

HAIL: um modelo de revisão de percepções de agentes inteligentes baseado em alucinação e ilusão

Gustavo E. Kundlatsch¹, Thiago Ângelo Gelaim¹, Elder Santos¹

¹Departamento de Informática e Estatística
Universidade Federal de Santa Catarina (UFSC)
Florianópolis – SC – Brazil

Abstract. *Perceptions are the main way to an entity to receive information from the environment. Each person has a different manner of percept and interpretate the world. However, it is known that in the human perception there are anomalies, incorrect perceptions that decieve the mind. So, how can we know if our perceptions are real or just tricks from our mind? And the follow up question is: what about computers? Intelligent agents are computational entities capable of making decisions based on the environment, and use sensors to perceive the world. But sensors can fail. In the current work, we present a generic model of perception revision, capable of treating anomalies received by the agent and creating new plans from them to adapt to the environment. This model was implemented and submitted to experiments to test it's behavior. The simulations demonstrate that the model is capable of detect and classify anomalies and that in environments with a high amount of anomalies the model can have more impact creating new plans through the automatic planning process and increasing the amount of valid perceptions received by the agent's cognition.*

Resumo. *Percepções são a principal maneira de uma entidade receber informações do ambiente. Cada pessoa possui uma maneira diferente de perceber e interpretar o mundo. Entretanto, sabe-se que na percepção humana existem anomalias, que são percepções incorretas que enganam a mente. Dito isso, como podemos saber se nossas percepções são reais ou se são apenas fruto de nossa imaginação? E a questão derivada disso é: e computadores? Agentes inteligentes são entidades computacionais autônomas capazes de tomarem decisões baseadas no ambiente no qual estão inseridos, e utilizam sensores para reconhecerem o mundo a sua volta. Mas esses sensores podem falhar. Neste trabalho, apresentamos um modelo genérico de revisão de percepções, capaz de tratar anomalias recebidas pelo agente e criar novos planos a partir delas para se adaptar ao ambiente. Esse modelo foi implementado e submetido a experimentos para testar seu funcionamento. As simulações realizadas demonstraram que o modelo é capaz de detectar e classificar as anomalias e que em ambientes com grande quantidade de anomalias o modelo consegue ter maior impacto criando novos planos através do planejamento automatizado e aumentando a quantidade de percepções válidas recebidas pelo raciocínio do agente.*

1. Introdução

Dentro da Inteligência Artificial (IA), agentes inteligentes são entidades capazes de raciocinar a respeito do ambiente em que estão inseridos e tomar decisões baseadas na situação

em que se encontram [Russell and Norvig 2016]. Dessa maneira, podemos descrever um agente pelo seus processos de percepção, raciocínio e atuação. O agente ocupa um ambiente, do qual recebe informações e no qual atua. O ambiente é o mundo em que o agente está inserido, podendo ser uma construção virtual como uma simulação ou uma parte do mundo real, no caso de um agente físico. Existem diversos tipos de ambientes que podem ser classificados de acordo com o seu fechamento (que determina se agentes de fora do ambiente podem afetar o sistema), dinamismo (a maneira como o ambiente evolui), determinismo (a consistência dos efeitos no ambiente) e cardinalidade (o número de objetos a serem afetados e percebidos) [Moya and Tolk 2007].

Uma das maneiras de um agente atualizar seu conhecimento a respeito do ambiente é a percepção, o processo de utilizar sensores para detectar o ambiente e transformar os dados coletados em informações úteis [Weyns et al. 2004]. O raciocínio, por sua vez, é o processamento das percepções baseado nos objetivos do agente, que resulta em um conjunto ações a serem tomadas através dos atuadores. O processo do raciocínio é comandado pela arquitetura cognitiva do agente, um modelo computacional inspirado na estrutura da mente humana [Dyachenko et al. 2018]. As arquiteturas cognitivas podem ser divididas em três categorias: simbólicas, emergentes e híbridas [Ye et al. 2018]. Arquiteturas simbólicas descrevem o ambiente através de símbolos armazenados em memória em uma base de conhecimentos, e utilizam lógica simbólica para realizar o ciclo de percepção, raciocínio e ação. Arquiteturas emergentes se baseiam na estrutura biológica do cérebro e normalmente utilizam redes neurais em uma estrutura hierárquica para lidar com situações de incerteza. Por fim, arquiteturas híbridas combinam o comportamento emergente e o processamento simbólico para resolver problemas de diversos domínios.

Todavia, sensores podem apresentar problemas para o processo de percepção por razões como campo de visão, distância do objeto observado, resolução dos sensores e leituras não confiáveis [Chrisman et al. 1991]. Tratar deste problema normalmente é responsabilidade da arquitetura cognitiva do agente, pois a arquitetura precisa ser capaz de fazer a ponte entre o ambiente e o conhecimento do agente [Langley et al. 2009].

O objetivo deste trabalho é apresentar um modelo genérico (independente da arquitetura do agente) que pode ser acoplado entre o processo de percepção e raciocínio, capaz de detectar e tratar percepções inválidas para transformá-las em informações úteis através de um processo de criação de novos planos. Esse modelo pressupõe um ambiente aberto (onde agentes externos podem influenciar o ambiente), dinâmico (mudanças no ambiente são causadas por eventos aleatórios) e não determinístico (ações do agente causam resultados diferentes no ambiente, mesmo em situações aparentemente idênticas, pois os resultados variam dependendo da percepção do agente daquele evento).

2. Fundamentação teórica

Alguns conceitos precisam ser bem definidos para que sejam utilizados mais tarde na formalização do modelo proposto. Esse capítulo apresenta essas definições iniciais.

2.1. Agente inteligente

Um agente inteligente é uma entidade autônoma, capaz de tomar as próprias decisões para atingir seus objetivos [Wooldridge 1999]. Apesar da definição intuitiva ser simples, assim como no termo inteligência artificial não existe um consenso da comunidade sobre o que é um agente. Para este trabalho, compomos a seguinte formalização de agente inteligente:

Definição 1. Um agente é uma tripla $Ag = \langle K, P, \gamma \rangle$, onde:

- K é uma base de conhecimentos, tal que $K = K_i \cup K_p$, onde K_i é o conjunto de conhecimentos iniciais do agente e K_p os conhecimentos adquiridos através das percepções. K_i é iniciado com valores arbitrários de acordo com a necessidade do agente e K_p é iniciado vazio. Uma base de conhecimentos é uma estrutura que representa fatos a respeito do mundo e apresenta formas de raciocinar a respeito desses fatos para deduzir novos conhecimentos [Hayes-Roth et al. 1983];
- P é o conjunto de planos do agente, sendo um plano definido como $plano = (\Psi, A, \Omega)$, onde Ψ é o conjunto união formado pelas pré-condições das ações que compõem o plano, A o conjunto de ações que compõe o plano e Ω o conjunto união formado pelas pós-condições das ações que compõem o plano. Por sua vez, uma ação é definida como $acao = (\psi, n, \omega)$, sendo ψ um conjunto de pré-condições, n um nome para a ação e ω um conjunto de pós-condições; e
- γ é a função de percepção, definida como $\gamma(p, K) \rightarrow P_i$, onde p é o conjunto de percepções recebidas, K a base de conhecimentos de Ag e P_i o retorno da função, que é um subconjunto próprio do conjunto P de planos do agente.

A partir dessa definição, podemos construir o conceito de contexto, que será importante mais tarde para o modelo de revisão de percepções. O contexto de um agente é o conjunto de todos os símbolos compreendidos pelo agente, e cuja percepção de cada um desses símbolos resulta na execução de um conjunto de ações diretamente mapeadas.

Definição 2. O contexto c de um agente Ag é o domínio de sua função γ .

Para este trabalho, serão utilizados símbolos compostos para representar percepções e ações. Esses símbolos são formados da maneira *predicado(característica)*, onde o predicado é o elemento principal do símbolo e a característica uma propriedade secundária.

2.2. Percepção

Existem diversas definições para o termo “percepção”. Podemos entender percepção como um conjunto de sensações que, através da maneira subjetiva que um dado agente o interpreta, representa determinadas entidades do ambiente [Gibson 1950]. Ou seja, a percepção não é simplesmente a representação direta das entidades reais que existem no mundo, mas um processo complexo que varia para cada indivíduo.

Para o modelo que iremos propor, baseado na Definição 1, o conceito de percepção pode ser simplesmente definido como as entradas da função de percepção γ de um agente. Vale destacar a diferença entre percepção e contexto, pois o contexto é constituído pelas percepções que fazem parte do domínio da função γ , ou seja, uma percepção é toda informação produzida pelo ambiente que o agente recebe, e contexto é o subconjunto das percepções que o agente possui planos para tratar.

2.2.1. Refinamento

Como o volume de percepções de um agente pode ser muito grande, e as percepções tomam determinado tempo para serem processadas, o número de percepções que chegam ao ciclo de raciocínio do agente pode ser reduzido para diminuir seu custo computacional. Neste trabalho, esse processo será chamado de refinamento, definido da seguinte maneira:

Definição 3. Refinamento de percepções é uma função θ tal que, dado o conjunto de entradas de percepções p , reduz tais percepções para um subconjunto próprio ρ .

2.2.2. Anomalias

Anomalias são as percepções consideradas inválidas, geradas por alguma falha no processo de percepção. Toda percepção para a qual o agente não possui uma resposta mapeada em sua função de percepção é considerada uma anomalia. O conjunto de anomalias de um agente é o conjunto de todas as percepções possíveis que não fazem parte de seu contexto. A seguinte definição será utilizada:

Definição 4. Uma percepção p de um agente Ag com contexto c , é uma anomalia caso $p \notin c$ de Ag .

Neste trabalho, as anomalias serão classificadas em dois grupos: alucinações e ilusões. Essa classificação foi inspirada no problema da percepção, um campo de estudo da filosofia [Crane and French 2017].

Uma ilusão é uma percepção da forma *predicado(característica)* onde ou o predicado ou a característica não faz parte do contexto do agente, ou seja, é uma percepção parcialmente correta. Uma alucinação é uma percepção da forma *predicado(característica)* onde nem o predicado nem a característica faz parte do predicado, ou seja, a percepção pode ser semanticamente correta mas ela não deveria ter sido realizada pelo agente.

2.3. Planejamento automatizado

Planejamento automatizado é um dos problemas fundamentais da Inteligência Artificial. As motivações para usar o planejamento automatizado são a capacidade de utilizar recursos de planejamento acessíveis e eficientes e reproduzir uma parte do processo cognitivo humano com um componente totalmente integrado de comportamento deliberativo [Ghallab et al. 2004].

Na Definição 5 é apresentada a noção abstrata de planejamento automático, descrita como um modelo conceitual simples que contém os elementos principais do problema, tendo sido originalmente apresentada por [Ghallab et al. 2004].

Definição 5. Um modelo conceitual de planejamento automatizado é descrito como a interação entre os seguintes três componentes:

- Um sistema de transição de estados Σ , especificado por uma função de transição de estados y , de acordo com os eventos e ações que ele recebe.
- Um *controlador*, que dado uma entrada de estados s do sistema, fornece como saída uma ação de acordo com algum plano.
- Um *planejador*, que dado uma entrada de uma descrição de sistema Z , uma situação inicial e alguns objetivos, sintetiza um plano para o controlador a fim de alcançar o objetivo.

Um sistema de transição de estados Σ é uma quádrupla $\Sigma = \langle S, A, E, \Gamma \rangle$, onde:

- $S = \{s_1, s_2, \dots, s_n\}$ é um conjunto finito ou recursivamente enumerável de estados;

- $A = \{a_1, a_2, \dots, a_n\}$ é um conjunto finito ou recursivamente enumerável de ações;
- $E = \{e_1, e_2, \dots, e_n\}$ é um conjunto finito ou recursivamente enumerável de eventos; e
- $\Gamma : S \times A \times E \rightarrow 2^S$ é uma função de transição de estados.

3. Trabalhos relacionados

Existem diversas abordagens para otimizar as percepções recebidas por um agente, isto é, garantir que todas as informações coletadas pelos sensores sejam utilizadas da melhor maneira possível. Diversos artigos apresentam processos de detectar percepções inválidas e tratá-las, de maneira similar ao modelo que será proposto neste trabalho.

Por exemplo, no artigo *Scalable perception for BDI-agents embodied in virtual environments* [van Oijen and Dignum 2011] é definido um *framework* onde os objetos do ambiente são definidos por classes e características e se organizam de maneira hierárquica, e a partir dessa organização o *framework* é capaz de decidir que tipo de percepções o agente deseja perceber de acordo com seus interesses, alterando dinamicamente ao longo do tempo.

Outro trabalho similar é o *PMK – a knowledge processing framework for autonomous robotics perception and manipulation* [Diab et al. 2019], onde o objetivo dos autores foi criar um mapa de todas as percepções possíveis no ambiente, de maneira que todas as percepções recebidas pelo agente estivessem em seu contexto, evitando anomalias. Todavia, isso não é viável em cenários abertos ou de alta complexidade, pois o mapeamento necessário é extenso demais. O modelo de programação lógica K-CoPMan (*Knowledge enabled Cognitive Perception for Manipulation* ou Percepção Cognitiva Ativada pelo Conhecimento para Manipulação) apresenta uma maneira do agente criar sua própria base de conhecimentos a partir da percepção passiva, eliminando as anomalias conforme elas são adicionadas ao contexto [Pangercic et al. 2010].

Esses exemplos são de artigos que buscam otimizar as percepções recebidas pelo agente de maneira simbólica, mas esse problema pode ser abordado de maneira conexionista também. No artigo *Understanding human intention by connecting perception and action learning in artificial agents* [Kim et al. 2017] os autores propõem um modelo, chamado OA-SMTRNN (*Object Augmented Supervised Multiple Timescale Recurrent Neural Network*), para entender a intenção do usuário e responder ativamente da maneira mais adequada, através do uso de redes neurais. Para implementar o reconhecimento de intenção, são focados dois processos cognitivos, a percepção da disponibilidade de objetos e a previsão da ação humana. Com o uso de redes neurais, a ideia é extrair semântica de percepções que seriam inicialmente inválidas para o agente.

O objetivo do modelo proposto neste trabalho é detectar anomalias de maneira simbólica, classificando as percepções recebidas com o uso do contexto do agente como referencial, para então utilizar um processo de planejamento automatizado para gerar novos planos para o agente a partir das percepções inválidas.

4. Modelo proposto

O modelo proposto foi inspirado pelos conceitos de ilusão e alucinação apresentados na Seção 2.2.2, que o nomeiam – o nome HAIL vem da junção das palavras *hallucination*

e *illusion*, alucinação e ilusão em inglês, respectivamente. Seu objetivo é identificar anomalias nas percepções recebidas por um agente qualquer e torná-las informação úteis na forma de novos planos.

O modelo HAIL foi desenvolvido para que seja possível adicioná-lo a qualquer agente, independente de arquitetura cognitiva, como um componente que conecta as percepções vindas do ambiente ao agente. O HAIL pode ser separado em dois módulos, como mostra a Figura 1. De maneira geral, o funcionamento e a comunicação desses módulos se dá da seguinte maneira:

1. As percepções recebidas pelo modelo são refinadas pelo módulo de refinamento;
2. As percepções refinadas passam pelo módulo de alucinação e ilusão onde são categorizadas entre: percepções válidas, alucinações, ilusões classe 1 e ilusões classe 2.
3. As percepções válidas são encaminhadas para o raciocínio do agente, enquanto as anomalias continuam no módulo de alucinação e ilusão armazenadas em estruturas chamadas de bloco avaliador;
4. Quando os requisitos estabelecidos pelo bloco avaliador são cumpridos, as anomalias são selecionadas para passarem pelo processo de planejamento automatizado, alimentando o agente com novos planos.

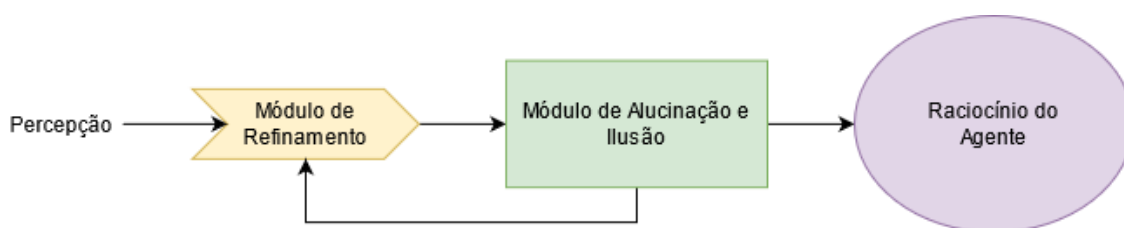


Figura 1. Visão geral do modelo HAIL.

4.1. Módulo de refinamento

O módulo de refinamento funciona como um primeiro filtro para que percepções indesejadas pelo agente não cheguem até seu ciclo de raciocínio. O processo de refinamento é descrito pela Definição 3.

O processo de refinamento não é obrigatório. Caso não seja de interesse de uma determina implementação do HAIL refinar suas percepções, basta que a função do módulo de refinamento seja a função identidade $f(x) = x$, possuindo assim $\rho = p$.

4.2. Módulo de alucinação e ilusão

A Figura 2 apresenta um diagrama do funcionamento do módulo de alucinação e ilusão. Sua função é receber todas as percepções que passaram pelo processo de refinamento, e detectar quais delas são anomalias. Para isso, esse módulo utiliza uma cadeia de decisores descrita pelo Algoritmo 1. Depois de serem classificadas pelos decisores, as anomalias são guardadas em seus respectivos blocos avaliadores.

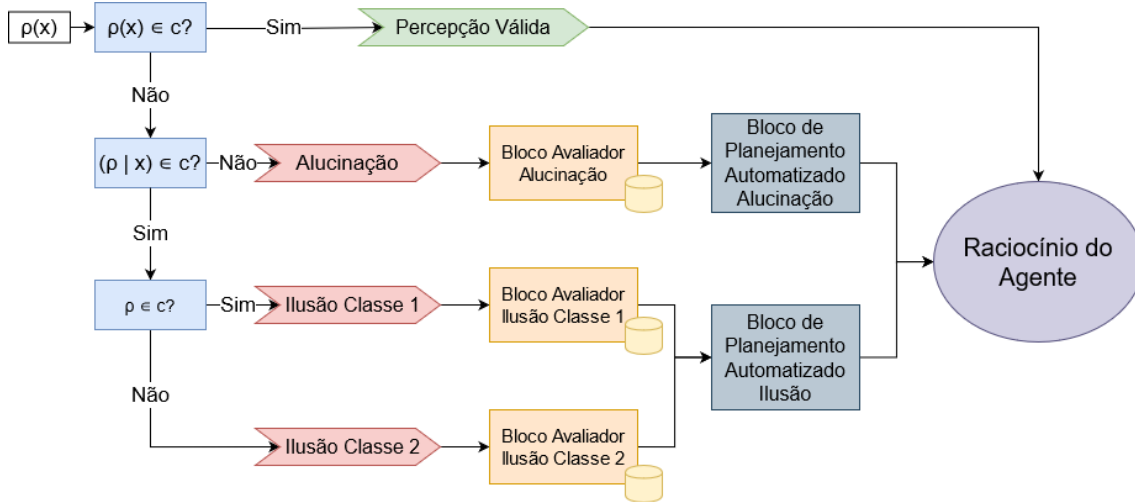


Figura 2. Módulo de alucinação e ilusão.

Entrada: contexto c do agente, percepção $\rho(x)$

```

1 início
2   se  $\rho(x)$  está em  $c$  então
3     |  $\rho(x)$  é uma percepção válida;
4   senão se nem  $\rho$  nem  $x$  estão em  $c$  então
5     |  $\rho(x)$  é uma alucinação;
6   senão se  $\rho$  está em  $c$  então
7     |  $\rho(x)$  é uma ilusão classe 1;
8   senão
9     |  $\rho(x)$  é uma ilusão classe 2
10  fim
11 fim

```

Algoritmo 1: Funcionamento dos decisores do módulo de alucinação e ilusão.

4.2.1. Bloco avaliador

O objetivo do bloco avaliador é decidir quando alucinações e ilusões que foram recebidas devem ser processadas, evitando que o planejamento automatizado que será realizado em seguida tenha impacto no tempo de execução de um ciclo de raciocínio do agente. Para isso, utilizamos uma lista ordenada por peso como escalonador. O princípio do funcionamento da lista ordenada por peso é o mesmo de uma fila, *first in first out* (FIFO), mas atribui um peso a cada entrada que aumenta quando novos elementos iguais são inseridos. Quando um elemento é inserido pela primeira vez na fila, ele recebe o peso 1, e quando uma cópia do mesmo elemento é inserida o elemento tem seu peso aumentado em 1. Quando uma operação de remoção é executada, o elemento de maior peso é removido. Se dois ou mais elementos tiverem o maior peso, aquele que foi inserido primeiro é removido.

O bloco avaliador seleciona quando uma percepção deve ser tratada através de uma função matemática, levando em conta o tempo médio de processamento de uma

percepção válida e de uma anomalia. Além desse funcionamento básico, o bloco avaliador ainda contém um mecanismo para remover anomalias classificadas com irrelevantes para o sistema, através de uma função de limpeza. Caso essa função retorne verdadeiro, todos os elementos de peso 1 da sua respectiva lista são removidos. Essas duas funções são descritas com mais detalhes na Seção 4.3.

4.2.2. Bloco de planejamento automatizado

O bloco de planejamento automatizado é potencialmente a parte mais custosa computacionalmente, o que pode ser um gargalo do sistema, principalmente caso o agente funcione em tempo real e receba um volume muito elevado de percepções por segundo. Um planejamento automatizado implementado de maneira puramente simbólica tende a ser complexo computacionalmente, uma vez que pode considerar milhares de alternativas para o estado de mundo atual, tentando chegar mais perto de seu objetivo. Um processo de planejamento automatizado conexionista é uma alternativa, uma vez que estamos tratando de uma análise incompleta do mundo. Caso seja possível, teorias com maior custo computacional (como criatividade computacional [Colton et al. 2012]) podem ser aplicadas aqui para um resultado ainda mais preciso.

Uma percepção chega ao bloco de planejamento automatizado uma vez que ela seja a primeira na fila ponderada e a função de processamento retorne verdadeiro em sua verificação. De um ciclo para outro, as percepções permanecem na fila, a não ser que sejam descartadas pelo mecanismo de limpeza. Nosso modelo não explicita qual é a ordem que os blocos avaliadores devem processar suas filas para mandar anomalias para o planejamento automatizado (isto é, se deve primeiro ser priorizada as anomalias do bloco avaliador de alucinações, ilusões classe 1 ou ilusões classe 2), ficando a cargo da implementação em questão tomar essa decisão.

4.3. Formalização

O bloco básico do modelo de revisão de percepções proposto, chamado de HAIL, é composto por um módulo para alucinação e ilusão M_{ai} e uma função de refinamento θ , conforme descrito na Definição 6. O módulo de ilusão e alucinação é uma quádrupla, apresentada na definição 7. A função de refinamento é uma função abstrata, cuja entrada é obtida através dos sensores do agentes e a saída é a entrada do módulo de alucinação e ilusão, conforme já foi descrito anteriormente na Seção 4.1.

Definição 6. O modelo de revisão de percepções HAIL é uma dupla $HAIL = \langle M_{ai}, \theta \rangle$, onde:

- M_{ai} é o módulo de ilusão e alucinação; e
- θ é a função de refinamento $\theta(p) = \rho$, onde p é um conjunto de percepções e ρ é um subconjunto próprio de p .

Após ter passado pela função θ , as percepções ρ irão passar pelo Algoritmo 1, e serão encaminhadas de acordo com sua classificação. O bloco de ilusão e alucinação é descrito por uma quádrupla, com conjuntos de decisores, blocos e uma função de transição.

Definição 7. O bloco de ilusão e alucinação é uma quádrupla $M_{ai} = \langle D, Ab, Ap, \Delta \rangle$, onde:

- D é o conjunto de decisores $D = \{d_a, d_h, d_i\}$, onde:
 - d_a é o decisor de anomalias, definido pela função:

$$d_a = \begin{cases} 0 & \text{se } \rho(x) \text{ está em } c^1; \\ 1 & \text{se } \rho(x) \text{ não está em } c. \end{cases}$$

- d_h é o decisor de alucinação, definido pela função:

$$d_h = \begin{cases} 0 & \text{se nem } \rho \text{ nem } (x) \text{ está em } c; \\ 1 & \text{se } \rho \text{ ou } (x) \text{ está em } c. \end{cases}$$

- d_i é o decisor de ilusão, definido pela função:

$$d_i = \begin{cases} 0 & \text{se } \rho \text{ está em } c; \\ 1 & \text{se } (x) \text{ está em } c. \end{cases}$$

- Ab é o conjunto de blocos avaliadores $Ab = \{Ab_h, Ab_{i1}, Ab_{i2}\}$, onde Ab_h é o bloco avaliador de alucinações, Ab_{i1} é o bloco avaliador de ilusões classe 1 e Ab_{i2} é o bloco avaliador de ilusões classe 2.
- Ap é o conjunto de blocos de planejamento automatizado $Ap = \{Ap_h, Ap_i\}$, onde Ap_h é o bloco de planejamento automatizado de alucinações e Ap_i é o bloco de planejamento automatizado de ilusões.
- Δ é a função de transição definido pela tabela abaixo, onde *out* é um estado final, que leva a percepção para fora do modelo de revisão de percepções, ou seja, pode tanto significar o encaminhamento de uma percepção válida para o raciocínio do agente quanto o fim da execução de um ciclo de revisão.

Tabela 1. Função de transição Δ do módulo de ilusão e alucinação.

Estado	0	1
d_a	<i>out</i>	d_h
d_h	Ab_h	d_i
d_i	Ab_{i1}	Ab_{i2}
Ab_h	<i>out</i>	Ap_h
Ab_{i1}	<i>out</i>	Ap_i
Ab_{i2}	<i>out</i>	Ap_i

O módulo de ilusão e alucinação é o artefato principal do modelo. Ele é subdividido em três partes principais: os decisores, os blocos avaliadores e os blocos de planejamento automatizado. Seu comportamento é descrito por uma função de transição que define o estado atual do módulo.

Definição 8. Um bloco avaliador é uma tripla $Ab_x = \langle L, Pf, Cf \rangle$, $x \in \{h, i1, i2\}$, onde:

¹ c é o contexto do agente, de acordo com a definição 2.

- L é uma lista ordenada pelo número de vezes que uma mesma anomalia é dada como entrada;
- Pf é a função de processamento, definida abaixo:

$$Pf = \begin{cases} 1 & \text{se } T_m(A) \leq T_m(V) * (|A| - |A_{pr}|); \\ 0 & \text{caso contrário.} \end{cases}$$

Onde:

- T_m é a função que retorna a média do tempo gasto para processar as percepções de um conjunto;
- A é o conjunto de anomalias, $A(x)$ é um elemento específico x e $|A|$ o número de anomalias do conjunto;
- A_{pr} é o conjunto de anomalias que já foram validadas para serem processadas pela função de processamento neste ciclo de raciocínio (A_{pr} é instanciada vazia a cada ciclo de raciocínio), e $|A_{pr}|$ o número de anomalias desse conjunto.
- V é o conjunto de percepções válidas.
- Cf é a função de limpeza definida abaixo com auxílio da função equação de limpeza Cf , sendo α um coeficiente de limpeza variável que precisa ser definido pela instância implementada do modelo (por padrão, toma-se $\alpha = 1$):

$$Cf = \begin{cases} 1 & \text{se } Ce = \text{Verdadeiro}; \\ 0 & \text{caso contrário.} \end{cases}$$

$$Ce = \sum_{i=1}^{|L|} W_n(L_i) > \alpha \sum_{j=1}^{|L|} W_1(L_j)$$

Onde:

- L é a lista ordenada do bloco, sendo $|L|$ seu número de anomalias e L_i a anomalia i da lista.
- W é a função peso da anomalia L_i definida como $W(L_i) = |L_i|$, sendo $|L_i|$ o peso da anomalia especificada (número de entradas recebidas dessa mesma anomalia na lista). A função W é utilizada para especificar as seguintes funções:

$$(i) W_1(L_i) = \begin{cases} 1 & \text{se } W(L_i) = 1; \\ 0 & \text{caso contrário.} \end{cases}$$

$$(ii) W_n(L_i) = \begin{cases} W(L_i) & \text{se } W(L_i) > 1; \\ 0 & \text{caso contrário.} \end{cases}$$

O bloco de planejamento automatizado recebe como entrada uma anomalia de um dos blocos avaliadores quando a função de processamento retorna verdadeiro, e pode ser definido da seguinte maneira:

Definição 9. Um bloco de planejamento automatizado é uma instância do modelo conceitual de planejamento automatizado (Definição 5).

5. Experimentos

Para este trabalho foram realizados três experimentos com o objetivo de averiguar se o modelo HAIL é funcional ou não. Esses experimentos não possuem um domínio específico, ou seja, não há um ambiente real onde o agente está inserido, ele recebe percepções aleatórias geradas pelo simulador e os planos iniciais do agente não são relevantes para os testes.

Os dois primeiros experimentos foram implementados utilizando o design 2^k fatorial [Jain 1990]. Esse tipo de design consiste em variar k fatores em 2 níveis diferentes, -1 e 1, que são extremos opostos. Os fatores e as variáveis livres utilizadas estão nas Tabelas 2 e 3, respectivamente. A análise do impacto dos fatores foi realizada utilizando a equação de regressão não linear do design 2^k fatorial.

Tabela 2. Fatores utilizados nos experimentos realizados com o modelo HAIL.

Fatores	Sigla	Nível -1	Nível 1
Porcentagem de Percepções Inválidas	PPI	5%	95%
Tempo Médio gasto pelo planejamento Automatizado	TMA	1/2 T	64 T
Tempo Médio gasto em um Ciclo de raciocínio	TMC	01 T	32 T
Número de Percepções recebidas por Ciclo	NPC	01	16

Tabela 3. Variáveis dependentes analisadas nos experimentos realizados com o modelo HAIL.

Variáveis	Motivação
Tempo Virtual Decorrido	Medir desempenho geral do modelo
Planos Criados	Avaliar potencial do modelo de inserir aprendizado em arquiteturas que não o possuem
Percepções Válidas Processadas	Analisar a capacidade do modelo de ganhar desempenho ao longo do tempo

Os experimentos consistem em diversas simulações, que submetem o modelo proposto ao processo de revisão de um grande número de percepções. Uma simulação possui 5000 ciclos de percepção, sendo que cada ciclo pode possuir uma ou várias percepções. Essas percepções podem ser válidas (pertencentes ao contexto do agente) ou inválidas (não pertencentes ao contexto do agente), sendo que a proporção entre o tipo de percepções é definido pela PPI. As percepções são produzidas aleatoriamente por um gerador de percepções.

O gerador de percepções cria as percepções válidas sorteando símbolos que pertencem ao contexto do agente, e as percepções inválidas são criadas utilizando o pacote *RandomWords* [ŚwiEicki 2012], que gera palavras em inglês aleatórias. As percepções são símbolos compostos da forma *corpo(argumento)*, e a quantidade de percepções válidas e inválidas geradas é baseado no PPI e TMA da simulação executada.

Para mensurar o tempo gasto pelos processos do agente, será considerada uma unidade de tempo genérica T, e iremos considerar o tempo virtual da execução das simulações com base nessa unidade.

Tabela 4. Análise dos fatores do experimento 1.

Fator	Efeito Tempo Virtual	Efeito Planos Criados
PPI	12.69%	66.10%
TMC	22.20%	0.50%
TMA	31.64%	2.95%
NPC	1.44%	8.67%
PPI + TMC	4.16%	3.76%
PPI + TMA	24.13%	0.05%
PPI + NPC	1.39%	2.13%
TMC + TMA	0.55%	0.50%
TMC + NPC	0.10%	0.50%
TMA + NPC	0.01%	2.99%
PPI + TMC + TMA	0.53%	3.76%
PPI + TMC + NPC	0.09%	3.75%
PPI + TMA + NPC	0.02%	0.06%
TMC + TMA + NPC	0.54%	0.50%
PPI + TMC + TMA + NPC	0.52%	3.76%

5.1. Experimento 1

O objetivo do primeiro experimento é analisar o impacto da mudança dos fatores no desempenho do modelo através da análise da variação no tempo virtual decorrido e da quantidade de planos criados pelo módulo de planejamento automatizado.

Cada simulação foi repetida dez vezes, e a média dos resultados foi obtida. Isso foi feito para garantir que os valores obtidos não foram afetados pela geração aleatória de percepções. Quando repetições são adotadas no experimento normalmente utiliza-se a metodologia 2^{k_r} fatorial, que leva em conta o erro obtido entre as diferentes execuções da simulação. Todavia, ao fazer a análise de erros o resíduo obtido foi descartável, com ordem de grandeza e^{-12} . Portanto, a metodologia de análise adotada foi a do 2^k fatorial.

Esse experimento demonstrou que o módulo de ilusão e alucinação funciona como previsto: simulações com mais percepções inválidas e menor tempo médio de planejamento automatizado resultam em menos tempo virtual gasto e mais planos criados. Em situações com um grande volume de percepções, implementações que possuem um bloco de planejamento automatizado que consome muito tempo para criar novos planos podem prejudicar o desempenho do agente caso ele seja muito sensível ao tempo. O resultado da análise de impacto dos fatores está exposto na Tabela 4.

5.2. Experimento 2

O segundo experimento foi realizado para analisar o ganho de performance de um agente ao utilizar os planos criados com o bloco de planejamento automatizado. Esse segundo experimento segue a metodologia do primeiro, porém foi realizado apenas uma iteração para cada configuração de fatores (ao invés das dez repetições). Após ser realizada uma simulação com uma configuração de fatores foi realizada uma nova simulação com a mesma configuração, mas usando o agente resultante da primeira. Assim, os planos criados pelo agente foram reaproveitados. Ao invés de analisar quais são os fatores que mais

Tabela 5. Alteração no valor de tempo virtual no experimento 2.

Simulação	PPI	TMC	TMA	NPC	R1	R2	RP
1.1	5%	1	64	1	20750T	20687T	99.70%
1.2	5%	1	64	16	20813T	20750T	99.70%
1.3	5%	32	64	1	168032T	167968T	99.96%
1.4	5%	32	64	16	168032T	168000T	99.98%
1.5	95%	1	64	1	227579T	10859T	4.77%
1.6	95%	1	64	16	304313T	177998T	58.49%
1.7	95%	32	64	1	273536T	162400T	59.37%
1.8	95%	32	64	16	312032T	250048T	80.14%
2.1	5%	1	1/2	1	4874.5T	4876.5T	100.04%
2.2	5%	1	1/2	16	5000T	5000T	100%
2.3	5%	32	1/2	1	152093.5T	152219.5T	100.08%
2.4	5%	32	1/2	16	153435.5T	152887T	99.64%
2.5	95%	1	1/2	1	3228.5T	4955T	153.48%
2.6	95%	1	1/2	16	5000T	4944T	98.88%
2.7	95%	32	1/2	1	48458.5T	157385.5T	324.78%
2.8	95%	32	1/2	16	160000T	140631T	113.77%

influenciam nas variáveis dependentes, comparamos os resultados dessas variáveis antes e depois da criação de novos planos. As variáveis dependentes analisadas foram novamente o tempo virtual decorrido e o número de planos criados.

O principal conclusão obtida a partir dos resultados do experimento 2 é que o principal fator para que o HAIL possa ter impacto no tempo virtual do agente e na quantidade de planos novos que ele recebe é a quantidade de percepções inválidas recebidas. Cenários com poucas percepções não tiveram um grande ganho nessas variáveis. Além disso, foi constatado que em cenários com grande número de anomalias (PPI 95% e NPC 16) o TMA se torna um gargalo para o funcionamento do HAIL, ou seja, quanto mais percepções inválidas são recebidas mais otimizado o bloco de planejamento automatizado precisa ser.

Os resultados das simulações estão expostos nas tabelas 5 (tempo virtual) e 6 (planos criados). As tabelas contém o índice da simulação, os níveis dos fatores, os resultados da primeira e da segunda iteração (R1 e R2, respectivamente) e a relação percentual (RP) entre os resultados, ou seja, a proporção de R2 em relação a R1, obtido através do cálculo $(R2 * 100) / R1$.

5.3. Experimento 3

O terceiro experimento foi executado para observar a capacidade de aprendizado do agente ao longo do tempo. Nele foram executadas 5 simulações com os mesmos fatores dos experimentos anteriores, mas com valores diferentes (como demonstra a Tabela 7). Entre uma simulação e outra o agente não foi recarregado, ou seja, manteve o aprendizado das simulações anteriores.

Nessa simulação foi possível observar que enquanto os planos criados diminuem a cada simulação, pois menos percepções são inválidas (uma vez que o agente aprende

Tabela 6. Alteração no valor de planos criados no experimento 2.

Simulação	PPI	TMC	TMA	NPC	R1	R2	RP
1.1	5%	1	64	1	250.0	249.0	99.60%
1.2	5%	1	64	16	251.0	250.0	99.60%
1.3	5%	32	64	1	251.0	249.0	99.20%
1.4	5%	32	64	16	251.0	250.0	99.60%
1.5	95%	1	64	1	3533.0	93.0	2.63%
1.6	95%	1	64	16	4751.0	2746.0	57.80%
1.7	95%	32	64	1	3548.0	75.0	2.11%
1.8	95%	32	64	16	4751.0	2814.0	59.23%
2.1	5%	1	1/2	1	251.0	247.0	98.41%
2.2	5%	1	1/2	16	502.0	500.0	99.60%
2.3	5%	32	1/2	1	251.0	247.0	98.41%
2.4	5%	32	1/2	16	2935.0	878.0	29.91%
2.5	95%	1	1/2	1	3543.0	90.0	2.54%
2.6	95%	1	1/2	16	9312.0	260.0	2.79%
2.7	95%	32	1/2	1	3541.0	83.0	2.34%
2.8	95%	32	1/2	16	4078.0	0.0	0.0%

com as simulações anteriores), a quantidade de percepções válidas processadas aumenta. Esse comportamento é demonstrado na Figura 5.3. Além disso, o tempo virtual cai na mesma proporção que os planos criados diminuem, uma vez que a criação de um plano é duas vezes mais custosa em tempo do que o processamento de uma percepção válida nesse experimento.

Tabela 7. Valor dos fatores no experimento 3.

Fator	Valor
PPI	50%
TMC	16
TMA	32
NPC	8

6. Conclusão

Agentes são entidades autônomas que, inseridas em um ambiente, são capazes de tomar decisões de maneira autônoma. Para se comunicar com o ambiente no qual estão inseridos, os agentes realizam o processo de percepção, utilizando sensores para receber diversos tipos de informações do mundo ao seu redor. Todavia, as percepções recebidas podem ser incorretas por diversos motivos.

Neste trabalho, apresentamos um modelo genérico capaz de detectar percepções inválidas, classificá-las de acordo com suas características e criar novos planos a partir delas. Esse modelo foi implementado e experimentos foram feitos para mostrar que seu funcionamento era de acordo com o esperado.

Para os trabalhos futuros, é necessário criar versões alternativas de cada um de

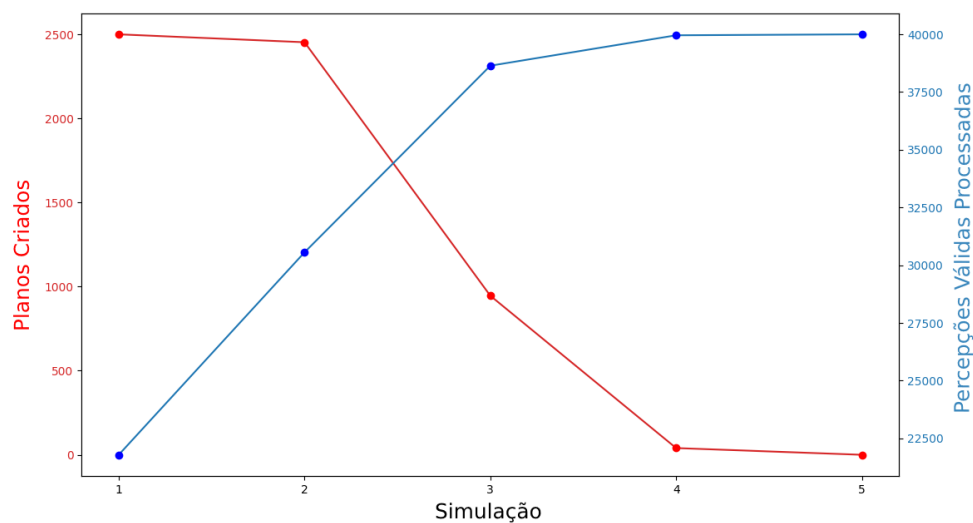


Figura 3. Evolução das percepções válidas processadas e dos planos novos criados ao longo das simulações.

seus componentes para validar se seu comportamento é o melhor possível. Além disso, é necessário testar o modelo em um ambiente real, acoplado a uma arquitetura específica, para testar seu desempenho em diferentes situações.

Referências

- Chrisman, L., Caruana, R., and Carriker, W. (1991). Intelligent agent design issues: Internal agent state and incomplete perception. In *Proceedings of the AAAI Fall Symposium on Sensory Aspects of Robotic Intelligence*. AAAI Press/MIT Press. Citeseer.
- Colton, S., Wiggins, G. A., et al. (2012). Computational creativity: The final frontier? In *Ecai*, volume 12, pages 21–26. Montpelier.
- Crane, T. and French, C. (2017). The problem of perception. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2017 edition.
- Diab, M., Akbari, A., Ud Din, M., and Rosell, J. (2019). Pmk—a knowledge processing framework for autonomous robotics perception and manipulation. *Sensors*, 19(5):1166.
- Dyachenko, Y., Nenkov, N., Petrova, M., Skarga-Bandurova, I., and Soloviov, O. (2018). Approaches to cognitive architecture of autonomous intelligent agent. *Biologically Inspired Cognitive Architectures*, 26:130 – 135.
- Ghallab, M., Nau, D., and Traverso, P. (2004). Chapter 1 - introduction and overview. In Ghallab, M., Nau, D., and Traverso, P., editors, *Automated Planning*, The Morgan Kaufmann Series in Artificial Intelligence, pages 1 – 16. Morgan Kaufmann, Burlington.
- Gibson, J. J. (1950). The perception of the visual world.
- Hayes-Roth, F., Waterman, D. A., and Lenat, D. B. (1983). Building expert system.

- Jain, R. (1990). *The art of computer systems performance analysis: techniques for experimental design, measurement, simulation, and modeling*. John Wiley & Sons.
- Kim, S., Yu, Z., and Lee, M. (2017). Understanding human intention by connecting perception and action learning in artificial agents. *Neural Networks*, 92:29–38.
- Langley, P., Laird, J. E., and Rogers, S. (2009). Cognitive architectures: Research issues and challenges. *Cognitive Systems Research*, 10(2):141–160.
- Moya, L. J. and Tolk, A. (2007). Towards a taxonomy of agents and multi-agent systems. In *SpringSim (2)*, pages 11–18.
- Pangercic, D., Tenorth, M., Jain, D., and Beetz, M. (2010). Combining perception and knowledge processing for everyday manipulation. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1065–1071.
- Russell, S. J. and Norvig, P. (2016). *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited,.
- van Oijen, J. and Dignum, F. (2011). Scalable perception for bdi-agents embodied in virtual environments. In *2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, volume 2, pages 46–53. IEEE.
- Weyns, D., Steegmans, E., and Holvoet, T. (2004). Towards active perception in situated multi-agent systems. *Applied Artificial Intelligence*, 18(9-10):867–883.
- Wooldridge, M. (1999). Intelligent agents. *Multiagent systems*, 35(4):51.
- Ye, P., Wang, T., and Wang, F. (2018). A survey of cognitive architectures in the past 20 years. *IEEE Transactions on Cybernetics*, 48(12):3280–3290.
- ŚwiEicki, T. (2012). Randomwords · pypi. <https://pypi.org/project/RandomWords/>. Acesso 21/07/2020, versão 0.3.0.