

UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ARTES, CIÊNCIAS E HUMANIDADES
PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO

PATRICIA GILAVERT FERNANDES

**Testes Adaptativos Computadorizados como um Processo de Decisão
Markoviano**

São Paulo

2020

Resumo

FERNANDES, Patricia Gilavert. **Testes Adaptativos Computadorizados como um Processo de Decisão Markoviano**. 2020. 38 f. Qualificação (Mestrado em Ciências) – Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2020.

Os Testes Adaptativos Computadorizados (Computerized Adaptive Testing - CAT) avaliam os indivíduos de forma adaptativa por meio de itens selecionados sequencialmente durante o teste, após cada resposta dada, de acordo com critérios de seleção de itens e critérios de parada do teste predefinidos. Esse tipo de avaliação se destaca em relação aos testes aplicados de forma convencional (papel e lápis) por apresentar um melhor compromisso entre precisão na estimação das habilidades e o tempo do teste. Em geral, os CATs consideram um modelo probabilístico de resposta para cada item, possibilitando encontrar a estimativa da habilidade do indivíduo; os modelos probabilísticos mais comuns são os modelos de Teoria de Resposta ao Item (TRI), que permitem descrever parâmetros dos itens, como dificuldade e discriminação. Tradicionalmente, os critérios de seleção de itens utilizados nos CATs são míopes, isto é, avaliam o próximo melhor item sem considerar os itens que ainda estão por vir em conjunto; além disso, consideram um critério de parada independente do critério de seleção de itens. Uma abordagem não míope que possibilita modelar os critérios de seleção e critérios de parada juntos são os Processos de Decisão Markovianos (Markov Decision Process - MDP). A proposta deste trabalho visa construir um CAT utilizando MDP para obter um melhor compromisso entre a qualidade da estimativa da habilidade dos indivíduos e a quantidade de itens realizados. O CAT baseado em MDP será avaliado em base de dados sintéticas e base de dados de testes reais; espera-se que o método proposto apresente resultados melhores no que se refere a número de itens e qualidade da estimativa da habilidade dos indivíduos quando comparado com CATs construídos de forma clássica.

Palavras-chaves: Testes Adaptativos Computadorizados, Teoria de Resposta ao Item, Processo de Decisão Markoviano.

Lista de figuras

Figura 1 – Fluxograma de um CAT.	12
Figura 2 – Sistema modelado como MDP.	14
Figura 3 – Fluxograma da abordagem ALICAT.	16
Figura 4 – Sistema de teste de domínio sequencial com RL.	18
Figura 5 – Critérios de seleção de item, com parâmetro b normal padrão e uniforme(-3,3), para habilidades baixas e critério de parada número de questões. .	25
Figura 6 – Critérios de seleção de item, com parâmetro b normal padrão e uniforme(-3,3), para habilidades medianas e critério de parada número de questões. .	25
Figura 7 – Critérios de seleção de item, com parâmetro b normal padrão e uniforme(-3,3), para habilidades altas e critério de parada número de questões. . .	26
Figura 8 – Critérios de seleção de item, com parâmetro b normal padrão e uniforme(-3,3), para habilidades baixas e critério de parada variância <i>a posteriori</i> . .	26
Figura 9 – Critérios de seleção de item, com parâmetro b normal padrão e uniforme(-3,3), para habilidades medianas e critério de parada variância <i>a posteriori</i> . .	27
Figura 10 – Critérios de seleção de item, com parâmetro b normal padrão e uniforme(-3,3), para habilidades altas e critério de parada variância <i>a posteriori</i> . .	27
Figura 11 – Critérios de seleção de item, com parâmetro b normal padrão e uniforme(-3,3), para habilidades baixas e critério de parada Δ variância.	28
Figura 12 – Critérios de seleção de item, com parâmetro b normal padrão e uniforme(-3,3), para habilidades medianas e critério de parada Δ variância. . . .	28
Figura 13 – Critérios de seleção de item, com parâmetro b normal padrão e uniforme(-3,3), para habilidades altas e critério de parada Δ variância.	29
Figura 14 – Critérios de parada para habilidade baixa, informação de Fisher como critério de seleção de itens e parâmetro b com distribuição normal padrão. .	29
Figura 15 – Critérios de parada para habilidade mediana, informação de Fisher como critério de seleção de itens e parâmetro b com distribuição normal padrão.	30
Figura 16 – Critérios de parada para habilidade alta, informação de Fisher como critério de seleção de itens e parâmetro b com distribuição normal padrão. .	31
Figura 17 – Ilustração dos estados no CAT com MDP.	32

Lista de tabelas

Tabela 1 – Cronograma do projeto.	35
---	----

Lista de abreviaturas e siglas

CAT	Computerized Adaptive Testing
EAP	Esperança <i>a posteriori</i>
ENEM	Exame Nacional do Ensino Médio
EP	Erro Padrão
MAP	Moda <i>a posteriori</i>
MDP	Markov Decision Process
ML3	Modelo logístico com três parâmetros
MMV	Método de Máxima Verossimilhança
MSE	Mean Squared Error
POMDP	Partially Observable Markov Decision Process
RL	Reinforcement Learning
SMT	Sequential Mastery Testing
TCT	Teoria Clássica do Teste
TRI	Teoria de Resposta ao Item

Lista de símbolos

θ	Escore verdadeiro. Estimativa do nível de habilidade dos indivíduos considerando a prova completa.
$\hat{\theta}$	Escore parcial. Nível de habilidade dos avaliados estimados durante a execução do CAT.
a	Parâmetro de discriminação do item
b	Parâmetro de dificuldade do item
c	Probabilidade de acerto aleatório do item
D	Constante numérica para que a função logística forneça resultados semelhantes com a da função ogiva normal
π	Política
s	Estado do processo
a	Ação que pode ser executada

Sumário

1	Introdução	7
1.1	<i>Objetivos</i>	8
1.2	<i>Hipóteses</i>	8
1.3	<i>Organização do documento</i>	8
2	Conceitos fundamentais	9
2.1	<i>Teoria de resposta ao item</i>	9
2.2	<i>Testes adaptativos computadorizados</i>	11
2.3	<i>Processo de Decisão Markoviano</i>	13
3	Trabalhos correlatos	16
4	Avaliação de compromissos	22
4.1	<i>CrITÉrios de seleção de itens</i>	23
4.2	<i>CrITÉrios de parada</i>	24
4.3	<i>Resultados</i>	24
4.4	<i>Conclusões</i>	30
5	Proposta de Pesquisa	32
5.1	<i>Tema</i>	32
5.2	<i>Motivação</i>	33
5.3	<i>Lacuna/problema</i>	33
5.4	<i>Justificativa</i>	33
5.5	<i>Método</i>	33
5.5.1	<i>Atividades</i>	34
5.5.2	<i>Cronograma</i>	34
5.6	<i>Avaliação</i>	35
5.7	<i>Contribuição</i>	35
5.8	<i>Escopo</i>	36
	Referências¹	37

¹ De acordo com a Associação Brasileira de Normas Técnicas. NBR 6023.

1 Introdução

As avaliações educacionais tem o propósito de coletar, interpretar e aferir resultados de grandes grupos populacionais, visando à certificação educacional ou ao ingresso em determinadas instituições. Com o avanço da tecnologia computacional nas últimas décadas, novas formas de avaliação têm ganhado destaque, como é o caso dos testes adaptativos computadorizados (Computerized Adaptive Testing - CAT). Esses testes adaptam a avaliação ao nível de conhecimento do respondente e apresentam inúmeras vantagens sobre os testes aplicados de forma convencional, denominados de papel e lápis (SPENASSATO; BORNIA; TEZZA, 2015).

O CAT é uma abordagem de medição que gera testes sob medida para os examinados, em tempo real, com base em suas respostas aos itens anteriores. Ao administrar itens que são mais informativos para um examinado, esses testes fornecem uma medida precisa da proficiência de um examinado com relativamente poucos itens (STAFFORD *et al.*, 2019).

Normalmente, utiliza-se a Teoria de Resposta ao Item (TRI) para encontrar a probabilidade de um indivíduo obter um escore em um item em função de sua habilidade. Essa probabilidade é influenciada pelos parâmetros dos itens, como dificuldade e discriminação.

Para construir um CAT, dois critérios fundamentais precisam ser definidos: critério de seleção de itens e critério de parada do teste. No geral, esses critérios são definidos previamente e de forma independente. Nos CATs clássicos, observa-se que os critérios de seleção de itens avaliam o próximo melhor item sem considerar os itens que ainda estão por vir.

O CAT pode ser entendido como um processo de decisão sequencial por selecionar itens sequencialmente durante o teste, após cada resposta dada, de acordo com os critérios de seleção de itens e critérios de parada. Uma abordagem muito utilizada para modelar problemas de decisão sequencial, que consegue ser não míope, olhando um horizonte longe de tempo são os Processos de Decisão Markovianos (Markov Decision Process - MDP). Dependendo da forma como é modelado, o MDP também consegue otimizar a escolha dos critérios de seleção de itens e critério de parada juntos.

Construir CAT como um processo de decisão MDP, aplicando técnicas de aprendizado por reforço (Reinforcement Learning - RL), pode melhorar a estimativa da habilidade

dos indivíduos e diminuir o número de itens quando comparado com CATs construídos de forma clássica.

1.1 Objetivos

O objetivo principal deste trabalho é construir um CAT como um MDP. Pretende-se aplicar o estudo para obter estimativa da habilidade $\hat{\theta}$ mais próxima da real habilidade θ com menor quantidade de itens possíveis.

Como objetivo específico, pretende-se fazer comparação dos resultados com dados simulados e reais.

1.2 Hipóteses

O uso do MDP na construção de um CAT proporciona vantagens, tais como a diminuição do erro de estimação da habilidade do aluno e a diminuição no número de itens de uma prova, comparadas com CATs clássicos.

Os métodos clássicos são míopes com relação a um horizonte longo, tomando decisões com base no ganho imediato. MDPs permitem olhar para um horizonte longo, realizando um compromisso entre ganhos imediatos e ganhos no futuro. Além disso, o MDP trabalha para encontrar o melhor critério de escolha do próximo item e melhor critério de parada ao mesmo tempo.

Esta hipótese será testada medindo-se as curvas de precisão da habilidade do aluno por itens realizados e a quantidade de itens necessários para atingir uma precisão predeterminada no CAT proposto e em um conjunto de CATs construídos de forma clássica.

1.3 Organização do documento

Este trabalho está organizado da seguinte forma: os conceitos fundamentais necessários para a compreensão da proposta são apresentados no Capítulo 2. No Capítulo 3 são apresentados os trabalhos correlatos, o Capítulo 4 apresenta uma avaliação de compromissos e o Capítulo 5 apresenta a proposta de pesquisa.

2 Conceitos fundamentais

Este capítulo apresenta os conceitos fundamentais relacionados ao escopo deste projeto de pesquisa, sendo dividido em três partes: Teoria de Resposta ao Item; Testes Adaptativos Computadorizados; e Processo de Decisão Markoviano.

2.1 Teoria de resposta ao item

A Teoria da Resposta ao Item (TRI) é uma metodologia que representa a relação entre a probabilidade de um indivíduo dar uma resposta a um item e seus traços latentes (características do indivíduo que não podem ser observadas diretamente).

A TRI trata de modelos matemáticos que procuram descrever a probabilidade de um indivíduo obter um certo escore em um item em função do nível do traço latente (ou dos traços latentes). Essa probabilidade também é influenciada por parâmetros do item, como dificuldade e capacidade de discriminação.

Os modelos ML1, ML2 e ML3 são muito utilizados para o cálculo da probabilidade de acerto para dados com respostas dicotômicas (ou seja, certo ou errado). O ML1, também chamado de modelo Rasch, possui uma dimensão que considera apenas o parâmetro de dificuldade do item (RASCH, 1960). O ML2 é o modelo logístico unidimensional com os parâmetros de dificuldade e discriminação (BAKER; KIM, 2004). O ML3 é o modelo logístico com três parâmetros, dificuldade, discriminação e probabilidade de acerto ao acaso (BIRNBAUM, 1968).

O modelo mais usual na literatura atualmente é o modelo ML3, dado por:

$$\Pr(X_{ij} = 1 \mid \theta_j; a_i, b_i, c_i) = c_i + \frac{(1 - c_i)}{1 + \exp[-Da_i(\theta_j - b_j)]}, \quad (1)$$

na qual,

- X_{ij} é uma variável dicotômica que assume o valor 1 se o aluno j responder corretamente ao item i e 0, caso contrário;
- $\Pr(X_{ij} = 1 \mid \theta_j)$ é a probabilidade do indivíduo j responder corretamente ao item i ;
- θ_j é o traço latente do indivíduo j , ou seja, a habilidade do indivíduo j ;
- a_i é o parâmetro de discriminação do item i ;
- b_i é o parâmetro de dificuldade do item i ;
- c_i é a probabilidade de acerto aleatório do item i ; e

- D é um fator de escala igual a 1 para o modelo logístico e 1,702 para que a função logística forneça valores parecidos com a da função ogiva normal.

Os modelos ML1 e ML2 são casos particulares do modelo ML3. Assim, quando $c_i = 0$ chega-se no modelo ML2 e quando, além de $c_i = 0$, o parâmetro de discriminação a_i for igual para todos os itens chega-se no modelo ML1.

A estimação dos parâmetro dos itens e estimação da habilidade geralmente são feitos pelo Método de Máxima Verossimilhança (MMV), por meio de um processo iterativo (Newton-Raphson, por exemplo), ou por procedimentos Bayesianos. Porém, o MMV apresenta problemas na estimação de itens que são respondidos corretamente, ou incorretamente, por todos os indivíduos, e também das habilidades de indivíduos que responderam corretamente, ou incorretamente, a todos os itens. Além disso, há a possibilidade de que as estimativas dos parâmetros dos itens caiam fora do intervalo esperado, tal como valores de a_i negativos, ou valores de c_i fora do intervalo $[0; 1]$ (ANDRADE; TAVARES; VALLE, 2000).

A estimação Bayesiana é uma alternativa para a solução desses problemas e consiste em estabelecer distribuições *a priori* para os parâmetros, construir uma distribuição *a posteriori* e estimar os parâmetros de interesse com base em alguma característica dessa distribuição. Os métodos Bayesianos mais utilizados para estimar os parâmetros são o da média *a posteriori* (EAP), que utiliza a média da distribuição *a posteriori*; e o da moda *a posteriori* (MAP), que utiliza a moda da distribuição *a posteriori*.

No caso da estimação da habilidade, define-se a distribuição *a priori* $f(\theta)$ para θ e após uma pergunta q , obtém-se uma resposta R ($R = 1$ para resposta correta e $R = 0$ para resposta incorreta) e por meio do modelo de TRI, é possível calcular $\Pr(R|\theta, q)$. Outra informação *a priori* que deve ser considerada é o modelo de itens (os itens podem ser dicotômicos, por exemplo).

Seguindo o teorema de Bayes, em que $\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}$ sendo A e B eventos, é possível calcular a distribuição *a posteriori* para θ :

$$f(\theta|q, R) = \frac{\Pr(R|\theta, q)f(\theta)}{\Pr(R|q)} \propto f(\theta) \Pr(R|\theta, q). \quad (2)$$

2.2 Testes adaptativos computadorizados

Dada a existência de uma metodologia consolidada e efetiva como a TRI para a elaboração de provas mais eficazes e justas com o desempenho do estudante, é possível construir um CAT baseado na TRI.

O CAT apresenta diversas vantagens, tais como: precisão na estimação da habilidade; redução de custos com materiais para o desenvolvimento de testes, armazenamento e correção; o desempenho do examinado guia o seu próprio conjunto de itens administrados; o tempo de teste é reduzido e mantém o mesmo nível de confiança de um teste convencional; pode ser utilizado como ferramenta para a educação a distância; e não requer que todos os examinados realizem a prova simultaneamente.

O método pode ser utilizado para avaliação de indivíduos e essa avaliação pode ser uma classificação ou uma regressão. Para classificação de indivíduos a resposta do CAT pode ser, por exemplo, aprovado ou reprovado, nível alto, médio ou baixo e para regressão é utilizado para dar um score ou uma nota para um indivíduo em um teste. Algumas avaliações em larga escala, já implementadas e conhecidas no mundo são: Graduate Record Examination (GRE), Graduate Management Admission Test (GMAT), Armed Services Vocational Aptitude Battery (ASVAB), Test of English as a Foreign Language (TOEFL), European Computer Driving Licence (ECDL), National Assessment of Educational Progress (NAEP) (SPENASSATO; BORNIA; TEZZA, 2015).

Os CATs são aplicados de forma adaptativa aos respondentes, via computador, cujos itens são selecionados sequencialmente durante o teste, após cada resposta dada a um item, de acordo com regras predefinidas do algoritmo (SPENASSATO; BORNIA; TEZZA, 2015). Segundo Van Linden e Glas (2000), um CAT clássico pode ser descrito pelos seguintes passos:

1. O primeiro item é selecionado;
2. Estima-se o traço latente baseado na resposta ao primeiro item;
3. Escolhe-se o próximo item a ser respondido. Tal item deve ser o mais adequado para a estimativa pontual interina da habilidade;
4. Recalcula-se a estimativa do traço latente baseado nas respostas anteriores
5. Repete-se os passos 3 e 4 até não ser mais necessária nenhuma resposta, segundo algum critério preestabelecido, chamado de critério de parada.

Em geral, o fluxograma do CAT é como na Figura 1.

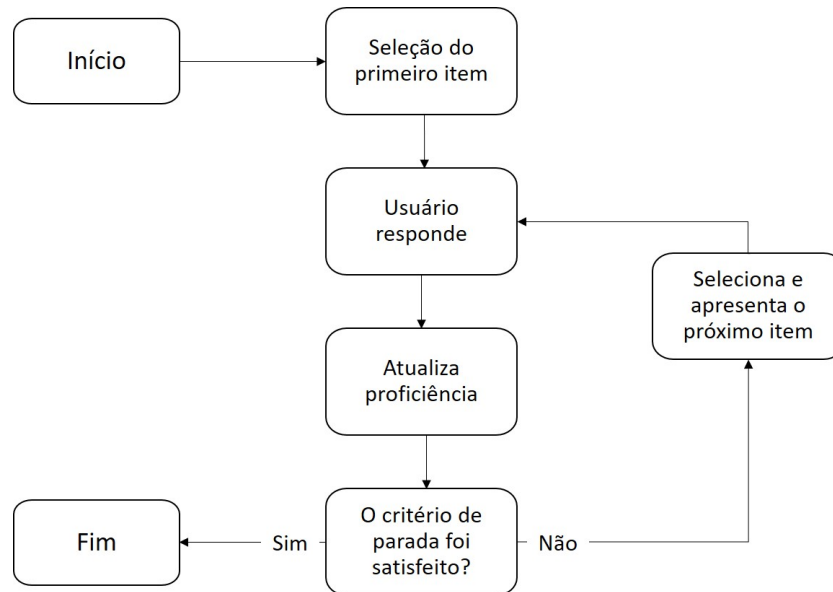


Figura 1 – Fluxograma de um CAT.

Fonte: Elaborado pela autora (2020).

O banco de itens deve ter informações suficientes para fornecer as informações necessárias sobre o teste e deve fornecer itens com diferentes níveis de dificuldade que consigam cobrir todos os tópicos que serão avaliados. Muitos trabalhos sugerem que em um CAT adequado, o número de itens deve ser de 5 a 10 vezes o tamanho da prova que se pretende aplicar.

A escolha do método de seleção de itens influencia a eficiência e precisão na estimativa da habilidade dos indivíduos. Existem duas abordagens de seleção consolidadas na área, uma baseada em informação e outra Bayesiana. As abordagens baseadas em informação selecionam itens que possuam maior informação a respeito da estimativa de um θ específico. Na abordagem Bayesiana, seleciona-se itens baseados em distribuições *a priori* e *a posteriori* das estimativas de um θ específico (JATOBÁ *et al.*, 2020). Os métodos de seleção de itens mais utilizados na literatura são: informação de Fisher, informação Intervalar de Fisher, informação ponderada pela máxima verossimilhança, informação ponderada *a posteriori* e informação de Kullback-Leibler.

Diferenças nos critérios de seleção de itens podem levar a diferentes escolhas de itens para os mesmos candidatos e, conseqüentemente, a diferentes decisões de classificação. A escolha do método de seleção de item é, portanto, uma das partes mais importantes do CAT.

A estimação dos parâmetros dos itens e das habilidades é comumente feita pelo MMV, máxima verossimilhança ponderada, esperança *a posteriori* ou a moda *a posteriori*.

Os itens iniciais podem ser selecionados a partir de um valor médio para o θ inicial quando não se sabe nada a respeito do respondente. Podem também iniciar a habilidade pela média da distribuição da habilidade da população ou estabelecer um valor aleatório (JATOBÁ *et al.*, 2020). Algumas aplicações podem submeter o usuário a um conjunto de itens iniciais fixos, e a partir das respostas, as habilidades são estimadas para que os próximos itens sejam adaptados. O teste também pode ser inicializado considerando um conjunto de informações *a priori*. Em geral, nas avaliações educacionais o primeiro item é variável, para evitar compartilhamento de respostas entre os participantes.

Os critérios de parada mais utilizados na literatura são: atingiu o limite predefinido de itens do banco de itens, atingiu um número fixo de itens a serem aplicados; atingiu um limite de variação do erro padrão da habilidade do respondente entre um item e outro ou atingiu o tempo limite. Alguns CATs podem adotar mais de um critério de parada.

A escolha do método mais adequado para cada componente de um CAT irá depender do contexto do teste, do modelo de TRI utilizado e das restrições que precisam ser incluídas (SPENASSATO; BORNIA; TEZZA, 2015).

De modo geral, os trabalhos que propõe construção de CAT funcionam de modo *greedy* com base em uma função score e não levam em conta que serão respondidos outros itens após o que está sendo avaliado.

2.3 Processo de Decisão Markoviano

Como mencionado, os critérios de seleção de itens utilizados nos CATs clássicos são míopes, ou seja, avaliam o próximo melhor item sem considerar os próximos itens, e o critério de parada é fixo. Uma forma de não ser míope e modelar os critérios de seleção e critérios de parada juntos é tratando o problema como um MDP.

Um MDP é um modelo de um agente que interage em sincronia com um ambiente. O agente recebe como entrada o estado do ambiente e gera como saída ações, que por sua vez afetam o estado do ambiente. Na estrutura do MDP, presume-se que, embora possa haver uma grande incerteza sobre os efeitos das ações de um agente, nunca há qualquer

incerteza sobre o estado atual do agente, por ter habilidades perceptivas completas e perfeitas (KAELBLING; LITTMAN; CASSANDRA, 1998).

O MDP pode ser descrito por uma tupla (S, A, T, R) , onde:

- S é um conjunto de estados em que o processo pode estar;
- A é um conjunto de ações que podem ser executadas em diferentes épocas de decisão;
- $T : S \times A \times S \mapsto [0,1]$ é uma função que dá a probabilidade de o sistema passar para um estado $s' \in S$, dado que o processo estava em um estado $s \in S$ e o agente decidiu executar uma ação $a \in A$ ($T(s'|s,a)$);
- $R : S \times A \mapsto \mathbb{R}$ é uma função que dá o recompensa (ou custo) por tomar uma decisão $a \in A$ quando o processo está em um estado $s \in S$.

Na propriedade de Markov, o próximo estado e a recompensa esperada dependem apenas do estado anterior e da ação realizada. O estado e a recompensa no tempo $t + 1$ dependem apenas do estado no tempo t e da ação no tempo t .

A Figura 2 mostra um sistema modelado como um MDP. O tomador de decisões verifica o estado atual do sistema (s), consulta uma política (π) e executa uma ação (a). Esta ação pode ter um efeito sobre o ambiente e modificar o estado atual. O tomador de decisões, então, verifica o novo estado para que possa tomar a próxima decisão.

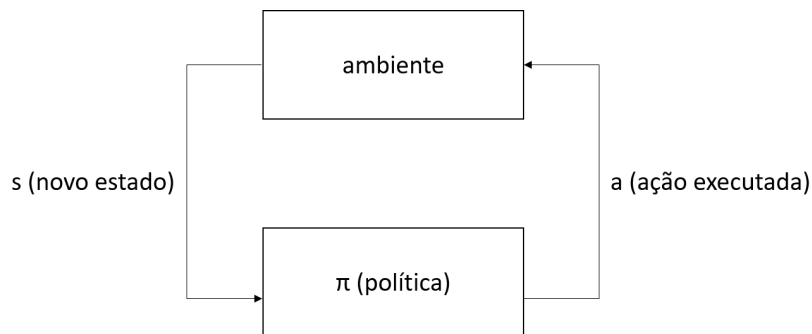


Figura 2 – Sistema modelado como MDP.

Fonte: Adaptado de Pellegrini e Wainer (2007).

A cada época de decisão, o tomador de decisões usa uma regra de decisão para escolher a próxima ação. O conjunto de todas as regras de decisão (uma para cada época de decisão) é chamado de política π . O MDP tem como objetivo encontrar a política ideal que otimiza uma determinada métrica de desempenho, resultando em maiores recompensas dado cada estado.

Uma recompensa imediata, que pode ser positiva, negativa ou zero, é obtida na transição de um estado para outro sob a influência de uma ação. A métrica de desempenho de uma política é geralmente uma função (a função objetivo) das recompensas imediatas obtidas quando a política associada é seguida ao longo de um horizonte de tempo predeterminado. O horizonte de tempo pode ser finito ou infinito (GOSAVI, 2009).

A função valor de uma política π para um MDP é uma função $V^\pi : S \mapsto \mathbb{R}$, tal que $V^\pi(S)$ fornece o valor esperado da recompensa para a política, de acordo com um critério de desempenho (ou de "otimalidade").

Quando o estado atual do sistema não é conhecido tem-se um MDP parcialmente observável (Partially Observable Markov Decision Process - POMDP) que é uma generalização de MDPs (SILVER; VENESS, 2010).

Um POMDP pode ser descrito por (S, A, T, R, Ω, O) , onde:

- S , A , T e R descrevem MDP;
- Ω é um conjunto de observações que são obtidas em cada época de decisão; e
- $O : S \times A \times \Omega \mapsto [0,1]$ é uma função que dá a probabilidade de uma observação o ser verificada, dados um estado s e a última ação a executada.

O POMDP faz uma observação com base na ação e no estado resultante. O objetivo do agente continua sendo maximizar a recompensa futura. O agente faz observações e gera ações. Ele mantém um estado de crença (*belief state*) interno, que resume sua experiência anterior. O estimador de estado é responsável por atualizar o estado de crença com base na última ação, na observação atual e no estado de crença anterior. A política, como antes, é responsável por gerar ações, mas desta vez em função do estado de crença do agente, e não do estado do ambiente (KAELBLING; LITTMAN; CASSANDRA, 1998).

Em um POMDP não há a possibilidade de se observar diretamente o estado em que o sistema está em um dado momento, ou seja, o tomador de decisões não sabe o estado atual s (ao contrário do que acontece em um problema modelado como MDP). No entanto, cada ação tem como resultado alguma observação que é probabilisticamente relacionada ao estado do sistema.

3 Trabalhos correlatos

Este capítulo trata de estudos encontrados na literatura com abordagens semelhantes as propostas aqui e com resultados promissores. Esses trabalhos motivam o desenvolvimento da pesquisa envolvendo CAT e MDP.

O trabalho em Jatobá *et al.* (2020) cria uma abordagem que personaliza o processo de seleção de itens no CAT, considerando o uso de mais de um critério. Após a análise do desempenho de diferentes critérios por meio de um estudo de caso, chega-se a conclusão de que a regra de seleção de Kullback-Leibler com distribuição *a posteriori* (KLP) possui o melhor desempenho na estimativa das habilidades em relação a outras regras testadas. Observou-se também que as regras de Informação de Fisher (F) e Informação Ponderada pela Máxima Verossimilhança (MLWI) tiveram melhor desempenho nos estágios iniciais do CAT, para estimar respondentes com níveis de habilidades extremos negativos e positivos, respectivamente. O critério de parada foi definido com base no ponto de estabilidade, ou seja, o momento do teste no qual a diferença do erro padrão entre o item atual aplicado e o item anterior é inferior a 1% do erro padrão do item anterior. A Figura 3 mostra o fluxograma da abordagem construída.

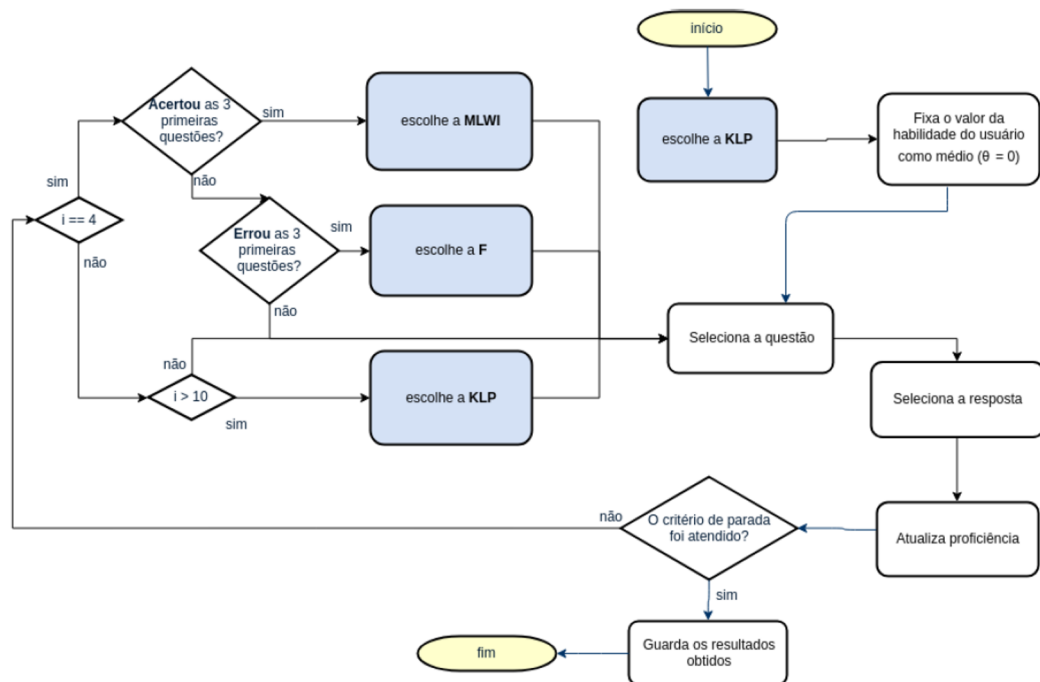


Figura 3 – Fluxograma da abordagem ALICAT.

Fonte: Adaptado de Jatobá *et al.* (2020).

No início do teste utiliza-se o método KLP, que mostrou ter o melhor desempenho na estimativa das habilidades. Após as 3 primeiras perguntas, utiliza-se o critério MLWI se o indivíduo acerta todas, já que o critério apresentou melhor desempenho para os extremos positivos. Caso o indivíduo erre as 3 primeiras perguntas, o critério utilizado será F. Quando não acontece os casos anteriores ou o número de itens administrados já é maior do que 10, volta-se a utilizar o método de melhor desempenho, o KLP.

O artigo citado contribui para o trabalho aqui proposto ao utilizar uma metodologia que a solução pode ser vista como uma política restrita, em que para cada estado decide-se qual critério de seleção usar e a ação é a regra escolhida.

Outro artigo que envolve ideias que contribuem para a proposta é o trabalho em El-Alfy (2011), que explora uma aplicação de técnicas de RL para testes de domínio sequencial (Sequential Mastery Testing - SMT), como é o caso dos CATs. O objetivo é classificar cada pessoa examinada, usando o número mínimo de itens de teste, como “mestre” ou “não mestre”. Usando RL, um agente inteligente aprende autonomamente a partir das interações para administrar testes de tamanho variável mais informativos e eficazes.

O trabalho compara a aplicação de RL com dois métodos que foram usados e estudados por pesquisadores para encontrar regras de decisão sequencial ótimas: o teste de razão de probabilidade sequencial (SPRT) e teoria da decisão Bayesiana sequencial. O SPRT é relevante em situações onde os itens são disponibilizados sequencialmente. Os itens de teste são selecionados aleatoriamente de um banco de itens calibrados, as respostas do examinando aos itens de teste são tratadas como uma sequência de ensaios de Bernoulli independentes e identicamente distribuídos e o SPRT é usado como um critério de encerramento do teste. A teoria de decisão Bayesiana com TRI tem sido usada para classificar examinandos em duas categorias. Essa teoria quando aplicada para testes de tamanho variável permite que os custos associados à administração de itens adicionais sejam levados em consideração.

No sistema de teste de domínio sequencial, sob a metodologia de RL, um agente de tomada de decisão interage com os examinandos. O tomador de decisão consiste em dois módulos. O primeiro módulo (SE) estima um estado de crença com base nas respostas do examinando a itens previamente administrados e conhecimento prévio. O estado para cada examinando é estimado sequencialmente a partir da estimativa anterior e da resposta para o item administrado recentemente. Portanto, a cada passo de tempo n , o agente observa a

resposta atual do examinando z_n para o item administrado q_n e então estimará um novo estado θ_n como função de z_n , q_n e θ_{n-1} . O segundo módulo (π) decide se deve parar o teste e declarar a categoria de examinando ou continuar o teste. Sempre que a decisão for continuar o teste, este módulo selecionará um item apropriado, $q_n + 1$, do conjunto de itens no banco de itens (B) que não foram apresentados anteriormente. O conjunto de todas as regras de decisão é a política π que pode ser predeterminada, aproximada e atualizada online, ou completamente aprendida online. O sistema descrito é mostrado na Figura 4.

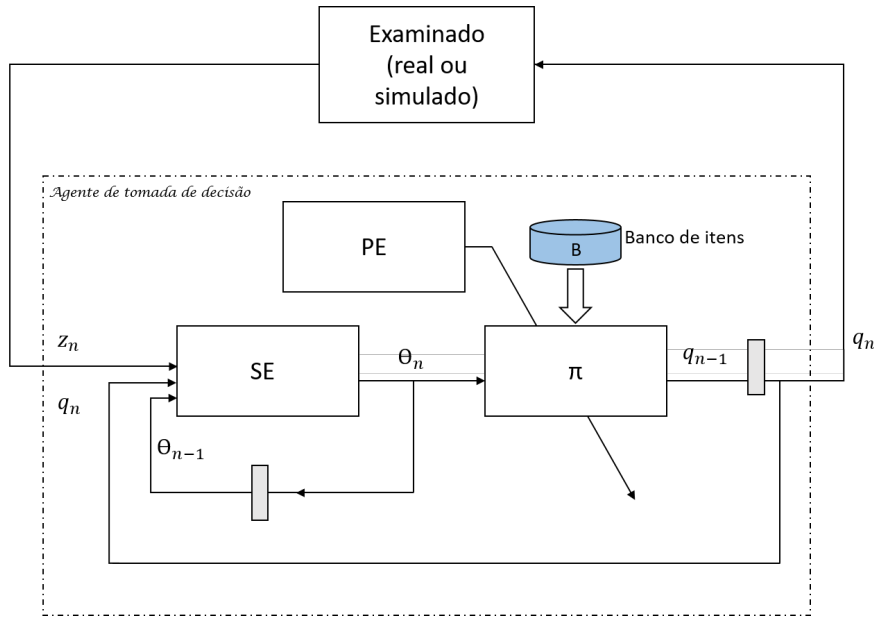


Figura 4 – Sistema de teste de domínio sequencial com RL.

Fonte: Adaptado de El-Alfy (2011).

A sequência de interação com os examinandos se divide em subsequências. Cada subsequência representa a interação com um examinando, onde o sistema começa em algum estado inicial e termina em um estado final. O agente começa a interagir com o examinando j no tempo t onde o estado é s_j^t . Uma ação a_j^t faz com que o estado mude para s_j^{t+1} e receba uma recompensa r_j^t . Então, o agente realiza uma ação a_j^{t+1} e a interação continua até o estado final s_j^{t+T} . O estado do sistema é definido pelo número de respostas corretas após k itens e é denotado por s_k . Definiu-se $d = 2k - s_k$ como a diferença entre o número de respostas corretas e o número de respostas erradas. O estado do sistema é completamente definido por d se estiver em um estado não terminal d e mais um item é apresentado ao examinando, então, com base em sua resposta, o próximo estado será $d + 1$ ou $d - 1$ dependendo se a resposta está correta ou incorreta, respectivamente. Em cada estado, foram definidas três ações: “continuar”, declarar “mestre”, declarar “não

mestre”. Se o tomador de decisão decidir continuar, haverá um pagamento igual ao custo de administração de um item. Se o tomador de decisão decidir parar e declarar o nível de domínio, um custo de classificação é dado com base no fato da classificação estar correta ou incorreta. Este sistema satisfaz a propriedade de Markov, pois o próximo estado é determinado pelo estado atual e pela ação executada.

Dado o estado do sistema, a distribuição preditiva da resposta do próximo item é definida pelas probabilidades de transição. A fim de determinar a estratégia de teste ideal usando RL, o tomador de decisão define uma função de utilidade que determina o valor esperado de ação i no estado s sob a política π . A melhor ação a ser realizada no estado s é aquela que tem o valor mínimo de ação. O tomador de decisão aprende sequencialmente a política durante a interação com os examinandos.

Usando dados simulados e diversas medidas de desempenho, o estudo avaliou o comportamento da abordagem com RL em comparação com SPRT e teoria da decisão Bayesiana sequencial. Verificou-se que os testes baseados em RL tem melhores falsos positivos e falsos negativos do que a abordagem Bayesiana. Embora RL tenha um comprimento médio de teste um pouco maior do que as abordagens SPRT e Bayesiana, ele apresenta ser melhor considerando a precisão e a correlação entre as categorias verdadeiras e preditas.

Os resultados obtidos com a abordagem de RL são promissores quando comparados a outras técnicas. Isso motiva uma investigação mais aprofundada do problema sob várias condições, usando MDP e RL.

O estudo em Nurakhmetov (2019) também utilizada uma abordagem de RL para CAT. Neste caso, o algoritmo de seleção sequencial de itens utilizado incorpora o RL em tarefas de classificação. Além disso, é apresentada uma investigação de como CATs podem ser considerados como um processo de decisão sequencial. O estudo explora a possibilidade de resolver problemas, como controle de exposição e balanceamento de conteúdo, que podem ocorrer em CATs, e simultaneamente tenta otimizar a seleção de itens.

O método proposto realiza uma seleção sequencial de itens que aprende quais itens são mais informativos, escolhendo o próximo item em função dos itens já administrados e da crença interna do classificador. Um estudo de simulação mostrou a eficiência do método para testes que requerem uma decisão de classificação confiável com o mínimo de itens possível. No teste de classificação, a seleção de itens é necessária como opção chave para bons resultados de classificação: filtrar itens menos informativos, ao mesmo tempo em que se presta atenção à exposição e ao controle de conteúdo. Reformulando para o aprendizado

de máquina clássica: a seleção de itens é um problema de otimização combinatória que tenta identificar esses itens, o que minimizará o erro de generalização, com diferentes restrições. Em particular, a seleção de itens pode ser vista como um processo que tenta reduzir a quantidade de dados redundantes.

Transformar a classificação em um processo de decisão sequencial resulta na seleção e classificação de itens um processo adaptativo: decidir qual item selecionar em seguida depende dos itens previamente selecionados e do comportamento do classificador sobre eles. Isso será alcançado usando um classificador treinado como um ambiente para um agente de RL, que aprende qual item selecionar a seguir, recebendo a recompensa na classificação correta do padrão de resposta de entrada parcialmente descoberto.

No primeiro momento, formula-se o CAT como um POMDP, tornando o problema sequencial e, portanto, acessível aos algoritmos de Aprendizado por Reforço. Um requisito do CAT como POMDP foi processar a sequência, uma entrada por vez no modo online, em vez de classificar a sequência inteira de uma vez. Além disso, como saída, um rótulo de classe foi adicionado a cada entrada. Portanto, o classificador sequencial requer algum tipo de memória. Redes Neurais Recorrentes (RNN) são conhecidas por terem memória implícita que pode armazenar informações sobre entradas vistas no passado. Lidar com um POMDP implica precisar extrair uma observação do classificador que resume o passado em uma crença estacionária.

O Estado s no passo de tempo t compreende o padrão de resposta de entrada atual x , o classificador sequencial e o histórico K^* do item anterior h_{t-1} , de modo que $s_t = (x, K^*, h_{t-1})$. A função de recompensa retorna a recompensa r_t no passo de tempo t para a transição do estado s_t para s_{t+1} com ação a_t .

A classificação CAT é formulada como um POMDP e, portanto, está acessível aos métodos RL. Para minimizar o número de itens (duração do teste) necessários para tomar uma decisão de classificação confiável, é feito treinando o agente RL para selecionar itens que levam a uma classificação rápida e precisa. O controle de exposição é mantido usando diferentes recompensas para a seleção de ações, que leva em consideração o histórico de seleção de itens para cada participante do teste. A distribuição de recompensa proposta penaliza o agente durante o processo de treinamento se ele escolher um item que já tenha sido administrado para determinado participante do teste. Além disso, penaliza o agente se o agente selecionar um item cujo limite da taxa de exposição seja excedido.

Por meio de estudos de simulação, o método para seleção adaptativa de itens em CATs mostrou ser eficiente para testes que requerem uma decisão de classificação confiável com o mínimo de itens possível. Além disso, ele pode resolver problemas como superexposição e subexposição com restrições de distribuição de conteúdo, por meio da alocação de recompensa.

Os artigos em El-Alfy (2011) e Nurakhmetov (2019) apresentam abordagens não míopes, porém diferem do trabalho aqui proposto ao tratarem de problemas de classificação. Além disso, as abordagens não tem como objetivo otimizar critério de seleção de itens e critério de parada juntos.

4 Avaliação de compromissos

Um estudo de simulação para um CAT com abordagem Bayesiana foi realizado com o objetivo de avaliar compromissos entre critérios de seleção de itens e critérios de parada.

Utilizando o *software* R Core Team (2019), foram simulados 500 bancos de itens de tamanho 100, considerando o modelo ML3. Para cada banco de itens foram simulados os parâmetros do modelo seguindo as distribuições:

- Parâmetro de discriminação do item i - $a_i \sim \log\text{-normal}(0,0.35)$;
- Parâmetro de dificuldade do item i - $b_i \sim \text{uniforme}(-3,3)$ ou $b_i \sim \text{normal}(0,1)$;
- A probabilidade de acerto aleatório do item i - $c_i \sim \text{beta}(2,5)$; e
- d do ML3 foi definido como 1.7 (modelo normal).

Para o parâmetro b_i foram testadas as duas distribuições: *uniforme* $(-3,3)$ e *normal* $(0,1)$ com o objetivo de verificar como cada distribuição se comporta, principalmente para aluno medianos, junto aos compromissos analisados.

Uma vez simulados os dados, foi construído um CAT utilizando a abordagem Bayesiana para a estimação da habilidade θ , como apresentado na seção 2.1. Nesta avaliação, a distribuição *a priori* considerada para $\Pr(\theta)$ foi *normal* $(0,1)$ e a distribuição *a posteriori* foi discretizada em 10.000 valores espaçados igualmente entre -5 e $+5$. Para cada valor associou-se uma densidade de probabilidade normal padrão ($N(0,1)$) e normalizou-se o resultado para que a soma fosse igual a 1. Dado um item q e uma resposta R ($R = 1$ para resposta correta, $R = 0$ para resposta incorreta), a *a posteriori* para θ discretizado é:

$$\Pr(\theta|q,R) = \frac{\Pr(\theta) \Pr(R|\theta)}{\Pr(R|q)} \propto \Pr(\theta) \Pr(R|\theta,q). \quad (3)$$

A cada iteração calcula-se:

- A média de θ discretizado;
- O desvio padrão e a variância de θ discretizado; e
- A entropia de $\Pr(\theta)$.

Para a habilidade verdadeira θ foi considerada uma sequência de número entre $-2,5$ e $2,5$ variando-se em $0,25$, obtendo 21 habilidades. Esse critério foi utilizado para garantir que o estudo contempla-se níveis de habilidade baixa, média e alta.

Para inicializar o teste, foi selecionado aleatoriamente os itens com parâmetro b entre $-0,5$ e $0,5$, ou seja, itens de dificuldade mediana. Após a administração do item, calcula-se a primeira estimativa para a habilidade θ .

4.1 Critérios de seleção de itens

A princípio, dois critérios de seleção de itens foram implementados. O primeiro foi a informação esperada ou informação de Fisher, que pode ser calculada como:

$$I_i(\theta) = \frac{[\frac{d}{d\theta} P_i(\theta)]^2}{P_i(\theta)Q_i(\theta)} \quad (4)$$

onde, $I_i(\theta)$ é a informação fornecida pelo item i no nível de habilidade θ , $\Pr_i(\theta) = \Pr(X_{ij}|\theta)$ e $Q_i(\theta) = 1 - \Pr_i(\theta)$. No caso do modelo logístico de 3 parâmetros, a equação pode ser escrita como:

$$I_i(\theta) = D^2 a_i^2 \frac{Q_i(\theta)}{P_i(\theta)} \left[\frac{P_i(\theta) - c_i}{1 - c_i} \right]^2. \quad (5)$$

A informação é maior:

1. quando b_i se aproxima de θ ;
2. quanto maior for o a_i ; e
3. quanto mais c_i se aproximar de 0.

Usando a informação de Fisher, o item a ser selecionado é o que fornece maior informação para um dado nível de habilidade.

O segundo critério de seleção de itens utilizado foi a medida de entropia. A medida de um conjunto de probabilidade p_1, \dots, p_n , também chamada por Shannon (2001) de medida de informação, é:

$$H = - \sum_{i=1}^n p_i \log p_i. \quad (6)$$

Quando há duas possibilidades de ocorrência de um evento: p e q , sendo $q = 1 - p$, a entropia do sistema é dada por:

$$H = -(p \log p + q \log q). \quad (7)$$

Usando a entropia, o item a ser selecionado é o que fornece maior entropia para um dado nível de habilidade.

4.2 Critérios de parada

Como critério de parada foram implementadas, a princípio, três configurações: número de questões (itens); variância *a posteriori* de θ ; e o chamado Δ variância. Para cada critério foi calculado o erro quadrático médio (MSE - mean squared error) e o número de questões (itens) necessários para atingir um conjunto de 20 precisões predefinidas para cada método.

No critério número de questões todos os itens foram administrados para todos os pontos de precisão e calculou-se o MSE.

Para o critério de variância, o teste é encerrado quando a variância da distribuição *a posteriori* de θ é menor do que o ponto de precisão considerado e então calcula-se o MSE e o número de questões que foram administrados até a parada.

No critério chamado de Δ variância o teste é encerrado se a variância *a posteriori* dos últimos 5 itens administrados for menor do que a precisão predeterminada. Em seguida, calcula-se o MSE e o número de questões que foram administrados até a parada.

4.3 Resultados

Os resultados dos compromissos entre critérios de seleção de itens e critérios de parada analisados, alterando-se também a distribuição do parâmetro b são apresentadas a seguir. Para todos os gráficos apresentados, as linhas na horizontal e vertical em cada ponto representam o desvio padrão do número de questões (eixo x) e o desvio padrão do MSE (eixo y), respectivamente.

Os gráficos das Figuras 5, 6 e 7 apresentam os resultados obtidos variando-se o critério de seleção de itens e a distribuição do parâmetro b e fixando o número de questões como o critério de parada para alunos com habilidades baixas, medianas e altas, respectivamente. Observa-se que o critério de seleção de itens informação de Fisher apresenta menor MSE quando comparado com o método de entropia para as duas distribuições de b . É possível notar que b com distribuição normal padrão é melhor para alunos com habilidades medianas, como mostra a Figura 6. Para alunos com habilidades baixas e altas o b com distribuição uniforme(-3,3) é melhor em quase todos os casos, como mostram as Figuras 5 e 7. Isso ocorre porque b com distribuição normal padrão gera mais itens com

difficuldade mediana, favorecendo alunos com esse nível de habilidade, o que não ocorre com b tendo distribuição uniforme.

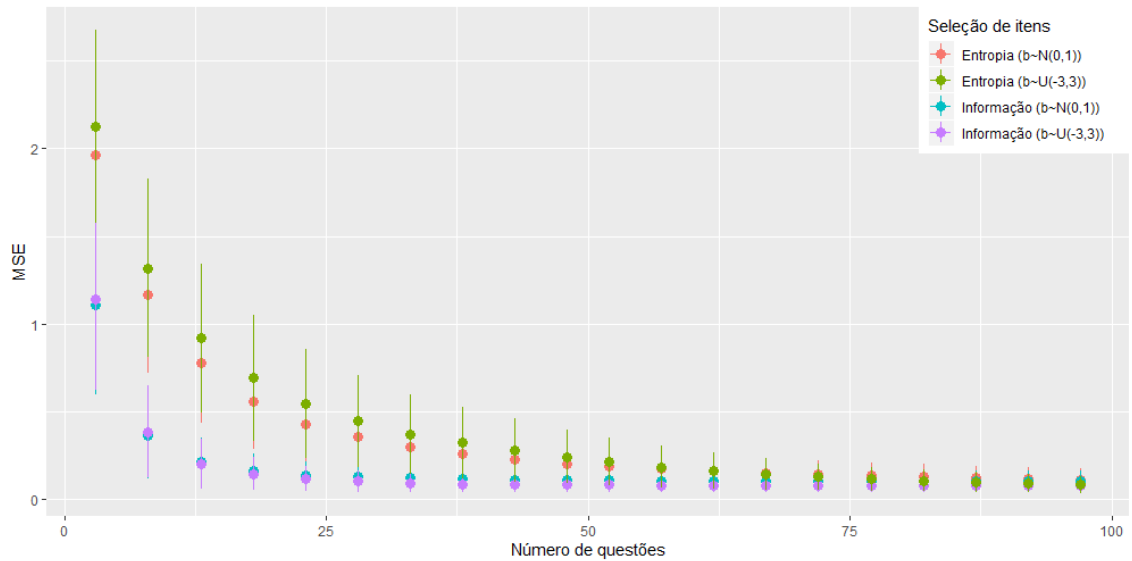


Figura 5 – Critérios de seleção de item, com parâmetro b normal padrão e uniforme(-3,3), para habilidades baixas e critério de parada número de questões.

Fonte: Elaborado pela autora (2020).

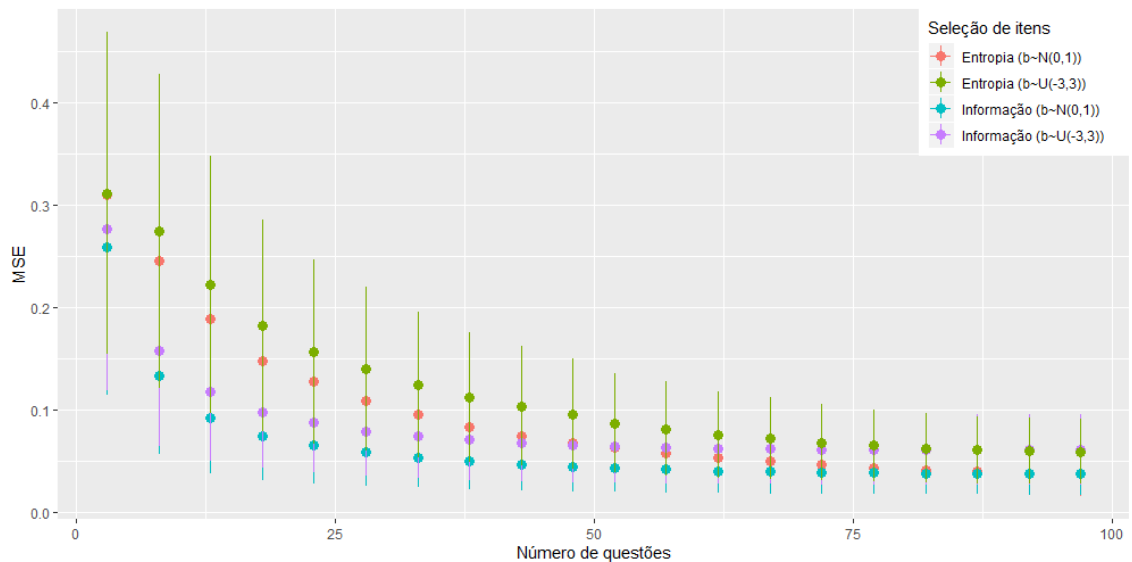


Figura 6 – Critérios de seleção de item, com parâmetro b normal padrão e uniforme(-3,3), para habilidades medianas e critério de parada número de questões.

Fonte: Elaborado pela autora (2020).

As Figuras 8, 9 e 10 apresentam os gráficos com os resultados obtidos variando-se o critério de seleção de itens e a distribuição do parâmetro b e fixando, neste caso, a variância *a posteriori* como o critério de parada para alunos com habilidades baixas, medianas e altas,

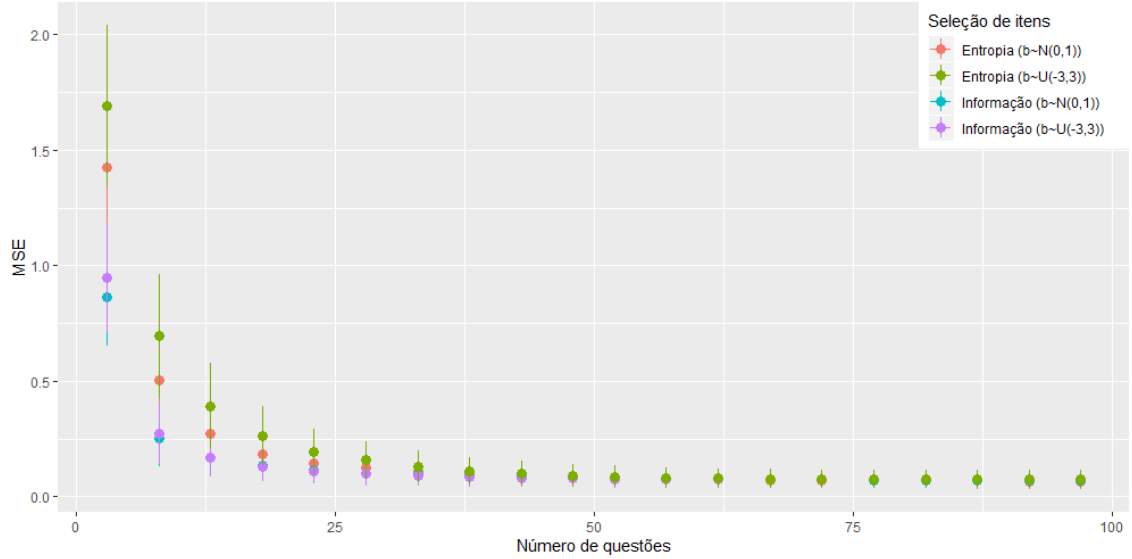


Figura 7 – Critérios de seleção de item, com parâmetro b normal padrão e uniforme(-3,3), para habilidades altas e critério de parada número de questões.

Fonte: Elaborado pela autora (2020).

respectivamente. Assim como observado anteriormente, aqui também observa-se que o critério de seleção de itens informação de Fisher apresenta menor MSE quando comparado com o método de entropia para as duas distribuições de b . Neste caso, também note-se que b com distribuição normal padrão é melhor para alunos com habilidades medianas, como mostra a Figura 9. Para alunos com habilidades baixas e altas o b com distribuição uniforme(-3,3) é melhor em quase todos os casos, como mostram as Figuras 8 e 10.

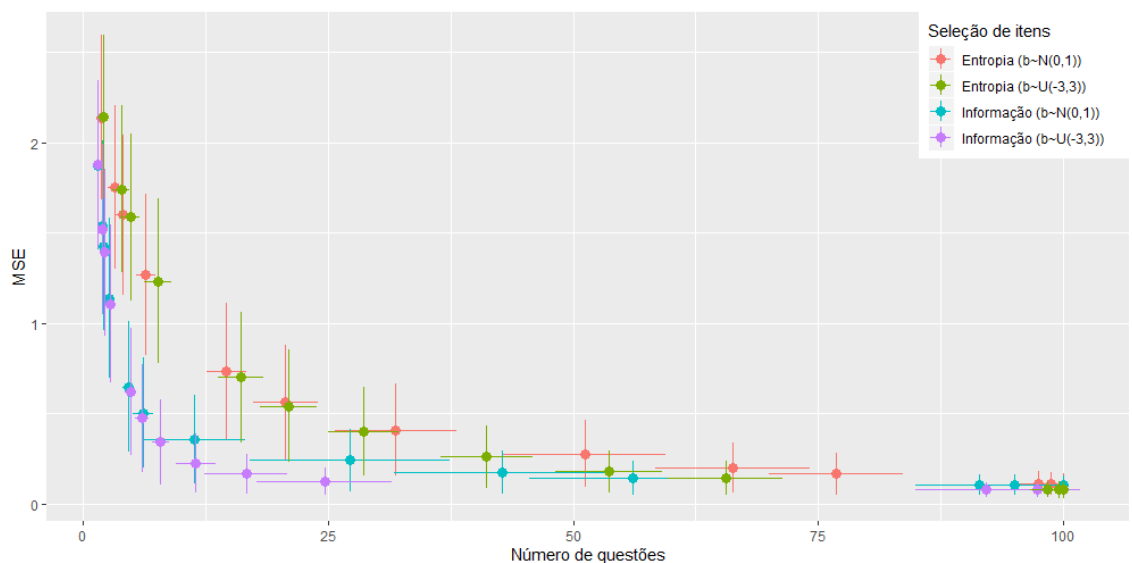


Figura 8 – Critérios de seleção de item, com parâmetro b normal padrão e uniforme(-3,3), para habilidades baixas e critério de parada variância *a posteriori*.

Fonte: Elaborado pela autora (2020).

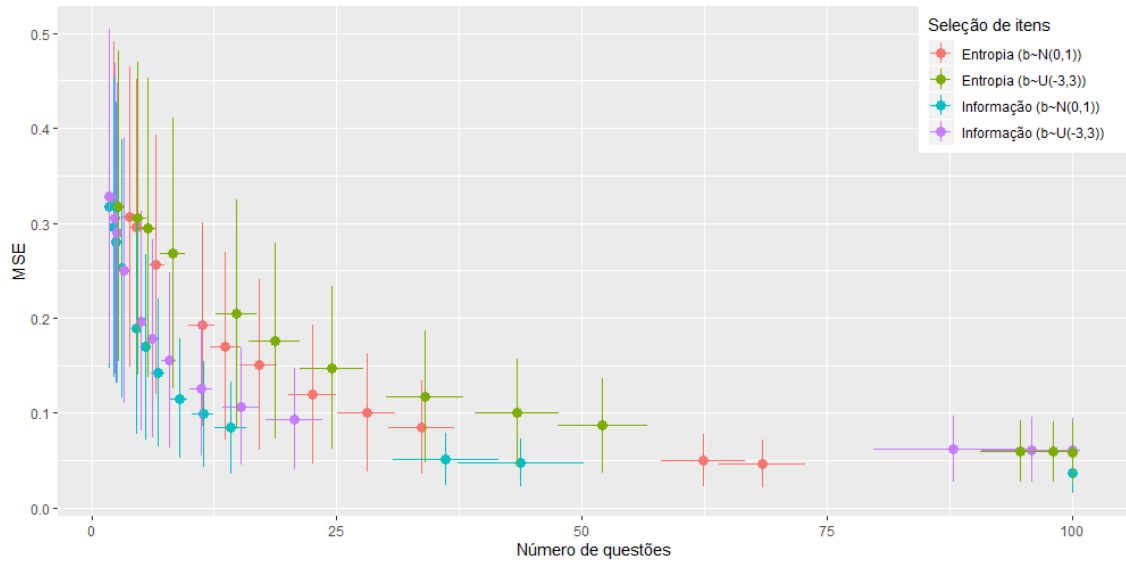


Figura 9 – Critérios de seleção de item, com parâmetro b normal padrão e uniforme(-3,3), para habilidades medianas e critério de parada variância *a posteriori*.

Fonte: Elaborado pela autora (2020).

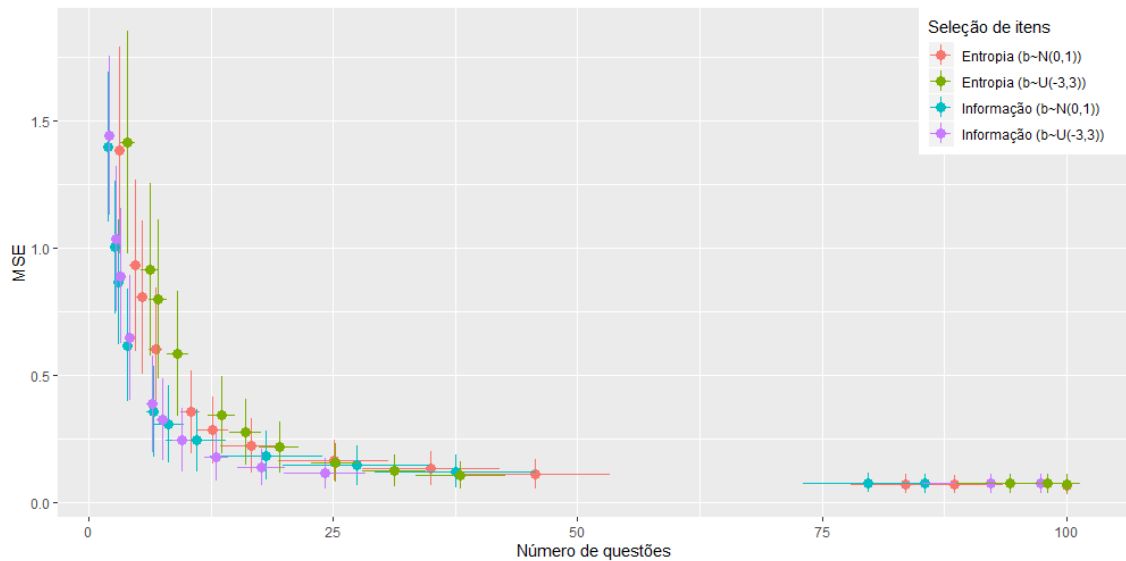


Figura 10 – Critérios de seleção de item, com parâmetro b normal padrão e uniforme(-3,3), para habilidades altas e critério de parada variância *a posteriori*.

Fonte: Elaborado pela autora (2020).

Para os gráficos das Figuras 11, 12 e 13 o critério de seleção de itens e a distribuição do parâmetro b também são variados e fixa-se o método Δ variância como o critério de parada para alunos com habilidades baixas, medianas e altas, respectivamente. As mesmas afirmações anteriores valem para essa configuração do CAT.

Os gráficos das Figuras 14, 15 e 16 apresentam os resultados obtidos trocando o critério de parada no CAT para alunos com habilidades baixas, medianas e altas,

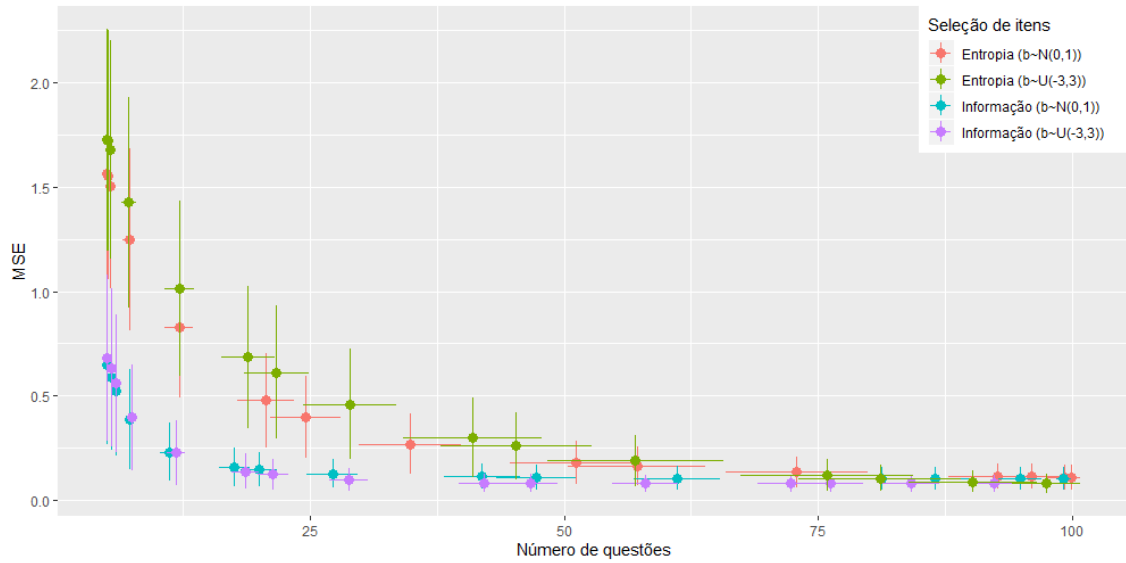


Figura 11 – Critérios de seleção de item, com parâmetro b normal padrão e uniforme(-3,3), para habilidades baixas e critério de parada Δ variância.

Fonte: Elaborado pela autora (2020).

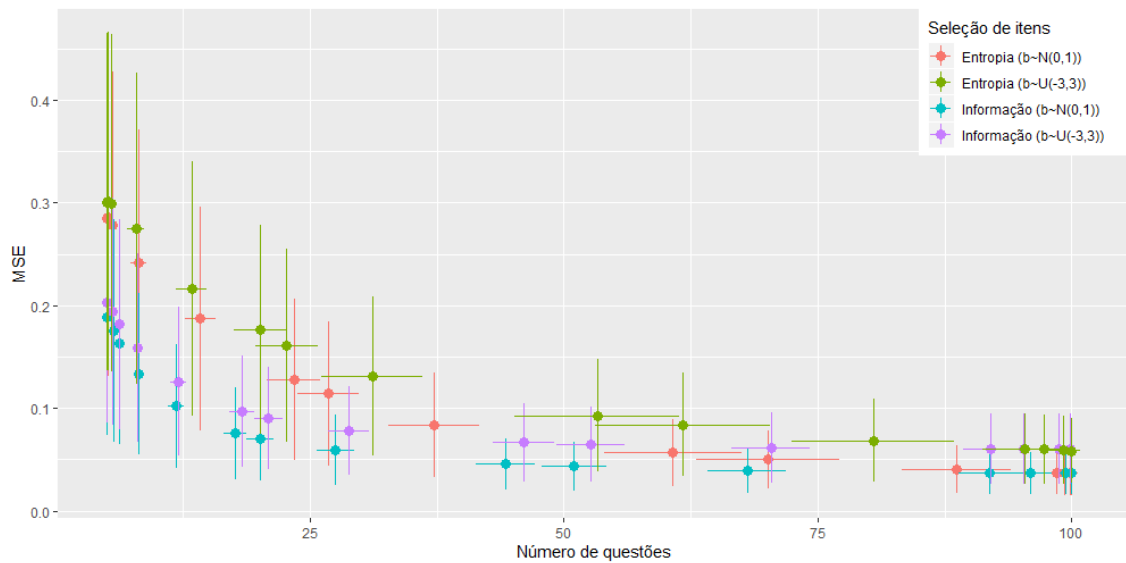


Figura 12 – Critérios de seleção de item, com parâmetro b normal padrão e uniforme(-3,3), para habilidades medianas e critério de parada Δ variância.

Fonte: Elaborado pela autora (2020).

respectivamente. Para analisar o comportamento do critério de parada, a informação de Fisher foi fixada como critério de seleção de itens e o parâmetro b foi fixado com distribuição normal padrão.

No gráfico da Figura 14, para um aluno com habilidade baixa, observa-se que para menos de 10 questões todos os critérios de parada apresentam alto desvio padrão para MSE e a medida que a precisão diminui e o número de questões aumenta, o MSE diminui e

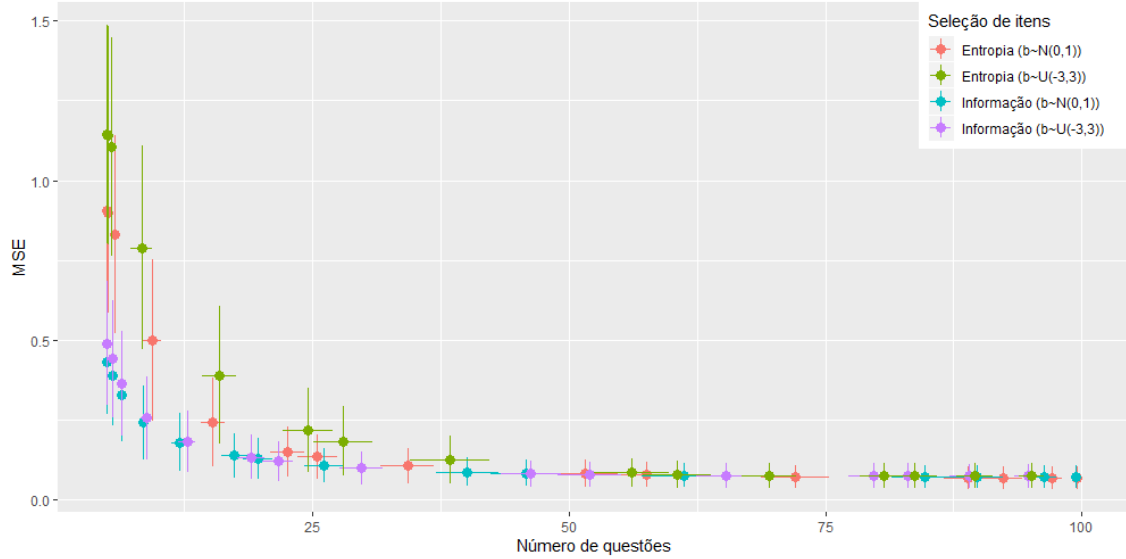


Figura 13 – Critérios de seleção de item, com parâmetro b normal padrão e uniforme(-3,3), para habilidades altas e critério de parada Δ variância.

Fonte: Elaborado pela autora (2020).

o desvio padrão dele também. Neste caso, o método de Δ variância e o método de número de itens parecem ser equivalentes quanto ao MSE, porém o método de número de questões apresenta menor desvio padrão. Observa-se que o método de variância *a posteriori* é o que apresenta maior desvio padrão no número de questões e também apresenta o maior MSE, na maioria das vezes.

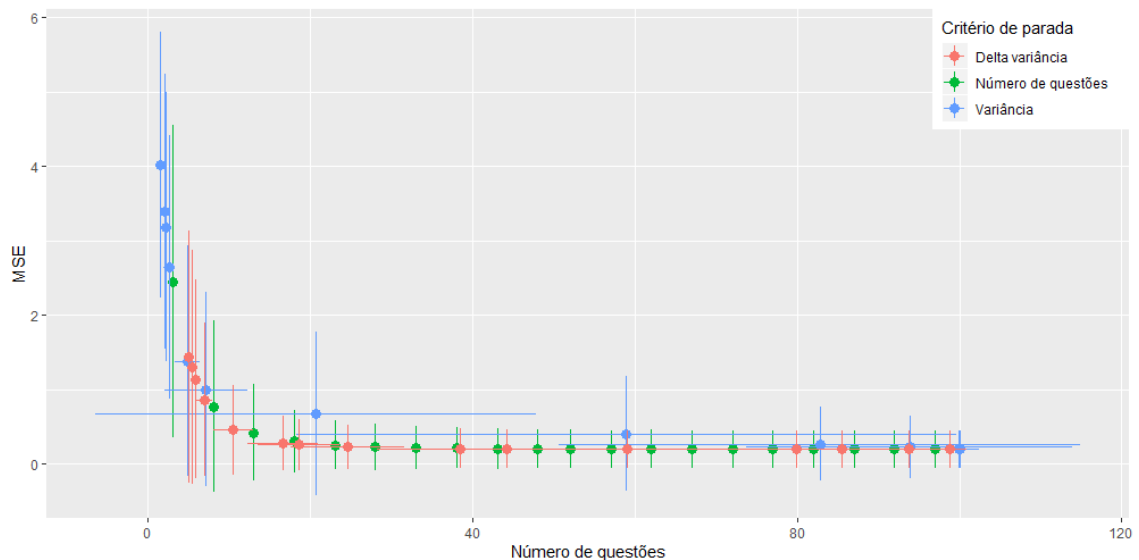


Figura 14 – Critérios de parada para habilidade baixa, informação de Fisher como critério de seleção de itens e parâmetro b com distribuição normal padrão.

Fonte: Elaborado pela autora (2020).

Na Figura 15, para um aluno com habilidade mediana, observa-se que para menos de 5 questões todos os critérios de parada apresentam alto desvio padrão para MSE e a medida que a precisão diminui e o número de questões aumenta, o MSE diminui e o desvio padrão dele também, mas não tanto quanto no caso da Figura 14. A partir de 10 itens, o método de Δ variância e o método de número de itens parecem ser equivalentes quanto ao MSE e seu desvio padrão. Observa-se, o método de variância *a posteriori* apresenta menor desvio padrão no número de itens quando comparado ao aluno de habilidade baixa.

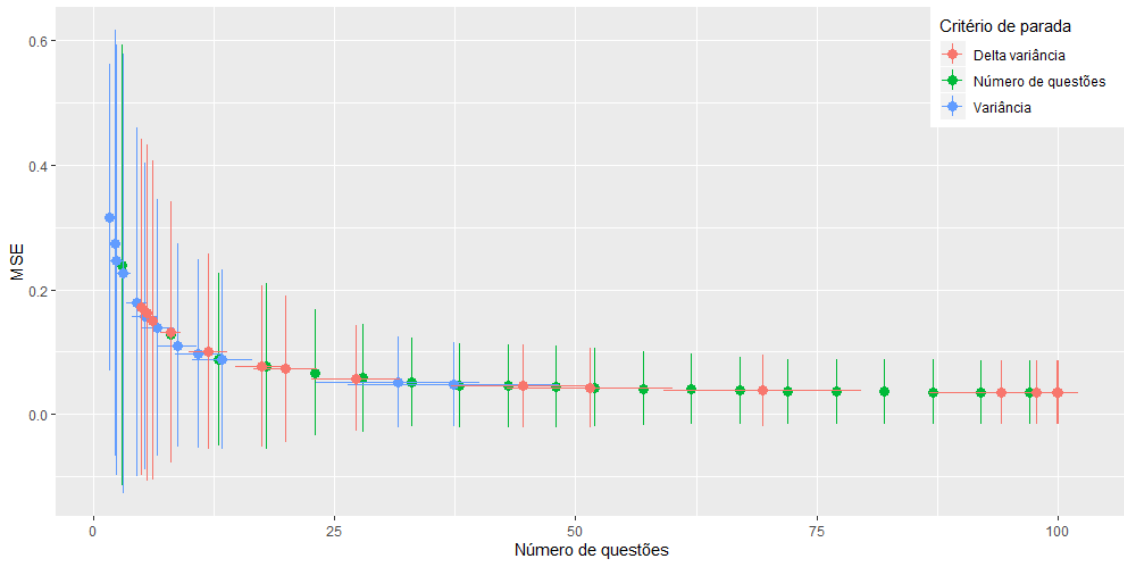


Figura 15 – Critérios de parada para habilidade mediana, informação de Fisher como critério de seleção de itens e parâmetro b com distribuição normal padrão.

Fonte: Elaborado pela autora (2020).

Para o gráfico da Figura 16, para um aluno com habilidade alta, observa-se que para menos de 10 questões o critério de parada da variância *a posteriori* apresentam alto desvio padrão para MSE e a medida que a precisão diminui e o número de questões aumenta, o MSE diminui e o desvio padrão dele também. Neste caso, assim como no caso do aluno com habilidade baixa, o método de Δ variância e o método de número de itens parecem ser equivalentes quanto ao MSE e o método de variância *a posteriori* é o que apresenta maior desvio padrão no número de questões.

4.4 Conclusões

A partir dos gráficos das Figuras de 5 até 13, observa-se que independente do critério de parada (considerando os três implementados) a informação de Fisher apresenta

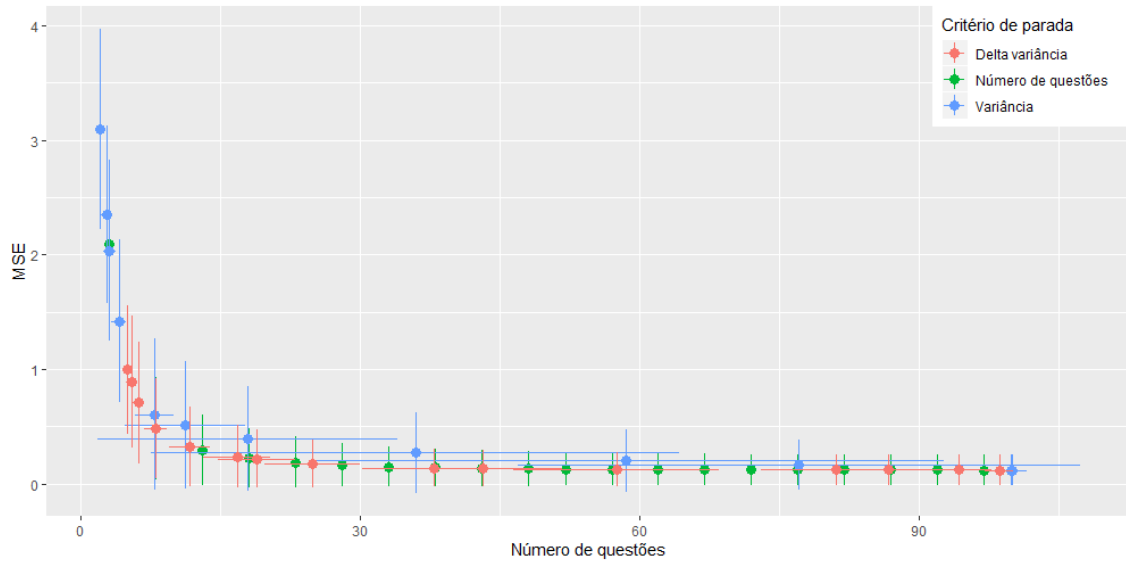


Figura 16 – Critérios de parada para habilidade alta, informação de Fisher como critério de seleção de itens e parâmetro b com distribuição normal padrão.

Fonte: Elaborado pela autora (2020).

menor MSE quando comparado com o método de entropia para as duas distribuições de b consideradas. Nos casos analisados o parâmetro b com distribuição normal padrão mostra ser melhor para alunos com habilidades medianas e com distribuição uniforme(-3,3) mostra-se melhor, em quase todos os casos, para alunos com habilidade baixas ou altas.

Para os gráficos das Figuras 14 e 16 o método da variância *a posteriori* apresentou altos desvios padrão para o número de questões, o que não foi observado na Figura 15 para o aluno com habilidade mediana. Para alunos com habilidade baixa e alta o critério de número de questões parece ser o melhor nos exemplos analisados. Para alunos com habilidade mediana o método de Δ variância e o método de número de itens parecem ser equivalentes quanto ao MSE e seu desvio padrão. No geral, o método de número de questões não apresentou tanta variância no processo.

Em todas as configurações de CAT apresentadas, fixa-se o critério de seleção de itens e varia-se o critério de parada ou vice-versa. O ideal, para encontrar o melhor compromisso entre critério de seleção de itens e critério de parada, é trabalhar com ambos juntos. Um dos objetivos proposta do CAT com MDP é tentar otimizar essa relação, encontrando o melhor critério de seleção de itens e melhor critério de parada juntos.

5 Proposta de Pesquisa

Neste trabalho pretende-se construir um CAT utilizando MDP. De acordo com o sistema da Figura 2, definiu-se que:

- Ambiente: aluno;
- a : item escolhido;
- π : para cada estado, qual o próximo item; e
- s : item submetido e o resultado. (*belief state*)

Inicialmente o estado assume uma distribuição *a priori* (Normal com média igual a 0 e variância igual a 1, por exemplo). A cada iteração, a resposta do aluno modifica a distribuição inicial, obtendo uma distribuição *a posteriori*. Caso o sistema decida interromper o teste, então a média da distribuição *a posteriori* pode ser uma estimativa para θ . A Figura 17 exemplifica a mudança de estado, após o primeiro item, a partir de uma resposta do aluno.

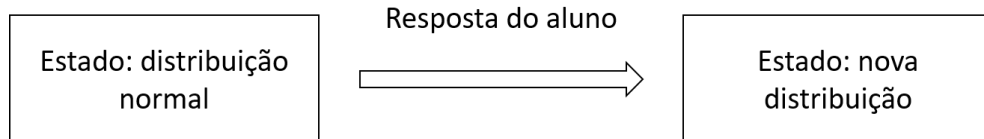


Figura 17 – Ilustração dos estados no CAT com MDP.

Fonte: Elaborado pela autora (2020).

A recompensa será uma função de quantidade de itens e precisão da avaliação do aluno, já que pretende-se encontrar a menor quantidade de itens e melhor estimativa de θ .

5.1 Tema

Os CAT selecionam gradativamente os itens a serem apresentadas ao indivíduo de acordo com o seu nível de conhecimento, também chamado de traço latente. Eles procuram otimizar o teste para cada indivíduo avaliado, estimando o traço latente a cada item respondido e a escolha do próximo item é baseada nesta estimativa.

5.2 Motivação

Diferentemente dos CATs construídos de forma clássica, o CAT como um MDP consegue trabalhar com critério de seleção de itens e critério de parada juntos. Além disso, o MDP permite construir uma abordagem não míope, realizando um compromisso entre ganhos imediatos e ganhos no futuro.

Encontram-se na literatura resultados promissores do uso do MDP nos CATs, por exemplo em Nurakhmetov (2019). Construir CATs como um MDP, aplicando técnicas de RL, pode melhorar as estimativas dos parâmetros do teste e diminuir o número de itens.

Acredita-se que este trabalho pode agregar ainda mais para essa área.

5.3 Lacuna/problema

A construção de CATs como um MDP para encontrar o escore $\hat{\theta}$ do indivíduo ainda não foi explorado.

O uso do MDP para trabalhar com critério de seleção de itens e critério de parada juntos também não foi explorado ainda.

5.4 Justificativa

Caso seja comprovado que a construção de um CAT como um MDP minimiza os erros na estimação da habilidade do aluno e a quantidade de itens de uma prova, então futuramente provas como as do ENEM poderão tirar proveito disso, sendo mais coerente na nota final do aluno de acordo com suas habilidades e sem ter que aplicar um número elevado de itens (180 itens).

5.5 Método

O método de pesquisa consiste em analisar o ganho de se construir CAT como um MDP, utilizando técnicas de RL. O uso do MDP permitirá que a sequência inteira de itens seja avaliada.

A metodologia será aplicada em conjunto de dados reais, bases do ENEM e bases internacionais, e simulados para comparar os resultados em diversos cenários. Todas as implementações computacionais serão realizadas utilizando o *software* R Core Team (2019) e Python.

5.5.1 Atividades

1. Revisão bibliográfica: revisão dos trabalhos relacionados com CATs e construção de CATs como um MDP.
2. Construção de um CAT com abordagem Bayesiana em R: necessário para o entendimento do CAT com estimação Bayesiana e implementação de diversos critérios de seleção de itens e critérios de parada.
3. Avaliação de compromissos: análise das combinações de critérios de seleção de itens e critérios de parada, além da variação da distribuição do parâmetro b do ML3.
4. Construção da estrutura do CAT como um MDP: construção do algoritmo para estabelecer um compromisso entre o número de itens e precisão da avaliação.
5. Experimentos em dados reais e simulados: aplicação do método aos dados do ENEM e outras bases com mais itens para posterior comparação com outros métodos e comparação com dados simulados.
6. Análise dos resultados: utilização de métricas de desempenho para o modelo, bem como comparação com resultados de outros trabalhos. Além de analisar se problemas clássicos do CAT estão sendo levados em conta, como tempo de seleção de itens e quantidade de itens. Após a avaliação, possíveis ajustes podem ser realizados para se alcançar os resultados esperados.
7. Dissertação: redigir a dissertação do trabalho, descrevendo detalhadamente cada etapa realizada.
8. Escrita de artigo: escrita dos resultados obtidos para a submissão de artigos.

5.5.2 Cronograma

Para cumprimento do objetivo principal, serão seguidas as Atividades enumeradas de 1 a 8 de acordo com o cronograma da Tabela 1.

Tabela 1 – Cronograma do projeto.

	2020										2021							
Atividades	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	
1	×	×																
2		×	×	×														
3				×	×	×												
4							×	×	×	×	×							
5											×	×						
6												×	×	×				
7			×	×	×	×	×	×	×	×	×	×	×	×				
8														×	×	×	×	

Fonte: Elaborado pela autora (2020).

5.6 Avaliação

Para avaliação da proposta do trabalho serão utilizadas métricas de desempenho para o modelo desenvolvido e teste de hipótese para verificar a significância, bem como comparação com resultados de outros trabalhos. Também serão analisados se problemas clássicos do CAT estão sendo levados em conta, tempo de seleção de itens e quantidade de itens, por exemplo número de itens resultante para aproximar o escore estimado do escore verdadeiro do aluno comparado com o número de itens de outros trabalhos que utilizam as mesmas bases de dados.

O modelo será testado nas bases de dados do ENEM e em bases internacionais com mais itens, todas de domínio público. Esses resultados também serão comparados com resultados do modelo em dados simulados.

5.7 Contribuição

Esse projeto representa uma contribuição para a área da educação e para a construção de testes adaptativos computadorizados ao propor um método de avaliação de estudantes inédito por meio de CAT como um processo de decisão Markoviano.

Os resultados esperados podem ser levados em conta na futura aplicação de provas com resposta dicotômica para a obtenção de um escore mais justo para o aluno com o uso de poucos itens.

O projeto disponibilizará especificamente:

- Uma construção de testes adaptativos computadorizados como um MDP; e

- Um CAT como um MDP para as bases de dados do ENEM e demais dados públicos abordados.

5.8 *Escopo*

A presente pesquisa propõe a construção de um CAT como um MDP, utilizando o TRI (ML3) e a abordagem Bayesiana, para provas com resposta dicotômica.

Referências¹

- ANDRADE, D. F. de; TAVARES, H. R.; VALLE, R. da C. Teoria da resposta ao item: conceitos e aplicações. *ABE, São Paulo*, 2000. Citado na página 10.
- BAKER, F. B.; KIM, S.-H. *Item response theory: Parameter estimation techniques*. Boca Raton, Florida, EUA: CRC Press, 2004. Citado na página 9.
- BIRNBAUM, A. L. Some latent trait models and their use in inferring an examinee's ability. *Statistical theories of mental test scores*, Addison-Wesley, 1968. Citado na página 9.
- EL-ALFY, E.-S. M. A reinforcement learning approach for sequential mastery testing. p. 295–301, 2011. Citado 3 vezes nas páginas 17, 18 e 21.
- GOSAVI, A. Reinforcement learning: A tutorial survey and recent advances. *INFORMS Journal on Computing*, INFORMS, v. 21, n. 2, p. 178–192, 2009. Citado na página 15.
- JATOBÁ, V. M. G.; FARIAS, J. S.; FREIRE, V.; RUELA, A. S.; DELGADO, K. V. Alicat: a customized approach to item selection process in computerized adaptive testing. *Journal of the Brazilian Computer Society*, Springer, v. 26, p. 1–13, 2020. Citado 3 vezes nas páginas 12, 13 e 16.
- KAELBLING, L. P.; LITTMAN, M. L.; CASSANDRA, A. R. Planning and acting in partially observable stochastic domains. *Artificial intelligence*, Elsevier, v. 101, n. 1-2, p. 99–134, 1998. Citado 2 vezes nas páginas 14 e 15.
- LINDEN, W. J. van der; GLAS, C. A. *Computerized Adaptive Testing: Theory and Practice*. Boston, MA: Springer Science & Business Media, 2000. Citado na página 11.
- NURAKHMETOV, D. Reinforcement learning applied to adaptive classification testing. In: *Theoretical and Practical Advances in Computer-based Educational Measurement*. Cham: Springer, 2019. p. 325–336. Citado 3 vezes nas páginas 19, 21 e 33.
- PELLEGRINI, J.; WAINER, J. Processos de decisão de markov: um tutorial. *Revista de Informática Teórica e Aplicada*, v. 14, n. 2, p. 133–179, 2007. Citado na página 14.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2019. Disponível em: <https://www.R-project.org/>. Citado 2 vezes nas páginas 22 e 34.
- RASCH, G. Probabilistic models for some intelligence and achievement tests. University of Chicago Press. Copenhagen, Denmark: Danish Institute for Educational Research, 1960. Citado na página 9.
- SHANNON, C. E. A mathematical theory of communication. *ACM SIGMOBILE mobile computing and communications review*, ACM New York, NY, USA, v. 5, n. 1, p. 3–55, 2001. Citado na página 23.

¹ De acordo com a Associação Brasileira de Normas Técnicas. NBR 6023.

SILVER, D.; VENESS, J. Monte-carlo planning in large pomdps. In: LAFFERTY, J. D.; WILLIAMS, C. K. I.; SHAWE-TAYLOR, J.; ZEMEL, R. S.; CULOTTA, A. (Ed.). *Advances in Neural Information Processing Systems 23*. Curran Associates, Inc., 2010. p. 2164–2172. Disponível em: <http://papers.nips.cc/paper/4031-monte-carlo-planning-in-large-pomdps.pdf>. Citado na página 15.

SPENASSATO, D.; BORNIA, A.; TEZZA, R. Computerized adaptive testing: A review of research and technical characteristics. *IEEE Latin America Transactions*, v. 13, n. 12, p. 3890–3898, 2015. Citado 3 vezes nas páginas 7, 11 e 13.

STAFFORD, R. E.; RUNYON, C. R.; CASABIANCA, J. M.; DODD, B. G. Comparing computer adaptive testing stopping rules under the generalized partial-credit model. *Behavior research methods*, Springer, v. 51, n. 3, p. 1305–1320, 2019. Citado na página 7.