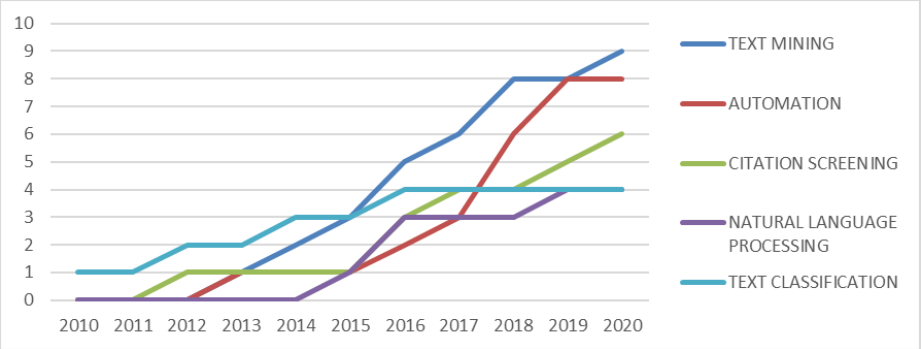
Figura 2 - Comparação dos termos utilizados: títulos, palavras-chave e palavras-chave extendidas

Fonte: Elaborado pelos autores (2020).

Em relação aos títulos dos documentos (à esquerda), os principais termos consistem em "*systematic*" com 24 ocorrências e "*reviews*" com 16 ocorrências. As palavras-chave do autor (centralizado) com o maior destaque são os termos "*systematic review*" e "*text mining*", com dez e nove ocorrências cada; e "*automation*" e "*sustematic review*", com oito e sete ocorrências. Como os termos de busca envolviam alguns destes termos, merecem destaque os termos das palavras-chave extendidas tais como: "*automation*" e "*data mining*".

Considerando a evolução dos termos em uma análise acumulativa das palavras-chave utilizadas pelos autores em todos os trabalhos, foram retirados os termos de busca: "*machine learning*" e "*systematic review*". O termo com maior frequência é *"text mining"* (42 ocorrências). Observa-se ainda crescimento acentuado do termo "*citation screening*", pois diversos estudos contemplam esta área, conforme apresentado na próxima seção. A evolução das cinco palavras-chave (com exceção dos termos de busca) é apresentada na Figura 3.

Figura 3 - Evolução das palavras-chave definidas pelos autores



Fonte: Elaborado pelos autores (2020).

A respeito dos documentos mais citados, são 20 artigos com um total de 575 citações. Os dez artigos com o maior número de citações são elencados na Tabela 3 uma vez que estes representam 72,52% do total (417/575).

Tabela 3 – Os dez documentos mais citados

#	Documento	Referência	Ano	Cit.	Cit./ ano
1	*The automation of systematic reviews*	(TSAFNAT *et al.*, 2013)	2013	67	8,4
2	*Robotreviewer: evaluation of a system for automatically assessing bias in clinical trials*	(MARSHALL; KUIPER; WALLACE, 2016)	2016	57	11,4
3	*Automatic text classification to support systematic reviews in medicine*	(GARCIA ADEVA *et al.*, 2014)	2014	56	8,0
4	*Automated content analysis: addressing the big literature challenge in ecology and evolution*	(NUNEZ-MIR *et al.*, 2016)	2016	46	9,2
5	*Topic detection using paragraph vectors to support active learning in systematic reviews*	(HASHIMOTO *et al.*, 2016)	2016	38	7,6
6	*Swift-review: a text-mining workbench for systematic review*	(HOWARD *et al.*, 2016)	2016	36	7,2
7	*Toward modernizing the systematic review pipeline in genetics: efficient updating via data mining*	(WALLACE *et al.*, 2012a)	2012	33	3,7
8	*Studying the potential impact of automated document classification on scheduling a systematic review update*	(COHEN; AMBERT; MCDONAGH, 2012)	2012	31	3,4
9	*Applications of natural language processing in biodiversity science*	(THESSEN; CUI; MOZZHERIN, 2012)	2012	29	3,2
10	*Making progress with the automation of systematic reviews: principles of the international collaboration for the automation of systematic reviews (ICASR)*	(BELLER *et al.*, 2018)	2018	24	8,0
			TOTAL	417	

Fonte: Elaborado pelos autores (2020).

Na Tabela 3, além do número de citações (Cit.), apresenta-se o ano, a referência e a média de citações por ano para cada registro (Cit. / ano). O trabalho realizado por Tsafnat *et al.* (2013) apresenta o maior número de citações (67). Entretanto, em média, os trabalhos com maior destaque foram elaborados por Marshall, Kuiper e Wallace (2016) e Nunez-Mir *et al.* (2016), com 11,4 e 9,2 citações em média por ano, respectivamente.

Em relação às citações em termos de países, os países mais citados são: a) Estados Unidos, com 195 citações; b) Austrália, com 67 citações c) Reino Unido, com 57 citações; d) Espanha, com 56 citações; e) Israel, com 14 citações; f) Canadá, com 13 citações; e Áustria, com 3 citações. No Brasil, identificou-se apenas um artigo com participação de um pesquisador brasileiro: Ricardo A. Correia, do Instituto de Ciências Biológicas e Saúde da Universidade Federal de Alagoas, cujo título do trabalho é "*Using machine learning to disentangle homonyms*" (ROLL; CORREIA; BERGER-TAL, 2018). Para este artigo, contudo, não há citações mencionadas.

RESULTADOS DAS ANÁLISES DE CONTEÚDO

Do total de artigos avaliados, 71,43% (25/35) caracterizam-se como *Systematic Literature Review* (SLR) (Revisão Sistemática de Literatura) e 28,57% (10/35) foram classificados como *screening* (triagem). No primeiro caso, considerou-se os estudos que analisaram o documento na íntegra; para o segundo caso, julgou-se os estudos que utilizaram título, resumo e palavras-chave como entrada dos experimentos. Um dos documentos (TSAFNAT *et al.*, 2013) é um editorial e foi classificado como SLR por realizar uma avaliação geral da área e foi incluído nesta avaliação ter recebido o maior número de citações (Tabela 3).

Dos registros recuperados, quatro comparam ferramentas. Tsou *et al.* (2020) comparam o Abstrackr e Eppi-Reviewer em nove relatórios médicos. Para três relatórios longos (de 2.706, 3.181 e 9.038 citações para *screening*) ambas as ferramentas tiveram bom desempenho, reduzindo o tempo de *screening* em 49% e 60,00%, respectivamente. Para relatórios curtos (entre 226 e 889 citações para *screening*), ambas as ferramentas tiveram bom desempenho mas em todos os experimentos, o Eppi-Reviewer foi ligeiramente superior.

Gates *et al.* (2019) comparam três ferramentas: Abstrackr, DistillerSR e RobotAnalyst. Em cada ferramenta submeteram um conjunto de 200 registros e calcularam a proporção de perdas, carga de trabalho e economia de tempo em comparação com os registros de *screening* não automatizados. Para testar as experiências dos usuários, oito pesquisadores utilizaram cada ferramenta e completaram uma pesquisa de avaliação. A economia de carga de trabalho proporcionada na simulação automatizada veio com o aumento do risco de registros relevantes não serem identificados.

O trabalho de O´Connor *et al.* (2019) discute a resistência de alguns revisores na adoação de ferramentas de *machine learning* que os auxiliem nas tarefas. Por meio do estudo, apontam como principais "barreiras": ferramentas não são compatíveis com os fluxos de trabalhos, não são confiáveis e as métricas que não

estão esclarecidas nas documentações dos sistemas. Schmidt *et al.* (2020) propõem uma "revisão viva" para rever processos e ferramentas disponíveis para extração de dados para semiautomatizar o processo de revisão sistemática. Trata-se de uma proposta de protocolo da referida pesquisa, ainda sem resultados. Um ponto de convergência entre os quatro trabalhos, está nas conclusão de que as decisões das ferramentas precisam ser combinadas com as de um revisor (análises semiautomatizadas) para minimizar os erros.

Alguns trabalhos a exemplode Gartlehner *et al.* (2019) e Sobocezenski *et al.* (2019) apenas registram experiências com o uso de ferramentas, em comparação às revisões manuais. Gartlehner *et al.* (2019) utilizaram a ferramenta DistillerAI para *screening* semiautomatizado. Para isto foi realizada uma comparação dos resultados da ferramenta com cinco equipes de revisores sistemáticos profissionais que examinaram os memos 2.472 resumos. Cada equipe treinou o DistellerAI com 300 resumos (treinamento e os demais resumos como conjunto de teste) selecionados de forma aleatória que a equipe realizou duas vezes o *screening*. As comparações foram: abordagem assistida por ferramenta, *screening* por revisor único e *screening* apenas com a ferramenta. As conclusões, considerando-se a precisão, foram de que o DistillerAI ainda não é adequado para substituir um revisor humano.

Sobocezenski *et al.* (2019) utilizaram o RobotReviewer e avaliaram a redução do tempo de revisões semiautomáticas com o seu uso em quatro artigos selecionados de forma aleatória. Para dois dos quatro artigos, o processamento foi semiautomatizado: eram fornecidas respostas da ferramenta e os 41 participantes poderiam sugerir alterações / correções necessárias. Como conclusões, são apontadas que as revisões semiautomatizadas são mais rápidas (média de 755 *versus* 824 segundos) e 62,00% dos destaques apontados pela ferramenta foram aceitos. Os revisores apontaram classificação entre "bom" e "excelente" para a melhoria do processo quando este é semiautomatizado.

Quanto à tarefa explicitada, 57,14% dos registros correspondem à tarefa de classificação (20/35); 11,43% (4/35) à tarefa de agrupamento (*clustering*); e 20,00% dos registros não informaram a tarefa (7/35). Aos trabalhos elaborados por Thessen, Cui e Mozzherin (2012), Beller *et al.* (2018), Tsafnat *et al.* (2013) e O´Connor *et al.* (2018), esta avaliação não se aplica, uma vez que o primeiro apenas verifica o uso de Processamento de Linguagem Natural (PLN) a uma base de biodiversidade, o segundo relata os resultados de um encontro no qual se definiram princípios para a área de revisão sistemática, o terceiro é um editoral que discute aspectos relevantes da área e o último apresenta um sumário de discussões no evento ICASR (*International Collaboration for Automation of Systematic Reviews*) ocorrido em outubro de 2016.

Quanto ao assunto específico, 37,14% (13/35) abordam a área de medicina e 28,57% (10/35) apresentam proposta de ferramentas. Destas dez propostas, cinco são de ferramentas relacionadas à medicina. Ou seja, 51,43% (18/35) dos trabalhos são específicos da área médica.

Uma síntese dos assuntos específicos é representada na Tabela 4.

Tabela 4 - Síntese dos assuntos específicos

Assunto específico	#
Medicina	13
Ferramenta	10
Text mining	2
Comparação de ferramentas	4
Ecologia	1
Filmes	1
Princípios	1
Processamento de linguagem natural	1
Protein data bank	1
Tomada de decisão	1

Fonte: Elaborado pelos autores (2020).

Dentre os métodos identificados nas pesquisas, nove trabalhos (WALLACE *et al.*, 2010, 2012a; COHEN; AMBERT; MCDONAGH, 2012; BUI *et al.*, 2016; MARSHALL; KUIPER; WALLACE, 2016; OLORISADE; BRERETON; ANDRAS, 2017; GATES; JOHNSON; HARTLING, 2018; KREINER; HAYN; SCHREIER, 2018; COHEN, 2015) utilizaram unicamente o algoritmo SVM; cinco trabalhos (GARCÍA ADEVA *et al.*, 2014; MO; KONTONATSIOS; ANANIADOU, 2015; YE *et al.*, 2016; BANNACH-BROWN *et al.*, 2019; MARSHALL; WALLACE, 2019) mencionaram o uso de SVM combinado com outros métodos. Desta forma, o algoritmo SVM foi empregado em 40,00% dos trabalhos.

Em seis trabalhos (STANSFIELD; THOMAS; KAVANAGH, 2013; HOWARD *et al.*, 2016; MORENO-GARCÍA; ACEVES-MARTINS; SERRATOSA, 2016; NUNEZ-MIR *et al.*, 2016; OLORISADE *et al.*, 2016; TSAFNAT *et al.*, 2018), não foram explicitados os métodos. Em outros trabalhos, este critério não se aplica, tais como: a) "*A machine learning approach for semi-automated search and selection in literature studies*" (ROS; BJARNASON; RUNESON, 2017) por tentar reproduzir – sem sucesso – seis experimentos já anteriormente relatados em outros artigos; b) "*Applications of natural language processing in biodiversity science*" (THESSEN; CUI; MOZZHERIN, 2012), por definir uma taxonomia para biodiversidade com o objetivo de otimizar o uso de ferramentas de PLN; e c) "*Making progress with the automation of systematic reviews: Principles of the international collaboration for the automation of systematic reviews* (ICASR)" (BELLER *et al.*, 2018) por ter como objetivo a apresentação de princípios e opções de ferramentas para revisão sistemática automática discutidas no primeiro encontro da ICASR, em Viena (2015).

A relação dos métodos empregados nos registros recuperados é apresentada na Tabela 5.

Tabela 5 - Métodos empregados

Método	#
SVM	9
Não informado	8
Não se aplica	5
Redes Neurais Artificiais	4
Naïve Bayes (NB)	2
Bit-Priori	1
Latent Dirichlet Allocation (LDA)	1
Louvain	1
NB, SVM, Regressão Logística (RL)	1
SVM, LDA	1
SVM, RL, *Random Forest*	1
SVM, NB	1

Fonte: Elaborado pelos autores (2020).

Dentre todos os estudos, nove relataram propostas de novas ferramentas. Três experimentos utilizaram a biblioteca libsvm - *A Library for Support Vector Machines*; dois empregaram a biblioteca *scikit-learn*, desenvolvida em Python; e um trabalho utilizou a ferramenta Weka. Os demais registros utilizaram ferramentas variadas tais como o RobotReviewer e o LINGO3G. Dentre os que fizeram suas próprias propostas, destacam-se: a) Abstrackr (GATES; JOHNSON; HARTLING, 2018); b) Twister (KREINER; HAYN; SCHREIER, 2018); c) SparkText (YE *et al.*, 2016); d) Swift-Review (HOWARD *et al.*, 2016); e) AFLEX Tag (RAMEZANI *et al.*, 2017) e f) uma ferramenta que utiliza Redes Neurais Artificiais para agrupamento de documentos (HASHIMOTO *et al.*, 2016).

Quanto às bases de dados, 34,29% (12/35) dos estudos explicitaram o uso da base PubMed como fonte; 5,71% (2/35) utilizaram o Medline; e outros 5,71% (2/35) usaram o O´Mara-Eves dataset. Os demais 5,71% (2/35) mencionaram os *datasets* sem especificar a origem destes.

Por fim, constatou-se o uso de técnicas de *text mining* para automação de revisões sistemáticas em diversas ferramentas propostas, dentre as quais destacam-se as já mencionadas: a) Abstrackr (GATES; JOHNSON; HARTLING, 2018), Twister (KREINER; HAYN; SCHREIER, 2018), b) SparkText (YE *et al.*, 2016); e c) Swift-Review (HOWARD *et al.*, 2016).

Além destas, destaca-se o trabalho desenvolvido por (TSAFNAT *et al.*, 2018) que utilizou o GATE (http://gate.ac.uk), uma ferramenta de mineração de texto que fornece por meio de interface gráfica funções comuns deste tipo de mineração de dados, como tokenização de palavras e sentenças.

CONSIDERAÇÕES FINAIS

Este artigo discorre sobre o resultado de uma revisão sistemática nas bases *Web of Science* e Scopus com os termos "*machine learning*", "*systematic review*" e "*tool*" sem restrição de localização destes no documento. Foram recuperados 328 documentos (116 da WoS e 212 da Scopus) e, após etapas de remoção de dados duplicados, triagem (*screening*) e leitura completa dos artigos, foram incluídos nas análises dos 35 documentos.

Como métodos, destaca-se, predominantemente, a utilização do algoritmo SVM, seguido pelo algoritmo Naïve Bayes e pelas Redes Neurais Artificiais. Esta pesquisa não apontou resultados que evidenciem o uso de, por exemplo, estratégias evolucionárias, que podem ser objeto de estudo em pesquisas futuras.

Percebe-se o interesse crescente dos cientistas pela utilização de técnicas de *text mining* na concepção de ferramentas automatizadas. Como exemplo, cita-se que as quatro ferramentas mencionadas - Abstrackr, Twister, SparkText e Swift-Review - utilizam *text mining* em seus projetos.

Apesar de diversas tentativas de desenvolvimento de ferramentas, como observadas por esta RSL , dentre as ferramentas pesquisadas nenhuma é aplicável a outros domínios que não seja a medicina, tampouco apresentam compatibilidade com outras bases de dados, além das mencionadas PUBMED, MEDLINE ou O´Mara-Eves dataset.

Outras ferramentas utilizadas como a LIBSVM, a scikit-learn e o Weka são aplicáveis a diversos domínios e não atingem o propósito do estudo conduzido: buscar por uma (ou diversas) ferramentas que utilizassem técnicas de *machine learning* para automatizar o processo de revisão sistemática em qualquer área do conhecimento.

Como continuidade dos estudos, pretende-se criar um protocolo para avaliar o uso das ferramentas de revisão sistemática em dois *corpora* controlados na área de inteligência artificial com a finalidade de registrar os resultados.

Machine learning and automated systematic literature review: a systematic review

ABSTRACT

In recent times, several systematic literature reviews have been published. However, to the best of our knowledge, no studies have been identified that summarize the papers published so far. For this reason, this research seeks to identify machine learning tools and methods, most cited articles, countries, reference authors, areas and datasets are used to automate systematic reviews. In this regard, we analyzed, based on the recommendation Preferred Reporting Items for Systematic Reviews, 35 of 328 scientific documents retrieved from the Web of Science (116) and Scopus (212) databases, without language or period constraints. The search terms were: (a) systematic review, machine learning and tool and (b) revisão sistemática, apredizagem de máquina and ferramenta. The search with Portuguese terms returned no articles. The results show that the production growth related to the theme is 74.29% of the records published after 2016 and identifies the interest of researchers for using text mining techniques, which is the most used keyword by the authors. Regarding the methods found, this research found Support Vector Machine algorithm as the most frequent, being used in nine studies, followed by Artificial Neural Networks and Naïve Bayes, with three and two applications each. One of the findings of the research is that the systematic reviews has a major application to the medical area, representing 51.43% of the documents recovered. As for the medical area databases, 34.29% (12/35) of the studies explained the use of PubMed as a source; 5.71% (2/35) Medline; and other 5.71% (2/35) O'Mara-Eves dataset. The other 5.71% (2/35) mentioned the use of datasets without specifying their origin. Finally, we stress that we could not found any tool able to be unrestrictedly used for systematic reviews. The best results were obtained with semi-automated experiments, partly performed by a tool and partly, manual.

KEYWORDS: Machine learning. Systematic review. Automation tools. Text Mining.

REFERÊNCIAS

ARIA, M.; CUCCURULLO, C. Bibliometrix: an R-tool for comprehensive science mapping analysis. **Journal of Informetrics**, v. 11, n. 4, p. 959–975, 2017.

BANNACH-BROWN, A. *et al. Machine learning* algorithms for systematic review: Reducing workload in a preclinical review of animal studies and reducing human screening error. **Systematic Reviews**, v. 8, n. 1, p. 1–12, 2019.

BELLER, E. *et al.* Making progress with the automation of systematic reviews: Principles of the International Collaboration for the Automation of Systematic Reviews (ICASR). **Systematic Reviews**, v. 7, n. 1, p. 1–7, 2018.

BUI, A. D. D.; FIOL, G. D.; JONNALAGADDA, S. PDF text classification to leverage information extraction from publication reports. **Journal of Biomedical Informatics**, v. 61, p. 141–148, 2016.

BORREGO, M.; FOSTER, M. J.; FROYD, J. E. Systematic literature reviews in engineering education and other developing interdisciplinary fields. **Journal of Engineering Education**, v. 103, n. 1, p. 45–76, 2014.

COHEN, A. M.; AMBERT, K.; MCDONAGH, M. Studying the potential impact of automated document classification on scheduling a systematic review update. **BMC Medical Informatics and Decision Making**, v. 12, n. 1, 2012.

COHEN, A.M. *et al.* Automated confidence ranked classification of randomized controlled trial articles: An aid to evidence-based medicine. **J. Am. Med. Inform. Assoc.**, v. 22, p. 707-717, 2015.

EGGUE, L. Theory and practise of the G-index. **Scientometrics**, v. 69, n. 1, p. 131–152, 2006.

ERCOLE, F. F.; MELO, L. S.; ALCOFORADO, C. L. G. C. Revisão integrativa versus revisão sistemática. **Revista Mineira de Enfermagem**, v. 18, n. 1, p. 9-12, 2014.

GARCÍA ADEVA, J. J. *et al.* Automatic text classification to support systematic reviews in medicine. **Expert Systems with Applications**, v. 41, n. 4, p. 1498–1508, 2014.

GARTLEHNER, G. *et al.* Assessing the accuracy of machine-assisted abstract screening with distillerai: A user study. **Systematic Reviews**, v. 8, n. 277, p. 1–10, 2019.

GATES, A.; JOHNSON, C.; HARTLING, L. Technology-assisted title and abstract screening for systematic reviews: A retrospective evaluation of the Abstrackr *machine learning* tool. **Systematic Reviews**, v. 7, n. 1, p. 1–9, 2018.

GATES, A. *et al.* Performance and usability of machine learning for screening in systematic reviews: a comparative evaluation of three tools. **Systematic Reviews**, v. 8, n. 1, p. 1–11, 2019.

HASHIMOTO, K. *et al.* Topic detection using paragraph vectors to support active learning in systematic reviews. **Journal of Biomedical Informatics**, n. 62, p. 59-65, 2016.

HIESCH, J. E. An index to quantify an individual's scientific research output. **PNAS**, v. 102, n. 46, p. 16569–16572, 2005.

HOWARD, B. E. *et al.* SWIFT-Review: A text-mining workbench for systematic review. **Systematic Reviews**, v. 5, n. 1, p. 1–16, 2016.

KREINER, K.; HAYN, D.; SCHREIER, G. Twister: A tool for reducing screening time in systematic literature reviews. **Studies in Health Technology and Informatics**, v. 255, n. 1, p. 5–9, 2018.

LEIMKUHLER, F. F. The Bradford distribution. **Journal of Documentation**, v. 23, n. 3, p. 197–207, 1967.

MARSHALL, I. J.; KUIPER, J.; WALLACE, B. C. RobotReviewer: Evaluation of a system for automatically assessing bias in clinical trials. **Journal of the American Medical Informatics Association**, v. 23, n. 1, p. 193–201, 2016.

MARSHALL, I.J.; WALLABE, B. C. Toward systematic review automation: A practical guide to using machine learning tools in research synthesis. **Systematic Reviews**, v. 8, n. 1, 2019.

MELNYK, B. M.; FINEOUT-OVERHOLT, E. Making the case for evidence-based practice. In: _____ (Ed.). **Evidence-based practice in nursing & healthcare**: a guide to best practice. 2. ed. Filadélfia: Lippincott Williams & Wilkins (LWW), 2011.

MO, Y.; KONTONATSIOS, G.; ANANIADOU, S. Supporting systematic reviews using LDA-based document representations. **Systematic Reviews**, v. 4, n. 1, 2015.

MOHER, D. *et al.* Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. **Annals of Internal Medicine**, v. 151, n. 4, p. 264–269, 2009.

MOHRI, M.; ROSTAMIZADEH, A.; TALWALKAR, A. **Foundations of machine learning**. London: The MIT Press, 2012. 427 p.

MORENO-GARCÍA, C. F.; ACEVES-MARTINS, M.; SERRATOSA, F. Unsupervised machine learning application to perform a systematic review and meta-analysis in medical research. **Computacion y Sistemas**, v. 20, n. 1, p. 7–17, 2016.

NUNEZ-MIR, G. *et al.* Automated content analysis: Addressing the big literature challenge in ecology and evolution. **Methods in Ecology and Evolution**, v. 7, n. 11, p. 1262–1272, 2016.

O´CONNOR, A.M. *et al.* Moving toward the automation of the systematic review process: a summary of discussions at the second meeting of International Collaboration for the Automation of Systematic Reviews (ICASR). **Systematic Reviews**, v. 7, n. 3, p. 1-5, 2018.

O´CONNOR, A.M. *et al.* A question of trust: Can we build an evidence base to gain trust in systematic review automation technologies?. **Systematic Reviews**, v. 8, n. 143, p. 1-8, 2019.

OLORISADE, B. K.; BRERETON, P.; ANDRAS, P. Reproducibility of studies on text mining for citation screening in systematic reviews: Evaluation and checklist. **Journal of Biomedical Informatics**, v. 73, p. 1–13, 2017.

OLORISADE, B. K. *et al.* A critical analysis of studies that address the use of text mining for citation screening in systematic reviews. In: 20th International Conference on Evaluation and Assessment in Software Engineering, 20, 2016, Limerick. **Proceedings…**, Limerick: ACM, 2016, p. 1–11.

RAMDHANI, A.; RAMDHANI, M. A.; AMIN, A. S. Writing a literature review research paper: A step-by-step approach. **International Journal of Basic and Applied Science**, v. 3, n. 1, p. 47–56, 2014.

RAMEZANI, K. *et al.* Rapid tagging and reporting for functional language extraction in scientific articles. In: WOSP 2017 - 6th International Workshop on Mining Scientific Publications. **Proceedings…**, p. 34-39, 2017.

ROLL, U.; CORREIA, R. A.; BERGER-TAL, O. Using *machine learning* to disentangle homonyms in large text corpora. **Conservation Biology**, v. 32, n. 3, p. 716–724, 2018.

ROS, R.; BJARNASON, E.; RUNESON, P. A *machine learning* approach for semi-automated search and selection in literature studies. In: 21th International Conference on Evaluation and Assessment in Software Engineering, 21., 2017, Karlskrona. **Proceedings…**, Karlskrona: ACM, 2017, p. 118–127.

SCHMIDT, L. *et al.* Data extraction methods for systematic review (semi)automation: A living review protocol. **F1000 Research 2020**, v. 4, n. 210, 2020.

SOBOCZENSKI, F. *et al.* Machine learning to help researchers evaluate biases in clinical trials: A prospective, randomized user study. **BMC Medical Informatics and Decision Making**, v. 19, n. 96, p. 1-12, 2019.

STANSFIELD, C.; THOMAS, J.; KAVANAGH, J. "Clustering" documents automatically to support scoping reviews of research: A case study. **Research Synthesis Methods**, v. 4, n. 3, p. 230–241, 2013.

THESSEN, A. E.; CUI, H.; MOZZHERIN, D. Applications of natural language processing in biodiversity science. **Advances in Bioinformatics**, v. 2012, p. 1–17, 2012.

TSAFNAT, G. *et al.* The automation of systematic reviews. **BMJ** (Online), v. 345, n. 7891, p. 9–10, 2013.

TSAFNAT, G *et al.* Automated screening of research studies for systematic reviews using study characteristics. **Systematic Reviews**, v. 7, n. 1, p. 64, 2018.

TSOU, A.Y. *et al.* Machine learning for screening prioritization in systematic reviews: Comparative performance of Abstrackr and Eppi-Reviewer. **Systematic Reviews**, v. 9, n. 1, p. 1-14, 2020.

VENABLE, G. T. *et al.* Bradford's law: identification of the core journals for neurosurgery and its subspecialties. **Journal of Neurosurgery**, v. 124, n. 2, p. 569–579, 2016.

WALLACE, B. C. *et al.* Toward modernizing the systematic review pipeline in genetics: Efficient updating via data mining. **Genetics in Medicine**, v. 14, n. 7, p. 663–669, 2012a.

WALLACE, B. C. *et al.* Deploying an interactive *machine learning* system in an evidence-based practice center. In: 2nd ACM SIGHIT International Health Informatics Symposium, 2., 2012b, Miami. **Proceedings...**, Miami: ACM, 2012b., p. 819 - 824.

WALLACE, B. C. *et al.* Modeling annotation time to reduce workload in comparative effectiveness reviews categories and subject descriptors active learning to mitigate

workload. In: 1st ACM International Health Informatics Symposium, 1., 2010, Arlington**. Proceedings…**, Arlington: ACM, 2010, p. 28–35.

WEBSTER, J.; WATSON, R. T. Analyzing the past to prepare for the future: Writing a literature review. **MIS Quarterly**, v.26, n. 2, p. xiii–xxiii, 2002.

XIONG, Z. *et al.* A *machine learning* aided systematic review and meta-analysis of the relative risk of atrial fibrillation in patients with diabetes mellitus. **Frontiers in Physiology**, v. 9, n. JUL, p. 1–12, 2018.

YE, Z. *et al.* SparkText: biomedical text mining on big data framework. **PLoS ONE**, v. 11, n. 9, p. 1–15, 2016.

Recebido: 28/04/2020
Aprovado: 30/07/2020
DOI: 10.3895/rts.v16n45.12119

Como citar: TSUNODA, D.F.; MOREIRA, P.S.C.; GUIMARÃES, A.J.R**.** Machine learning e revisão sistemática de literatura automatizada: uma revisão sistemática. **Rev. Tecnol. Soc**., Curitiba, v. 16, n. 45, p. 337-354, out./dez., 2020. Disponível em: https://periodicos.utfpr.edu.br/rts/article/view/12119. Acesso em: XXX.

Correspondência: