

Avaliação e *Feedback* Automático em Educação apoiada por Tecnologia: Um mapeamento sistemático da literatura

Romário L. Batista¹, Nilmara S. Salgado¹, Raimundo S. Barreto¹

¹Universidade Federal do Amazonas (UFAM) - Instituto de Computação (IComp)

rlb@icompu.ufam.edu.br, nilmarasalgado@gmail.com, rbarreto@icompu.ufam.edu.br

Abstract. *The task of evaluating and classifying the performance of students who are part of the technology-supported education scenario can be a tedious and error-prone task. Providing manual feedback based on these assessments may compromise students' learning. In view of this, it is adopted the automation of these two tasks through the use of computational methods available in the literature. In order to identify these methods, as well as the scenarios in which they are applied, this paper presents a systematic mapping, which classifies 11 methods and 4 scenarios applicable to this context, distributed in 25 papers, and categorized into 3 different types of studies.*

Resumo. *A tarefa de avaliar e classificar o desempenho de estudantes integrantes do cenário de educação apoiada por tecnologia pode ser uma tarefa tediosa e propensa a erros. Fornecer feedback manual com base nessas avaliações pode comprometer o aprendizado desses estudantes. Em vista disso, vem-se adotando a automatização dessas duas tarefas por meio da utilização de métodos computacionais disponíveis na literatura. De modo a identificar esses métodos, bem como os cenários em que são aplicados, este trabalho apresenta um mapeamento sistemático, que classifica 11 métodos e 4 cenários aplicáveis à esse contexto, distribuídos em 25 trabalhos, e categorizados em 3 tipos de estudos diferentes.*

1. Introdução

A tarefa de melhorar nosso sistema educacional, dinâmico e complexo, exige atuação em múltiplas dimensões e decisões fundamentadas, seguras e criativas. Conforme explica [Valente 1999], busca-se melhorias nas condições de atendimento às novas gerações, traduzidas por adequação nos currículos e nos recursos para seu desenvolvimento, num nível tal que provoquem ganhos substanciais na aprendizagem dos estudantes em diversos cenários.

Sob contexto da educação apoiada por tecnologia, usualmente se utilizam ambientes virtuais de aprendizagem (AVA) como meio de adequação desses recursos e automação de tarefas direcionadas aos estudantes. [Silva et al. 2014] classifica dois pontos essenciais à essa adequação: o fator mais importante para que esse tipo de abordagem seja concretizada e aplicada a um cenário real é a presença de *feedback* automático. No entanto, este mostra ser consequência do processo de avaliação feita sobre o desempenho do aluno em cima da sua interação com o AVA.

Avaliação automática diz respeito à corrigir e classificar o desempenho do aluno sobre um determinado tema, utilizando algum tipo de tecnologia ou método, sem

necessidade de intervenção humana. *Feedback* automático refere-se a transformação dos dados obtidos com a avaliação em informações úteis que contribuam com o aprendizado de um estudante.

Como forma de entender e expor o estado-da-arte sobre a automatização desses itens essenciais para o aprendizado de estudantes, métodos utilizados e cenários de aplicação, este trabalho apresenta um mapeamento sistemático da literatura (MSL). Assim, este trabalho é apresentado como segue: a Seção 2 apresenta a metodologia usada no MSL, a Seção 3 apresenta a condução do MSL, a Seção 4 discute os resultados e as análises dos dados e a Seção 5 faz as considerações finais deste MSL.

2. Metodologia do mapeamento sistemático

Para definir o estado da arte com base confiável e sólida ao trabalho, optou-se pelo mapeamento sistemático da literatura (MSL). O MSL é uma forma de identificar, avaliar e interpretar todas as pesquisas disponíveis relevantes para uma questão de pesquisa particular [Kitchenham 2004]. Para esta pesquisa, empregou-se a ferramenta StArt (*State of the Art through Systematic Reviews*)¹ na versão 3.3 desenvolvida pelo Laboratório de Engenharia de Software (LAPES) da Universidade Federal de São Carlos (UFSCAR).

2.1. Objetivo e questões de pesquisa

O objetivo deste MSL pode ser descrito de acordo com o paradigma *Goal-Question-Metric* [Basili et al. 1994], como sendo: Analisar publicações relacionadas a avaliação e geração de *feedback* automático visando identificar as tecnologias, termo utilizado por [Santos et al. 2012] para generalizar métodos, técnicas, modelos, ferramentas, abordagens, linguagens e outros tipos de propostas elaboradas, do ponto de vista do pesquisador no que diz respeito a educação em contexto digital.

Com isso, algumas questões de pesquisa (QP) foram elaboradas para este trabalho, mas a principal QP fundamenta-se no seguinte questionamento: **Quais tecnologias relacionados a avaliação e geração de *feedback* automático podem ser encontrados na literatura dentro do contexto de educação apoiada por tecnologia?**

A QP principal norteia outras QP relacionadas ao objetivo do MSL, são elas:

- QP1: Quais os métodos aplicados para produzir a avaliação automática de estudantes em contexto à educação apoiada por tecnologia?
- QP2: Quais métodos são capazes de gerar *feedback* automáticos baseados nos métodos usados como avaliação automática, satisfazendo critérios previamente definidos?
- QP3: Em que cenários educacionais as tecnologias identificadas para QP1 e QP2 são aplicadas na literatura?

2.2. Estratégia de busca

O ambiente utilizado para execução da revisão foi o de bibliotecas digitais. Logo, optou-se pela biblioteca digital Scopus². Com esta biblioteca foi possível consultar e acessar diferentes fontes digitais via web por meio de expressões de busca pré-estabelecidas.

¹http://lapes.dc.ufscar.br/tools/start_tool

²<https://www.scopus.com>

2.3. Expressão de busca

A expressão de busca, apresentada na Tabela 1 foi contruída com base em palavras chaves que representam População (*P*), Intervenção (*I*) e Resultados (*R*). *P* consiste nos termos relacionados a contextos em que há suporte de tecnologias para educação. *I* consiste nos termos relacionados à abordagens presentes nos trabalhos. Já *R* consiste no que se busca extrair dos trabalhos.

Tabela 1. String de busca utilizada.

Expressão de busca
(“education” OR “educational environment” OR “virtual learning environment” OR “e-learning” OR “e-education” OR “engineering education”) AND (“grading” OR “automatic grading” OR “automatic evaluation” OR “grading system” OR “learning evaluation” OR “computerized grading”) AND (“feedback” OR “automatic feedback”) AND (“approach” OR “method” OR “technique” OR “methodology” OR “formal”)

2.4. Critérios de inclusão e exclusão

Os critérios de inclusão (CI) e exclusão (CE) utilizados para direcionar o MSL ao assunto escolhido QP principal são descritos na Tabela 2.

Tabela 2. Critérios de inclusão e exclusão utilizados no MSL.

	Critérios de Inclusão	Critérios de Exclusão
1	Descrever tecnologias (técnicas, métodos, modelos, ferramentas, abordagens) utilizadas para prover avaliação automática;	Trabalho não é aplicável a nenhum dos critérios de inclusão;
2	Descrever tecnologias (técnicas, métodos, modelos, ferramentas, abordagens) utilizadas para prover <i>feedback</i> automático;	Publicação não disponível;
3	Descrever customização de técnicas ou métodos aplicados a geração de avaliação automática em cenários específicos;	Publicação não atende a delimitação de idioma;
4	Descrever customização de técnicas ou métodos aplicados a geração de <i>feedback</i> automático em cenários específicos.	Publicação não apresenta tecnologia (em uma visão educacional) como população;
5	-	Publicação dispõe apenas de conceitos em seu conteúdo;
6	-	A metodologia usada é restrita a uma área específica.

2.5. Extração de dados

Os principais dados extraídos com o MSL para responder às QP da Seção 2.1 foram: Método usado para produzir avaliação automática; Método usado para gerar *feedback* automático; Cenário de aplicação; Tipo de estudo; Sub-áreas de pesquisa; Limitações do(s) método(s); e Perspectivas futuras. Também foram considerados itens referentes aos dados de publicação (título, autores, ano de publicação e palavras-chave).

Vale ressaltar que os tipos de estudos serão classificados de acordo com a definição de [Wazlawick 2017], havendo portanto três tipos de estudos:

- Pesquisas formais, que exigem a elaboração de uma teoria e de uma prova formal de que essa teoria é correta;
- Pesquisas empíricas, que apresentam uma nova abordagem, comparando-a com outras por meio de testes aceitos pela comunidade;
- Pesquisas exploratórias, que não conseguem provar uma teoria ou resultados estatisticamente aceitos. São comuns dessa categoria os estudos de caso e as análises qualitativas.

2.6. Limitações da metodologia

Duas limitações foram identificadas sobre o MSL. Primeiro, existe a possibilidade de alguma publicação relevante não ter retornado durante a revisão. Em vista disso, optou-se por utilizar a biblioteca digital *Scopus* por esta ser uma das maiores bases de dados de publicações, incluindo editoras internacionais amplamente reconhecidas (como *Springer*³, *IEEE Xplore Digital Library*⁴ e *ACM Digital Library*⁵), [Kitchenham and Brereton 2013] adotaram e recomendam a biblioteca *Scopus*.

Segundo, por mais que apenas uma pessoa tenha extraído os dados de todas as publicações e isto possa ter introduzido viés, outros pesquisadores somaram com sutis contribuições ao revisar os dados reunidos, ajudando a tomar decisões sobre a aceitação de alguns trabalhos. Adicionalmente, o índice de concordância Kappa foi aplicado para mensurar o grau de concordância sobre o MSL.

3. Condução do mapeamento

Finalizado o planejamento da revisão, iniciou-se o processo de condução do MSL em janeiro de 2018. Da busca, 130 publicações foram retornadas. Sobre essas 130 publicações foram aplicados dois filtros de revisão.

Na Figura 1 é possível observar que dos 130 trabalhos retornados e aplicado o primeiro filtro, 82 foram aceitos para o próximo, 45 foram rejeitados e 3 foram considerados duplicados. O primeiro filtro analisa as seguintes informações: título, resumo/*abstract* e palavras-chave. Já no segundo filtro, 25 trabalhos foram aceitos e 57 foram rejeitados, não havendo duplicados durante essa etapa. Essa fase consistiu em uma leitura minuciosa dos trabalhos. Esses 25 trabalhos aceitos passaram então para a fase de extração.

³<http://www.springer.com/br/>

⁴ieeexplore.ieee.org/

⁵<https://www.acm.org/>

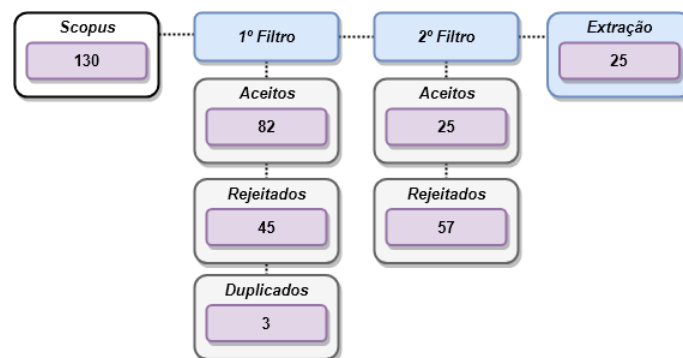


Figura 1. Processo de seleção das publicações.

3.1. Concordância Kappa

Com o intuito de agregar confiabilidade a pesquisa, o método de coeficiente Kappa foi utilizado, com base na configuração apresentada por [Nakagawa et al. 2017], para calcular o grau de concordância entre dois pesquisadores sobre a seleção dos trabalhos retornados por esta revisão, comparando resultados e solucionando conflitos por consenso. A concordância pode variar entre valores abaixo de zero e um. De acordo com [Gisev et al. 2013], o coeficiente Kappa é um dos mais originais e conhecidos índices de concordância utilizado pela academia.

Foram selecionados aleatoriamente 20% do total de trabalhos (26 trabalhos), sobre os quais os dois filtros foram aplicados por dois pesquisadores separadamente. No primeiro filtro obteve-se uma concordância de 0.76 e no segundo filtro obteve-se uma concordância de 0.72. Ambos indicam uma concordância substancial. Com isso, este mapeamento apresenta um índice satisfatório de confiabilidade e precisão.

4. Resultados e análises dos dados

Esta seção detalha os resultados obtidos no MSL, com o objetivo de analisar publicações relacionadas a avaliação e geração de *feedback* automático visando identificar conceitos, metodologias e métodos do ponto de vista do pesquisador no que diz respeito a educação suportada por tecnologia. No que concerne à QP principal descrita na Seção 2.1, de um total de 130 publicações, foram aceitas 25 trabalhos seguindo os critérios de inclusão e exclusão.

A Figura 2 mostra o gráfico do número de publicações aceitas por ano, indicando uma linha de tendência linear crescente com o passar dos anos. Para realização do MSL, o ano de publicação não foi delimitado. Sendo selecionados trabalhos publicados entre 2009 e 2017, demonstrando que as pesquisas nessa área são recentes principalmente por ter o maior índice no ano de 2017.

Foram identificados 11 métodos diferentes que respondem a QP principal. Dentre eles, destacam-se métodos que envolvem aprendizado de máquina, cálculos de similaridade, teste de software e grafos, conforme mostra Figura 3. Alguns métodos são aplicados em conjunto com outra abordagem formando uma customização do método.

Essas customizações são novas metodologias propostas que vinculam dois ou mais métodos com a finalidade de avaliar e gerar *feedbacks* relacionados à essas avaliações.

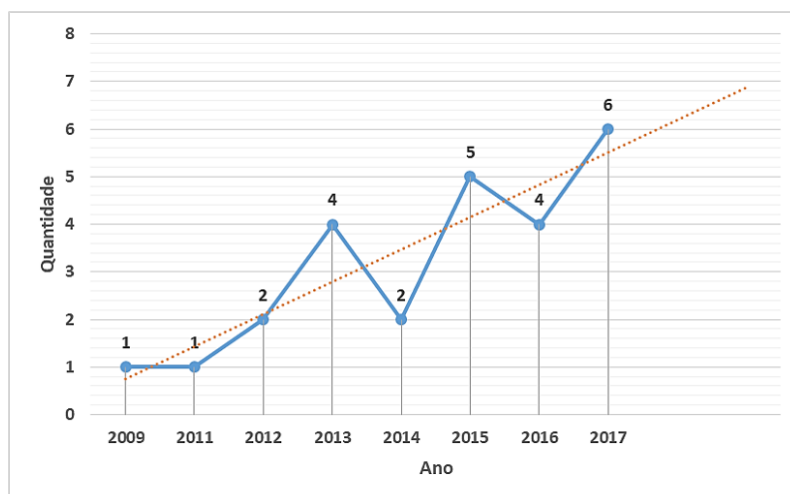


Figura 2. Número de publicações aceitas por ano.

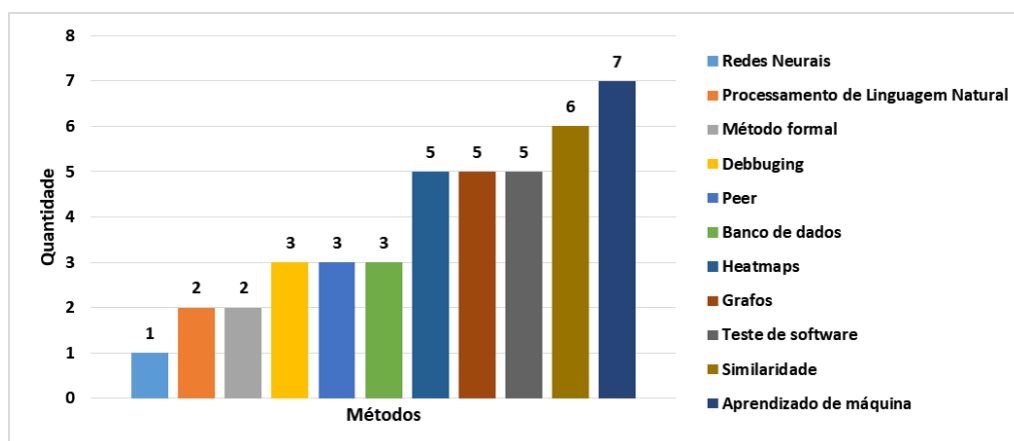


Figura 3. Quantidade das aplicações dos métodos identificados durante o MSL.

QP1: Quais os métodos aplicados para produzir a avaliação automática de estudantes em contexto à educação apoiada por tecnologia?

Esta questão de pesquisa investigou quais dos 11 métodos identificados na QP principal são utilizados para produzir avaliações automáticas de desempenho de estudantes, em função dos 25 trabalhos selecionados.

Como pode ser visto na Figura 4, não são todos os métodos identificados na QP principal que são aplicados a avaliação automática. Nesse caso, somente 10 métodos puderam ser identificados como meios de fornecer avaliação automática. Dando destaque para aprendizado de máquina e cálculos de similaridade, seguidos de grafos e teste de software. Os métodos pouco identificados nas abordagens foram redes neurais e processamento de linguagem natural. No entanto, são as mais recentes, o que nos leva a refletir que ambas podem estar no advento de suas aplicações a esse contexto.

QP2: Quais técnicas são capazes de gerar *feedback* automáticos baseados nos métodos usados como avaliação automática, satisfazendo critérios previamente definidos?

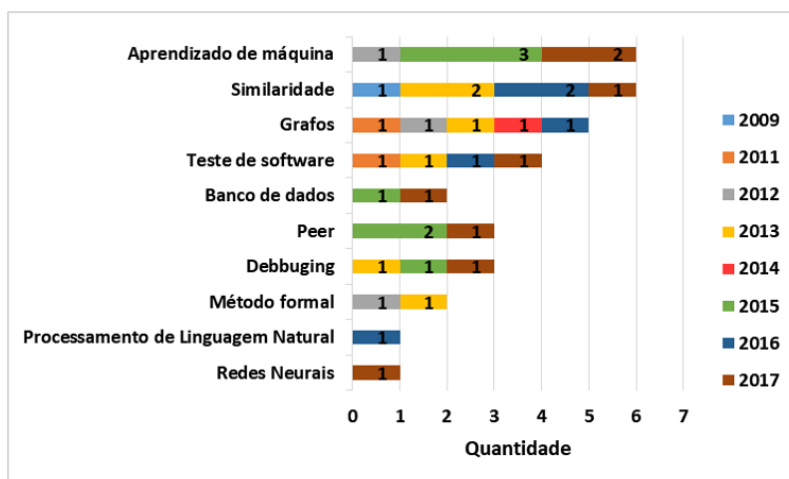


Figura 4. Quantidade dos métodos identificados aplicados ao contexto de avaliações automáticas distribuídos por ano.

Semelhante a QP1, esta teve o objetivo de caracterizar os métodos utilizados para gerar *feedbacks* úteis sobre o desempenho que o estudante obteve, baseado nos critérios de sua avaliação.

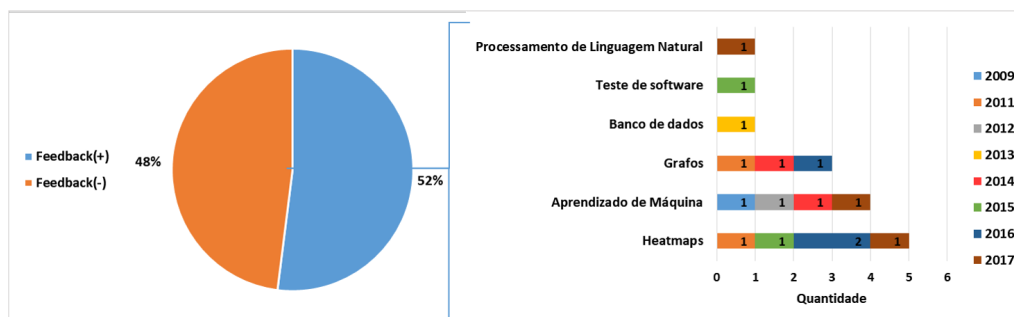


Figura 5. Percentual e quantidade dos métodos identificados aplicados que geram *feedback* automático distribuídos por ano.

Como pode ser visto na Figura 5, dos 25 trabalhos aceitos, 48% identificados como “*Feedback(-)*” são trabalhos que não abordam uma metodologia de geração automática de *feedback*, esses trabalhos apenas apresentam o resultado de suas avaliações para que seja trabalhado pelo responsável competente.

Os outros 52% compõem o grupo dos trabalhos que utilizam um método para gerar *feedback* automático com base na avaliação (distribuídos por ano). Dentro desse grupo foi possível distinguir 6 dos 11 métodos identificados para a QP principal. Dá-se destaque para a técnica de *heatmaps*, seguida de aprendizado de máquina e grafos.

QP3: Em quais cenários educacionais as tecnologias identificadas para QP1 e QP2 são aplicadas na literatura?

Ao analisar as publicações aceitas, e após identificar quais os métodos utilizados para avaliar o desempenho de estudantes e gerar automaticamente *feedback*, foi possível definir quais os cenários de aplicação presentes na literatura.

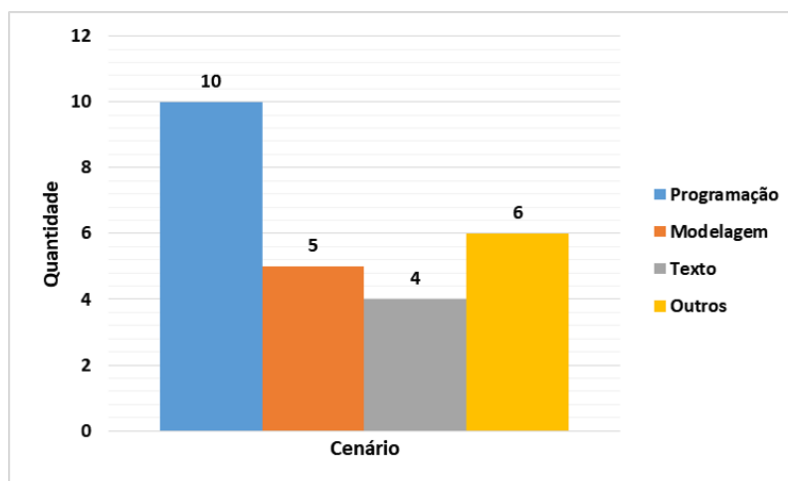


Figura 6. Quantidade dos cenários identificados durante o MSL.

Como pode ser visto na Figura 6, foram identificados 4 grupos de cenários:

- Programação: refere-se aos cenários em que o processo de avaliação e *feedback* automático é realizado sobre atividades que envolvem produção de códigos-fontes, programação, teste e áreas afins. Geralmente associado aos métodos de aprendizado de máquina, grafos, *debugging*, teste de software e similaridade;
- Modelagem: trata dos cenários em que o processo de avaliação e *feedback* automático é realizado sobre atividades de cunho criativo. Essas atividades envolvem esquematização, projeto visual, descrições técnicas, entre outras. Associa-se aos métodos de aprendizado de máquina, grafos, similaridade, banco de dados e *heatmaps*;
- Texto: está relacionado aos cenários cuja a avaliação e o *feedback* automático são aplicados sobre atividades de cunho dissertativo, ou seja, atividades cuja a solução será algum tipo de texto a ser verificado. Tem relação com os métodos de redes neurais, processamento de linguagem natural, aprendizado de máquina e banco de dados;
- Outros: caracteriza os cenários de propósitos gerais. Nesse caso, o processo de avaliação e *feedback* automático é feito com base em atividades cuja taxonomia de execução é definida por questionários de múltipla-escolha. Nesse caso, os métodos podem ser adaptados para a abordagem que se deseja.

4.1. Tipos de estudos

A classificação dos 11 métodos em função do seu tipo de estudo, conforme visto na Seção 2.5 e apresentado pela Figura 7, corrobora que a maior parte dos trabalhos trata de estudos exploratórios, ou seja, não provam uma teoria mas apresentam argumentações qualitativas de seus resultados por meio de estudos de caso.

Em seguida, os estudos empíricos caracterizam a segunda maior quantidade de aplicações. Esses estudos apresentam uma nova abordagem e comparam-na com outras presentes na literatura, além de apresentarem testes estatísticos como evidências de seus resultados.

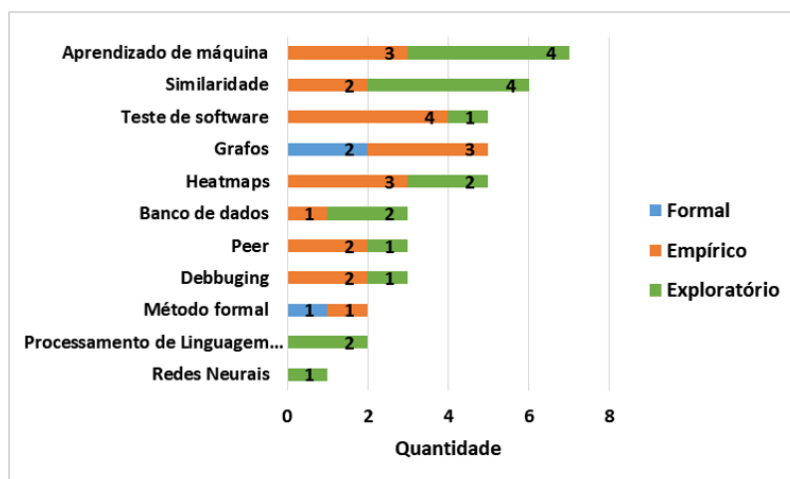


Figura 7. Distribuição dos métodos por tipo de estudo.

Já os estudos formais constituem a menor parcela dos métodos. O que é justificável uma vez que é preciso a elaboração de uma teoria e de uma prova formal de que tal teoria é correta (o que não é trivial), comprovando seus resultados.

Enfatizamos com as observações acima que os estudos exploratórios são mais comuns devido aos métodos utilizados facultativamente exigirem testes estatísticos que comprovem seus argumentos ao contrário das pesquisas empíricas e formais que requerem uma quantificação de seus resultados.

5. Considerações finais

Nesse trabalho foi realizado um mapeamento sistemático da literatura objetivando responder questões de pesquisa relacionadas aos métodos disponíveis que produzam avaliações e gerem *feedbacks* automáticos do desempenho de alunos em contexto a educação apoiada por tecnologia.

Dentre os 11 métodos identificados nos trabalhos analisados destacam-se os métodos de aprendizado de máquina, similaridade, grafos e teste de software para o contexto de avaliações automáticas. Ainda assim, os métodos de processamento de linguagem natural e redes neurais mostraram ser pesquisas mais recentes.

Já relacionado ao *feedback* gerado, o destaque se dá para a técnica conhecida como *heatmaps*, seguida de aprendizado de máquina e grafos. Vale ressaltar que uma parte dos trabalhos não apresentou uma metodologia científica de geração de *feedback*.

Foi identificado que a busca e publicação dentro da área cresce, principalmente se redes neurais estiver de fato em seu despertar nesse contexto. O mapeamento ainda mostrou ser válido por meio da aplicação do coeficiente kappa que obteve um índice substancial de confiabilidade em ambos os dois filtros de seleção aplicados. Contribui-se para a sociedade acadêmica como fonte de busca de trabalhos bases de aplicações de cada um dos métodos identificados.

Referências

Basili, V. R., Caldiera, G., and Rombach, H. D. (1994). Experience factory. *Encyclopedia of software engineering*.

- Gisev, N., Bell, J. S., and Chen, T. F. (2013). Interrater agreement and interrater reliability: key concepts, approaches, and applications. *Research in Social and Administrative Pharmacy*, 9(3):330–338.
- Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University*, 33:1–26.
- Kitchenham, B. and Brereton, P. (2013). A systematic review of systematic review process research in software engineering. *Information and software technology*, 55:2049–2075.
- Nakagawa, E. Y., Scannavino, K. R. F., Fabbri, S. C. P. F., and Ferrari, F. C. (2017). *Revisão Sistemática da Literatura em Engenharia de Software: Teoria e Prática*. Elsevier Brasil.
- Santos, G., Rocha, A. R., Conte, T., Barcellos, M. P., and Prikladnicki, R. (2012). Strategic alignment between academy and industry: a virtuous cycle to promote innovation in technology. In *2012 26th Brazilian Symposium on Software Engineering*, pages 196–200. IEEE.
- Silva, T. T., Valente, J. A., and Dias, P. (2014). Diferentes abordagens da Educação a Distância mediada por computador e via Internet. *Revista Scitis*.
- Valente, J. A. (1999). O computador na sociedade do conhecimento. *Oea Nied/Unicamp*.
- Wazlawick, R. (2017). *Metodologia de pesquisa para ciência da computação*, volume 2. Elsevier Brasil.