

TriMElo: Uma Proposta de Modelo para Avaliação Multi-Habilidades

Fabiana Zaffalon¹, André Prisco¹, Ricardo Souza², Jean Luca Bez³,
Neilor Tonin³, Rafael Penna¹, Silvia Botelho¹

¹Programa de Pós-Graduação em Educação em Ciências
Universidade Federal do Rio Grande (FURG) – Rio Grande – RS – Brasil

²Programa de Pós-Graduação em Modelagem Computacional
Universidade Federal do Rio Grande (FURG) – Rio Grande – RS – Brasil

³Universidade Regional Integrada do Alto Uruguai e das Missões
Erechim – RS – Brasil

{fabinhazaffalon,prisco.c3,rapennas,silviacb.botelho}@gmail.com,

rcrdsou@hotmail.com, {bez, neilor}@urionlinejudge.com.br

Abstract. *The assessment methods that stand out the most are those that present a precise overview of the evolution of the multiple skills of students. In this way, we present the TriMElo model that aims to simultaneously estimate the general and ad sub-skills of students according to the skills involved in the problems. The model is concerned with verifying which are the sub-skills that are decisive for the error or correctness of the question so that its updates occur differently. To validate the model we use a database from the URI Online Judge platform. The empirical results show that the TriMElo model updates the sub-skills more smoothly, according to the wrong or correct answer.*

Resumo. *Os métodos de avaliação que mais se destacam são aqueles que apresentam um panorama preciso da evolução das múltiplas habilidades dos estudantes. Dessa forma, apresentamos o modelo TriMElo que tem por objetivo estimar simultaneamente a habilidade geral e as sub-habilidades dos estudantes de acordo com as habilidades envolvidas nos problemas. O modelo se preocupa em verificar quais são as sub-habilidades determinantes para o erro ou acerto da questão para que suas atualizações ocorram de forma distintas. Para validar o modelo utilizamos uma base de dados da plataforma URI Online Judge. Os resultados empíricos apontam que o modelo TriMElo atualiza as sub-habilidades de forma mais suave, conforme resposta errada ou correta.*

1. Introdução

Com o aumento do uso de plataformas *online* aplicadas ao ensino, os métodos de avaliação educacional que vem se destacando são os que visam apresentar dados precisos sobre a construção das competências ou habilidades adquiridas, não por meio de uma avaliação pontual, mas sim por avaliações que acompanhem o progresso do estudante ao longo do tempo, levando em consideração as várias habilidades envolvidas para a resolução dos problemas, sejam de estudantes presenciais ou à distância.

Muitos exercícios práticos exigem mais de uma habilidade para que possam ser resolvidos. Os estudantes da área de computação, além da habilidade da programação em si referente às particularidades da linguagem de programação, também precisam da habilidade matemática, raciocínio lógico, interpretação, entre outras [Moreira et al. 2018], [Robins 2010]. Para esses estudantes há plataformas que oferecem materiais e exercícios direcionados para o aprendizado de programação.

Sistemas de recomendação incorporados à educação são recursos importantes porque oferecem fácil acesso aos materiais de estudo e auxiliam os estudantes na busca e seleção de um conteúdo focado em seu perfil [Cazella et al. 2012]. Para *sistemas de recomendação educacional* que propõem resoluções de exercícios, um modelo para estimar pontos fortes e fracos no conjunto de habilidades dos estudantes, pode proporcionar uma indicação de exercícios mais adequada, evitando a frustração de problemas demasiadamente difíceis ou o conforto dos muito fáceis [Vargas et al. 2019]. Assim, um primeiro passo em direção a esses sistemas é perceber e representar métricas que permitam traçar a evolução da habilidade e sub-habilidades dos aprendizes de maneira precisa.

Competência pode ser definida como o domínio global ou prático de uma situação cotidiana, e habilidade como domínio de uma operação específica ou ações que atendam à uma ou mais competências [Perrenoud and Magne 1999]. No presente trabalho assume-se como sub-habilidade as várias habilidades necessárias para resolver determinado problema.

Existem vários métodos que estimam as habilidades dos estudantes, entre eles a Teoria de Resposta ao Item (TRI) [Baker 2001] e o *Elo Rating System* [Elo 1978]. A TRI, é um conjunto de modelos matemáticos que procura representar a probabilidade de um indivíduo responder corretamente um item, em função dos parâmetros do item e habilidade do respondente [Andrade et al. 2000], [Baker 2001]. Há modelos de TRI Multidimensional (TRIM), entre eles o modelo compensatório, que contemplam casos em que mais de uma habilidade é exigida [Andrade et al. 2000], [Baker 2001].

Por outro lado, o Elo [Elo 1978], muito utilizado para avaliação em *rankings* internacionais de xadrez, tem como objetivo classificar os jogadores através de seus históricos de jogos, utilizando uma classificação estatística que calcula a habilidade de competidores. Na educação, o Elo estabelece que um estudante é considerado um jogador e o problema é considerado seu adversário [Pelánek 2016]. No entanto tem sido argumentado que o Elo pode ser problemático devido à natureza das habilidades incorporadas nos materiais de aprendizagem, pois destina-se a rastrear apenas uma única habilidade [Park et al. 2019].

O modelo M-ERS proposta por [Park et al. 2019], utiliza um modelo TRIM Compensatório e o modelo Elo para estimar as múltiplas habilidades dos estudantes em sistemas de aprendizagem adaptativa, assumindo que um item pode envolver mais de uma habilidade e que uma baixa habilidade pode ser compensada por outra mais alta. Dessa forma, no presente trabalho, definiu-se habilidade como a habilidade geral em programação, e sub-habilidades como as múltiplas habilidades utilizadas na elaboração da solução.

Apesar desses modelos estimar e acompanhar a evolução das múltiplas habilidades dos estudantes, ao atualizar as habilidades os modelos aumentam ou diminuem seus

valores na mesma proporção de acordo com sua importância no problema, eles não analisam qual das sub-habilidades do estudante foram determinantes para o erro do problema.

No presente artigo propõe-se um modelo, fundamentado no modelo M-ERS, de forma a abordar o problema em que as sub-habilidades sejam atualizadas de formas distintas, não somente de acordo com a importância (ou relevância) das mesmas no problema, mas levando em conta também as sub-habilidades determinantes.

Na próxima seção os modelos TRI, Elo e M-ERS serão detalhados. Na sequência o modelo proposto TriMElo será apresentado. Em seguida, a avaliação do desempenho do modelo através de experimentos utilizando uma base de dados de uma plataforma *Online Judge* do Brasil [Bez et al. 2014]. Por fim, conclusões e trabalhos futuros são apresentados.

2. Teoria de Resposta ao Item (TRI)

A Teoria de Resposta ao Item (TRI) é uma abordagem que vem gradativamente sendo inserida no espaço da educação por ser considerada um instrumento importante no processo quantitativo de avaliação. O modelo permite a mensuração de características do indivíduo, difíceis de serem medidas diretamente [Andrade et al. 2000], [Baker 2001]. No Brasil, o Ministério da Educação adotou o modelo na avaliação do Exame Nacional do Ensino Médio (Enem).

Para o cálculo da estimativa da habilidade θ , a TRI se baseia em métodos estatísticos e modelos matemáticos que consideram não só as respostas dos indivíduos como também as propriedades dos itens [Andrade et al. 2000], [Baker 2001]. Quanto maior a habilidade do indivíduo, maior será a probabilidade de acerto ao item [Baker 2001].

Existem vários modelos propostos da TRI, que dependem de vários fatores [Andrade et al. 2000], [Baker 2001]. Para itens dicotômicos, há 3 modelos que diferem entre si pela quantidade de parâmetros utilizados para descrever o item. São eles: Modelo Logístico de 1 Parâmetro ou Modelo de Rasch (considera somente a dificuldade do item), Modelo Logístico de 2 Parâmetros (considera a dificuldade e a discriminação do item) e Modelo Logístico de 3 Parâmetros (considera a dificuldade, a discriminação e a probabilidade de acerto ao acaso) [Andrade et al. 2000], [Baker 2001].

Esses modelos consideram que o teste seja um instrumento unidimensional que implica na existência ou predominância de apenas uma habilidade, o que não se aplica em muitas situações práticas. Como exemplo pode-se citar um teste de matemática que exige a interpretação de texto antes mesmo de exigir o desenvolvimento matemático e, nesse caso, trata-se de um teste bidimensional, pois exige duas habilidades [Nojosa 2002].

Pesquisas têm mostrado que o modelo TRI Multidimensional (TRIM) se adapta melhor aos dados reais do que os modelos unidimensionais, pois na educação as respostas dos sujeitos são determinadas por mais de uma habilidade ao mesmo tempo [Pasquali 2018].

Os modelos TRIM podem ser separados em duas classes: os compensatórios e os não compensatórios. Um modelo é dito compensatório quando a probabilidade de acerto do item é mantida ou aumentada mesmo que uma das habilidades seja baixa, sendo essa compensada por outra habilidade mais alta [Nojosa 2002].

O modelo compensatório da TRIM é representado pela Equação 1 [Reckase 2006]:

$$P(u_i = 1|\theta_j) = \frac{e^{a_{ik}\theta_{jk}+d_i}}{1 + e^{a_{ik}\theta_{jk}+d_i}} \quad (1)$$

onde u_i é a resposta ao item i ; a_{ik} é o parâmetro de discriminação do item i na dimensão k ; θ_{jk} é o traço latente da pessoa j na dimensão k e d_i é um escalar que indica a dificuldade do item. O expoente e na Equação 1 pode ser escrito de acordo com a Equação 2:

$$a_{i1}\theta_{j1} + a_{i2}\theta_{j2} + \dots + a_{im}\theta_{jm} + d_i \quad (2)$$

3. Sistema de Classificação Elo

Inicialmente, o sistema de classificação Elo foi proposto para analisar e classificar o desempenho de jogadores de xadrez [Elo 1978]. Através de métodos estatísticos, cada jogador recebe uma classificação inicial θ_i e, à medida que vai participando dos jogos, essa classificação vai sendo atualizada de acordo com os resultados. Quando utilizado na educação, o Elo estabelece que um estudante é considerado um jogador e o problema é considerado seu adversário [Pelánek 2016]. A estimação acontece de maneira contínua, pois a atualização da classificação acontece ao término de cada resolução [Pelánek 2016].

O modelo Elo trabalha em função da expectativa e do resultado, a probabilidade esperada de que o jogador vença a partida é dada pela função logística em relação à diferença das classificações estimadas, conforme Equação 3 [Pelánek 2016]:

$$P(R_{ij} = 1) = \frac{1}{1 + 10^{\frac{\theta_j - \theta_i}{400}}} \quad (3)$$

onde $R = \{0,1\}$ é o conjunto de resultados de um jogo: 1 (ganhar) e 0 (perder). Dada uma partida entre o jogador i e o jogador j , com Elo θ_i e θ_j , respectivamente, ao final da partida novos Elos são calculados de acordo com as expectativas dos resultados, os Elos anteriores e uma constante k . Quanto maior o k , maior é a mudança do Elo, conforme Equação 4 [Pelánek 2016]:

$$\theta_i = \theta_i + k(R_{ij} - P(R_{ij} = 1)) \quad (4)$$

O uso do sistema de classificação Elo oferece vantagens, tais como simplicidade de uso em ambientes *online* e implementação em sistemas educacionais, além de apresentar baixo quantitativo de parâmetros para ajustar [Pelánek 2016]. Todavia, destina-se a rastrear apenas uma única habilidade [Park et al. 2019].

Prisco (2018) apresenta uma adaptação do Elo para estimar as múltiplas habilidades para aprendizagem de alunos de programação. O objetivo é fornecer meios para combinar objetos de aprendizagem e estudantes em um ambiente de recomendação com problemas de programação, considerando que cada estudante possui algumas habilidades mais desenvolvidas e outras que precisam ser aprimoradas. No modelo clássico, Elo é um valor escalar para cada aluno e para cada objeto de aprendizado, já o modelo estendido da métrica Elo para torná-la multidimensional, considera que cada dimensão é uma

habilidade que o aluno deve ter em algum nível. Cada estudante tem suas habilidades representadas pela Equação 5:

$$\vec{\theta}_s = (S_1, S_2, \dots, S_N) \quad (5)$$

onde S_i é o Elo na habilidade i . Cada objeto de aprendizagem L tem seu nível de exigência representada por $\vec{\sigma}$ e por relevância \vec{m}_i , onde n é o número de objetos de aprendizagem. O conjunto de objetos de aprendizagem pode ser representado pela Equação 6:

$$L = \{(\vec{\sigma}_1, \vec{m}_1), (\vec{\sigma}_2, \vec{m}_2), \dots, (\vec{\sigma}_n, \vec{m}_n)\} \quad (6)$$

onde relevância m é um número real, entre 0 e 1, que indica quão importante é uma habilidade para interação com um objeto de aprendizado. Considerando a interação do estudante i com objeto de aprendizagem j , a adaptação para o modelo é representada pela Equação 7. Para cada habilidade s :

$$\begin{aligned} \theta_{i_s} &= \theta_{i_s} + m_{j_s} k(R_{ij} - P(R_{ij} = 1)) \\ \sigma_{j_s} &= \sigma_{j_s} + m_{j_s} k(R_{ji} - P(R_{ji} = 1)) \end{aligned} \quad (7)$$

Cada interação é uma tupla $I = (S, L, R)$, onde $R = \{0, 1\}$, com 1 indicando que o estudante acertou o problema ou 0 indicando que ele errou.

4. Modelo M-ERS

O modelo proposto por Park (2019) apresenta uma abordagem que incorpora o modelo TRIM compensatório ao modelo Elo, para rastrear as estimativas dos parâmetros de habilidade em sistemas de aprendizagem adaptativa.

Em vez de assumir um traço unidimensional de respostas a itens, a abordagem assume que um único item pode envolver mais de uma habilidade. Dessa forma, os autores estenderam o Elo padrão, que atualiza uma única habilidade, de forma a permitir uma atualização simultânea de m habilidades distintas. A Equação 8 estima a probabilidade de resposta correta P_{ij} :

$$P_{ij} = P(Y_{ij} = 1) = \frac{e^{(\sum_{m=1}^M \alpha_{jm} \theta_{im} - \beta_j)}}{1 + e^{(\sum_{m=1}^M \alpha_{jm} \theta_{im} - \beta_j)}} \quad (8)$$

onde θ_{im} é a habilidade m do estudante i ($m=1, \dots, M$), α_{jm} é a discriminação do item j correspondente à m dimensão de habilidade e β_j é o nível de dificuldade geral do item j .

A diferença entre o desempenho observado Y_{ij} e o esperado P_{ij} , com base nos modelos TRIM, é usado para atualizar os parâmetros de habilidade após cada resposta de item. P_{ij} dentro do Elo para a m -ésima capacidade do indivíduo i é atualizado de acordo com a Equação 9:

$$\begin{aligned} \hat{\theta}_{im(t)} &= \hat{\theta}_{im(t-1)} + D_{m(t)} K \{Y_{ij(t)} - P_{ij(t)}\} \\ \hat{\beta}_{j(t)} &= \hat{\beta}_{j(t-1)} - D_{m(t)} K \{Y_{ij(t)} - P_{ij(t)}\} \end{aligned} \quad (9)$$

onde $D_{m(t)}$ é um peso para especificar se a habilidade m é indicada pelo item dado na t -ésima etapa. Para a habilidade que é indicada pelo item, $D_{m(t)}$ é igual a 1. Para a habilidade que não é indicada pelo item, o peso leva valores entre 0 e 1. K diminui linearmente, entre 0,4 e 0,1, em função do número total de itens respondidos.

5. Modelo TriMElo

Neste *paper* propõe-se o modelo TriMElo que tem por objetivo estimar simultaneamente a habilidade e as sub-habilidades dos estudantes que utilizam uma plataforma *online* de exercícios de programação. A proposta utiliza o modelo M-ERS como base, porém, antes de atualizar cada sub-habilidade do estudante, identifica-se quais delas são as sub-habilidades determinantes.

Entende-se por determinante aquela sub-habilidade que é a mais provável causa do erro do problema. Tomando-se como exemplo um problema que envolva duas sub-habilidades, “matemática” e “programação”, e considerando-se um estudante que tenha sub-habilidade “matemática” maior que a relevância (importância da sub-habilidade no problema) e sub-habilidade “programação” menor que a relevância no problema: no caso de uma resposta errada podemos pressupor que a sub-habilidade determinante foi “programação”, pois é menor que a necessária para a resolução correta do problema. Assim, compreende-se que as sub-habilidades determinantes devem ter uma atualização diferenciada das demais, de forma que a sub-habilidade “matemática” não sofra uma atualização na mesma proporção que “programação”.

Para estimar a probabilidade de resposta correta do estudante i ao item j alguns parâmetros da Equação 8 do modelo M-ERS foram ajustados, conforme a Equação 10:

$$P_{ij} = P(Y_{ij} = 1) = \frac{e^{(\sum_{m=1}^M \alpha_{jm}(\theta_{im} - \beta_j))}}{1 + e^{(\sum_{m=1}^M \alpha_{jm}(\theta_{im} - \beta_j))'}} \quad (10)$$

onde α_{jm} é a relevância da sub-habilidade m no problema j , θ_{im} é a sub-habilidade m do estudante i e β_j é a dificuldade do item j .

Os parâmetros de habilidade geral dos estudantes e dificuldade dos itens são atualizados conforme Equação 11:

$$\begin{aligned} \hat{\theta}_{i(t)} &= \hat{\theta}_{i(t-1)} + \sigma_j K \{Y_{ij} - P_{ij}\} \\ \hat{\beta}_{j(t)} &= \hat{\beta}_{j(t-1)} + \sigma_j K \{Y_{ji} - P_{ji}\} \end{aligned} \quad (11)$$

onde σ_j é a discriminação do problema j , estimada pela TRI modelo logístico de 2 parâmetros. K é uma constante igual a 0,4 [Wauters et al. 2012], que define o quanto a estimativa pode ser afetada pela diferença entre a resposta atual e a esperada.

Para a atualização de cada sub-habilidade é preciso identificar quais delas são as determinantes, para que elas sejam atualizadas conforme cada caso:

- Sub-habilidade maior que a relevância no problema, Equação 12:

$$\hat{\theta}_{im(t)} = \hat{\theta}_{im(t-1)} + \alpha_{jm} \Phi \{Y_{ij} - P_{ij}\} \quad (12)$$

- Sub-habilidade menor ou igual à relevância no problema, Equação 13:

$$\hat{\theta}_{im(t)} = \hat{\theta}_{im(t-1)} + \alpha_{jm} K \{Y_{ij} - P_{ij}\} \quad (13)$$

onde $\Phi = \alpha_{jm} * K/10, 0$ [Prisco et al. 2018]. O princípio básico é, em vez de utilizar a mesma constante para atualização nos dois casos, utiliza-se valores distintos para cada situação de modo que as sub-habilidades não determinantes sofram um ajuste menor na atualização. A seguir é apresentado o experimento realizado para validação do modelo.

6. Experimento

Para validação do modelo, utilizou-se os dados disponibilizados pela plataforma URI *Online Judge* [Bez et al. 2014], que possui um repositório de problemas de programação onde os usuários ou estudantes submetem suas soluções em uma das linguagens de programação aceitas pela plataforma, que dá um *feedback* de acerto ou erro. Os usuários escolhem os problemas a serem resolvidos e podem submeter várias soluções para o mesmo exercício.

A base de dados foi composta por 20.000 submissões, organizadas em ordem cronológica, de 99 problemas realizadas por 195 usuários da plataforma URI.

Foram adotadas as sub-habilidades de Prisco (2018) : **Matemática** (domínio em álgebra e geometria); **Básico** (envolve problemas sequenciais com operadores condicionais e *loops*); **Modularização** (capacidade de dividir um problema em subproblemas); **String** (processamento de texto, *substring*, pesquisa e outros que especificamente envolvem essa estrutura); **Estruturas Lineares** (pilhas, filas, listas); **Estruturas não Lineares** (domínio e aplicação de grafos e árvores) e **Algoritmo Avançado** (abrange conceitos de recurso computacional e a análise de complexidade do algoritmo). Cada problema recebeu um valor de relevância, entre 0,0 e 1,0, para cada uma das sub-habilidades [Prisco et al. 2018].

Para estimar e calibrar os parâmetros dos problemas, aplicou-se a TRI modelo logístico de 2 parâmetros, que estima a dificuldade e discriminação dos itens. Para isso, os dados foram tabulados de forma a relacionar usuários, problemas e as respectivas respostas (1 - correto ou 0 - incorreto).

De posse dos parâmetros de dificuldade e discriminação, estimados pela TRI, e da relevância das sub-habilidades dos problemas, o próximo passo foi atribuir o valor igual a 0 (zero) aos parâmetros de habilidade e sub-habilidades dos usuários que, posteriormente, foram sendo atualizados a cada submissão, assim como a dificuldade dos problemas.

O modelo proposto TriMElo e o modelo M-ERS foram aplicados na mesma base com o objetivo de comparar o desempenho e resultados gerados, apresentados na próxima seção.

7. Resultados e Discussões

Após a aplicação dos modelos, analisou-se de forma preliminar os resultados obtidos a cada submissão. No presente trabalho tomou-se como o exemplo os dados de um usuário que submeteu 80 soluções para 64 problemas distintos. Dessas submissões, obteve 56 acertos e 24 erros. Analisou-se dois casos desse usuário: uma submissão cuja resposta estava incorreta e, outra, com resposta correta.

Na submissão com a resposta errada, o problema resolvido exige três sub-habilidades: Básico (relevância igual a 1,0), String (relevância igual a 0,3) e Linear (relevância igual a 0,7).

Comparou-se os valores das sub-habilidades do usuário imediatamente antes da atualização (usuário pré) com os valores após a atualização (usuário pós). Estão relacionados na Tabela 1, os valores de relevância das sub-habilidades no problema com os valores das sub-habilidades do usuário (pré e pós) obtidos pelos modelos TriMElo e M-ERS.

	Sub-habilidades TriMElo			Sub-habilidades M-ERS		
	Básico	String	Linear	Básico	String	Linear
Problema	1,0	0,3	0,7	1,0	0,3	0,7
Usuário (Pré)	$\cong 1,326$	$\cong 0,158$	$\cong 0,346$	$\cong 2,812$	$\cong 0,103$	$\cong 0,216$
Usuário (Pós)	$\cong 1,286$	$\cong 0,038$	$\cong 0,0661$	$\cong 2,443$	$\cong -0,008$	$\cong -0,043$
Perda	$\cong 0,040$	$\cong 0,120$	$\cong 0,280$	$\cong 0,369$	$\cong 0,111$	$\cong 0,258$

Tabela 1. Relevância das sub-habilidades do problema e do usuário obtidos pelos modelos TriMElo e M-ERS - Resposta Incorreta

De acordo com os modelos, quando o usuário erra uma resposta os valores de suas sub-habilidades devem diminuir, o que de fato acontece tanto no TriMElo quanto no M-ERS. Analisando cada sub-habilidade do usuário e comparando com a relevância exigida no problema, pode-se supor quais as sub-habilidades contribuíram para que ocorresse o erro da questão. De acordo com a Tabela 1, o usuário possui a sub-habilidade “Básico” maior que a relevância exigida no problema e as sub-habilidades “String” e “Linear” abaixo do que é exigido, logo acredita-se que a sub-habilidade “Básico” não contribuiu para o erro. Assim, tal sub-habilidade não será reduzida na mesma proporção que as demais. Observando os resultados dos dois modelos nota-se que o TriMElo foi o que menos “penalizou” a sub-habilidade.

Na submissão com a resposta correta, o problema resolvido também exigia três sub-habilidades: Básico (relevância igual a 1,0), String (relevância igual a 0,3) e Linear (relevância igual a 0,7), conforme Tabela 2.

	Sub-habilidades TriMElo			Sub-habilidades M-ERS		
	Básico	String	Linear	Básico	String	Linear
Problema	1,0	0,3	0,7	1,0	0,3	0,7
Usuário (Pré)	$\cong 1,297$	$\cong 0,071$	$\cong 0,143$	$\cong 2,705$	$\cong 0,071$	$\cong 0,141$
Usuário (Pós)	$\cong 1,325$	$\cong 0,156$	$\cong 0,341$	$\cong 3,062$	$\cong 0,178$	$\cong 0,391$
Ganho	$\cong 0,028$	$\cong 0,085$	$\cong 0,198$	$\cong 0,356$	$\cong 0,107$	$\cong 0,250$

Tabela 2. Relevância das sub-habilidades do problema e do usuário obtidos pelos modelos TriMElo e M-ERS - Resposta Correta

Com a resposta correta o usuário aumentou suas sub-habilidades nos dois modelos. Novamente a sub-habilidade “Básico” do usuário é superior à exigida pelo problema e as demais são inferiores. Nesse caso, o modelo TriMElo não aumentou o valor da sub-habilidade “Básico” na mesma proporção que as demais, valorizando mais as que estavam abaixo. O que não ocorreu com o modelo M-ERS que entre todas as sub-habilidades a sub-habilidade “Básico” foi a que mais teve ganho.

A Figura 1 apresenta a evolução, nos dois modelos, da sub-habilidade “Básico” que está presente em todos os problemas submetidos pelo usuário. Percebe-se que a discrepância entre os dois modelos ocorre a partir da oitava submissão, onde o usuário ultrapassa o valor 1,0 para a sub-habilidade, mesmo valor que a pertinência do problema em tal submissão. Essa diferença está relacionada à forma com que cada modelo ajusta seus valores; o modelo TriMElo continua variando o valor da habilidade, porém em menor escala.



Figura 1. Evolução da Sub-Habilidade "Básico"

A partir da análise com um usuário, foi realizado o cálculo do coeficiente de correlação de Pearson em todas as submissões. Com o resultado igual a 0,841564 sugere que há uma forte correlação nas sub-habilidades dos dois modelos. Tal correlação se justifica pela maneira que os modelos realizam as estimativas, ambos aumentam as sub-habilidades com os acertos e diminuem com os erros. O que os difere é a forma com que cada um trata a variação das sub-habilidades.

8. Conclusão

No presente trabalho foi proposto o modelo TriMElo, alicerçado no modelo M-ERS, com o objetivo de estimar dinamicamente o progresso das sub-habilidades dos usuários de uma plataforma *online* de programação. O modelo proposto objetiva atualizar as sub-habilidades considerando-se sua influência na resposta além de considerar as sub-habilidades determinantes, de forma a ajustar os seus valores de maneira adequada. Dessa forma, uma resposta correta deve majorar a valorização da habilidade determinante, assim como uma resposta incorreta não deve penalizar em demasia as habilidades não determinantes.

Aplicou-se os dois modelos na mesma base de dados a fim de analisar e comparar o comportamento da evolução das sub-habilidades de um usuário. Identificou-se diferenças nos valores das sub-habilidades estimadas, principalmente nas situações em que a sub-habilidade do usuário é superior à exigida no problema, casos em que o modelo TriMElo atualizou em menor escala.

Acredita-se que o uso do TriMElo em sistemas de recomendação de exercícios de programação possa permitir uma melhor recomendação ao analisar cada sub-habilidade do usuário e as sub-habilidades envolvidas nos exercícios. Assim, como trabalho futuro, visa-se aplicar o modelo TriMElo, através de um estudo de caso, recomendando exercícios de programação para alunos iniciantes nos cursos de computação.

Agradecimentos

Universidade Federal do Rio Grande (FURG) e Instituto Federal de Educação, Ciência e Tecnologia Sul-rio-grandense (IFSul). O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Referências

- Andrade, D., Tavares, H., and Valle, R. (2000). Teoria da resposta ao item: conceitos e aplicações. In *Simpósio Nacional de Probabilidade e Estatística (SINAPE)*.
- Baker, F. (2001). *The basics of Item Response Theory*. ERIC.
- Bez, J. L., Tonin, N. A., and Rodegheri, P. R. (2014). URI Online Judge Academic: A tool for algorithms and programming classes. In *2014 9th International Conference on Computer Science Education*, pages 149–152.
- Cazella, S. C., Behar, P., Schneider, D., da Silva, K. K., and Freitas, R. (2012). Desenvolvendo um sistema de recomendação de objetos de aprendizagem baseado em competências para a educação: relato de experiências. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 23.
- Elo, A. (1978). *The rating of chessplayers, past and present*. Arco Pub.
- Moreira, G. L., Holanda, W., Coutinho, J. C. d. S., and Chagas, F. S. (2018). Desafios na aprendizagem de programação introdutória em cursos de TI da UFRSA, campus Pau dos Ferros: um estudo exploratório. *Anais do Encontro de Computação do Oeste Potiguar ECOP/UFRSA (ISSN 2526-7574)*, (2).
- Nojosa, R. T. (2002). Teoria da Resposta ao Item (TRI): modelos multidimensionais. *Estudos em Avaliação Educacional*, (25):123–166.
- Park, J. Y., Cornillie, F., van der Maas, H. L., and Van Den Noortgate, W. (2019). A multidimensional irt approach for dynamically monitoring ability growth in computerized practice environments. *Frontiers in psychology*, 10.
- Pasquali, L. (2018). *TRI—Teoria de resposta ao item: Teoria, procedimentos e aplicações*. Editora Appris.
- Pelánek, R. (2016). Applications of the Elo rating system in adaptive educational systems. *Computers Education*, 98:169–179.
- Perrenoud, P. and Magne, B. (1999). *Construir as competências desde a escola*. Artmed, Porto Alegre.
- Prisco, A., Santos, R., Botelho, S., Tonin, N., and Bez, J. (2018). A multidimensional Elo model for matching learning objects. In *2018 IEEE Frontiers in Education Conference (FIE)*.
- Reckase, M. D. (2006). Multidimensional item response theory. *Handbook of statistics*, 26:607–642.
- Robins, A. (2010). Learning edge momentum: A new account of outcomes in cs1. *Computer Science Education*, 20(1):37–71.
- Vargas, A. P., dos Santos, R. A. P., Bez, J., Tonin, N., and da Costa Botelho, S. S. (2019). Um modelo de mediação pedagógica para ambientes massivos. *RENOTE-Revista Novas Tecnologias na Educação*, 17(1):93–102.
- Wauters, K., Desmet, P., and Van Den Noortgate, W. (2012). Item difficulty estimation: An auspicious collaboration between data and judgment. *Computers & Education*, 58(4):1183–1193.