Descoberta automática de atributos salientes para obtenção de macro-ações

Rafael Lemes Beirigo[*], Valdinei Freire da Silva[†], Anna Helena Reali Costa[‡]

***Abstract*— A obtenção de soluções para problemas de decisão sequenciais pode ser realizada através de métodos de Aprendizado por Reforço. Entretanto devido à necessidade de interação através de tentativa e erro, esses métodos podem apresentar extrema lentidão. Para contornar esse problema são propostas técnicas de Transferência de Conhecimento que permitem explorar a estrutura hierárquica dos problemas através de abstração temporal na forma de macro-ações. Métodos de descoberta de macro-ações que utilizam eventos salientes apresentaram resultados promissores, entretanto demandam a definição de atributos salientes por parte do projetista, o que é indesejável. O presente trabalho propõe métodos de descoberta automática de atributos salientes no intuito de facilitar a descoberta automática de macro-ações.**

***Keywords*— Aprendizado por Reforço, Transferência de Conhecimento, Macro-ações, Eventos salientes, Atributos salientes.**

I. Introdução

SOLUÇÕES para problemas de decisão sequenciais sem conhecimento *a priori* podem ser obtidas com sucesso através de técnicas de Aprendizado por Reforço (em inglês, *Reinfocement Learning* (RL)) (Sutton e Barto, 1998). Entretanto, as técnicas de RL necessitam que o agente interaja com o ambiente, no intuito de descobrir as ações que, ao serem executadas, maximizam o valor esperado de recompensas recebidas. Como esse processo se dá por tentativa e erro, a aplicação das técnicas de RL na solução de problemas pode apresentar extrema lentidão.

Buscando contornar esse problema, foram propostas na literatura técnicas de Transferência de Conhecimento (Taylor e Stone, 2009; Torrey e Shavlik, 2009; Pan e Yang, 2010; Bergamo *et al.*, 2011; Da Silva e Costa, 2011; Da Silva, Selvatici e Costa, 2011; Matos, Bergamo, Da Silva e Costa, 2011; Matos, Bergamo, Da Silva, Cozman, *et al.*, 2011; Beirigo *et al.*, 2012; Da Silva, Pereira e Costa, 2012; Lazaric, 2012; Koga *et al.*, 2013), em que o agente acumula conhecimento na solução de tarefas de forma a poder utilizá-las em tarefas futuras e dessa forma alcançar um melhor desempenho.

Um elemento que possui papel fundamental no proceso de transferência é a representação do conhecimento adquirido, pois está relacionada a todas as etapas do reuso. Uma forma de representação frequentemente utilizada na literatura que explora a estrutura do domínio é a representação de *habilidades* (em inglês, *skills*) (Gullapalli, Franklin e Benbrahim, 1994; Thrun e Schwartz, 1995). A ideia por trás do conceito é a de que é possível identificar habilidades presentes em uma tarefa que são comuns a outras, permitindo ao agente aprender e solucionar cada uma dessas sub-tarefas de forma independente e posteriormente transferir de forma modular o conhecimento adquirido entre problemas que compartilhem sub-tarefas.

Uma forma usual de formalizar o conceito de habilidades em aplicações de RL tem sido usar o conceito de *macro-ações* (em inglês, *options*) (Sutton, Precup e Singh, 1999), cuja utilização possui a vantagem de permitir a aplicação do arcabouço teórico investigado na literatura de RL. Dessa forma, os algoritmos de aprendizado e provas de convergência se aplicam também quando da utilização de macro-ações. Macro-ações exploram a estrutura hierárquica dos problemas por efetuarem uma abstração temporal da solução.

Embora a abstração por macro-ações possibilite um impacto positivo no desempenho do agente (Chentanez, Barto e Singh, 2004; Konidaris, Scheidwasser e Barto, 2012), a definição das mesmas fica a cargo do projetista, o que pode implicar em uma sobrecarga em tempo de projeto.

Para solucionar esse problema, métodos foram propostos nos quais a descoberta de macro-ações se dá através da utilização de *eventos salientes* (em inglês, *salient events* (SE)) (Chentanez, Barto e Singh, 2004), que correspondem a *sub-goals* presentes na tarefa. Um evento saliente ocorre quando há uma variação *drástica* no valor de um ou mais atributos presentes na descrição fatorada de estados. Esses atributos utilizados para verificar se houve uma alteração drástica são denominados *atributos salientes* (em inglês, *salient features* (SF)) que serão utilizados para detecção dos SE.

Apesar dos resultados positivos apresentados no referido trabalho em relação à melhora de desempenho, o método exige que o projetista defina previamente os SF, o que pode representar uma indesejável sobrecarga de trabalho para o projetista, além de um conhecimento prévio do domínio. Dessa forma, seria interessante investigar métodos que descubram os SF de forma automática.

O presente trabalho propõe dois métodos de descoberta automática dos SF, que avaliam a utilização seletiva dos atributos da descrição fatorada de estados do problema e o consequente impacto no valor da aplicação de uma política.

Nas próximas seções são apresentados a conceituação

[*] R. L. Beirigo, Escola Politécnica, Universidade de São Paulo, São Paulo, SP, Brasil, rafaelbeirigo@usp.br.
[†] V. F. da Silva, EACH, Universidade de São Paulo, São Paulo, SP, Brasil, valdinei.freire@usp.br.
[‡] A. H. Reali Costa, Escola Politécnica, Universidade de São Paulo, São Paulo, SP, Brasil, anna.reali@usp.br.

teórica do trabalho e o detalhamento dos métodos envolvidos na comparação.

II. Background Teórico

A. MDP e Aprendizado por Reforço

Um problema de decisão sequencial modelado por um *Markovian Decision Process* (MDP) (Puterman, 2009) é definido pela quádrupla $\langle S, A, T, R \rangle$, na qual S compreende um conjunto de estados do ambiente, A contém as ações possíveis de serem executadas em cada estado, ambos finitos, $T: S \times A \times S \rightarrow [0,1]$ e $R: S \times A \times S \rightarrow \Re$ representam, respectivamente, a probabilidade de transição e a recompensa recebida ao, partindo de um estado $s_i \in S$ alcançar o estado $s_j \in S$ executando-se a ação $a \in A$.

Uma política $\pi: S \times A \rightarrow [0,1]$ associa a cada par (s, a) uma probabilidade de se executar a ação a estando estado s, sendo o valor de uma política π obtido através da função $V^{\pi}: S \rightarrow \Re$, onde $V^{\pi}(s)$ fornece o valor esperado de recompensas futuras ao se seguir a política π partindo do estado $s \in S$.

A solução para um MDP consiste em uma política ótima π^*, tal que $V^{\pi^*}(s) = \max_{\pi} V^{\pi}(s), \forall s \in S$, ou seja, π^* maximiza o valor esperado de recompensas recebidas para todos os estados aí presentes.

Métodos de RL (Sutton e Barto, 1998) podem ser utilizados para solucionar MDPs. Ao utilizar uma técnica de RL, um agente interage com o ambiente, realizando uma ação para cada estado e observando a recompensa recebida. O intuito do agente é o de maximizar o valor esperado de recompensas recebidas.

Para isso, o agente necessita experimentar ações em estados ainda não visitados através de exploração, mas deve também guiar sua trajetória utilizando o conhecimento já adquirido sobre os estados visitados anteriormente.

Dessa forma, há um compromisso entre a *descoberta* de informação sobre o ambiente e a *utilização* do conhecimento adquirido até o momento, muitas vezes referido como o compromisso *exploração*/*explotação*.

Apesar de utilizados frequentemente e com sucesso para a solução de MDPs, podem apresentar lentidão, por dependerem de interação do agente com o ambiente baseada em tentativa e erro.

Esse fato promoveu a investigação de métodos de transferência de conhecimento (Taylor e Stone, 2009), cuja ideia básica é a de que o agente enfrenta problemas que possuem elementos semelhantes em sua estrutura, que podem ser aprendidos e armazenados de forma conveniente para reuso posterior.

Assim, o agente acumula conhecimento ao solucionar uma sequência de tarefas e transfere esse conhecimento para novas tarefas que, apesar de diferentes, apresentam semelhança com as já resolvidas previamente.

No entanto, para que a transferência seja possível, o conhecimento armazenado pelo agente durante seu aprendizado necessita de uma estrutura que permita esse reuso. A codificação desse conhecimento possui então papel fundamental na aplicação das técnicas de transferência de conhecimento, pois interferem diretamente tanto na possibilidade quanto na eficiência da reutilização do conhecimento adquirido.

Dessa forma, a *representação* do conhecimento desempenha papel fundamental no processo de transferência, pois está relacionado tanto ao conhecimento que está sendo adquirido quanto ao armazenado, estabelecendo a ponte entre a aquisição e o reuso.

B. Transferência de Habilidades

A utilização de *habilidades* é uma forma de representação que procura explorar a estrutura hierárquica presente em tarefas nas quais o objetivo pode ser desmembrado em objetivos intermediários, ou sub-tarefas. A intuição por trás do conceito se baseia no fato de que a solução de uma tarefa possui elementos repetidos na sua execução, ou seja, atividades que precisam ser desempenhadas mais de uma vez durante a solução do problema.

Cada uma dessas atividades pode ser vista como uma habilidade, sendo que, do ponto de vista do agente, é interessante aprender essa habilidade e armazená-la para posterior reuso. Isso porque, uma vez aprendidas, as habilidades podem ser utilizadas posteriormente com mais rapidez, pois não será necessário que o agente as aprenda novamente, o que tornará o aprendizado mais rápido e eficiente.

Por outro lado, uma habilidade pode ser enxergada como uma *sub-tarefa* de aprendizado, a qual pode ser resolvida separadamente, gerando-se soluções para cada sub-tarefa, que são soluções parciais da tarefa principal. Como há uma repetição das mesmas na tarefa a ser resolvida, é possível que habilidades sejam transferidas entre tarefas (Sutton, Precup e Singh, 1999).

Assim, a identificação e aprendizado de habilidades poderia auxiliar o agente tanto no aprendizado de uma tarefa individual, quanto no reuso de habilidades já aprendidas previamente em tarefas que necessitem delas.

Para formalizar o conceito de habilidades pode-se utilizar a estrutura de *macro-ações* (Sutton, Precup e Singh, 1999), que correspondem a políticas parciais que podem ser utilizadas pelo agente na solução das sub-tarefas. Uma vantagem significativa da utilização do formalismo de macro-ações é a possibilidade de utilização do arcabouço teórico desenvolvido na literatura de RL (Sutton, Precup e Singh, 1999).

Formalmente, uma macro-ação é definida por três elementos: uma *política* $\pi: S \times A \rightarrow [0,1]$, uma função que define a *condição de terminação da macro-ação* $\beta: S \rightarrow [0,1]$ e um *conjunto-iniciação* $\Im \subseteq S$, que define a disponibilidade da macro-ação somente para os estados $s \in \Im$.

Assim, a solução de um MDP, inicialmente definida pela obtenção de uma política ótima para o problema, pode ser desmembrada na obtenção de *políticas parciais* referentes às macro-ações aí presentes. Uma vantagem adicional é obtida pela estruturação do problema de forma *hierárquica*, onde o agente ganha a possibilidade de executar não somente ações *atômicas*, ou seja, as presentes em A, mas também as macro-ações, o que pode implicar em um aumento de eficiência no aprendizado (Sutton, Precup e Singh, 1999; Chentanez, Barto e Singh, 2004; Konidaris e Barto, 2007; Konidaris e Barreto, 2009; Konidaris, Scheidwasser e Barto, 2012).

C. *Descoberta de Habilidades*

Apesar do impacto positivo no desempenho de aprendizado do agente implicado pela utilização de habilidades (Chentanez, Barto e Singh, 2004; Konidaris, Scheidwasser e Barto, 2012), o processo de definição das mesmas pode representar uma sobrecarga ao projetista. Dessa forma, foram propostos métodos de descoberta automática de habilidades.

No trabalho presente em (Chentanez, Barto e Singh, 2004) um agente interage com o ambiente enquanto busca descobrir eventos que indiquem a necessidade da construção de uma nova *macro-ação*. Esses eventos são aí denominados eventos salientes (SE) e ocorrem quando, durante a transição de um estado a outro, há uma alteração significativa nos valores de determinados atributos da descrição fatorada de estados, sendo esses atributos denominados atributos salientes (SF), dados *a priori* pelo projestista.

O método de descoberta de macro-ações a partir de SE proposto em (Chentanez, Barto e Singh, 2004) ocorre da seguinte maneira: durante sua interação com o ambiente, a cada transição de estado s_t para s_{t+1}, o agente verifica se, ao atingir o estado s_{t+1} alguma das SF sofreu uma modificação significativa (que ultrapassa um limiar pré-definido). Em caso afirmativo, o agente cria uma macro-ação cujo estado terminal é s_{t+1}, ou seja, uma macro-ação cujo objetivo é alcançar s_{t+1}.

Além disso, analisa-se o conjunto de macro-ações definidas para verificar se alguma delas está disponível para o estado s_{t+1}, mas não para o estado s_t. Caso isso ocorra, o estado s_t é incluído na macro-ação, tendo como resultado a construção incremental das macro-ações com base nas transições que ocorrem durante a interação do agente com o ambiente. Dessa forma, as macro-ações são criadas levando em consideração variações nas SF pré-definidas para o domínio.

O trabalho apresenta resultados positivos na melhora de desempenho do aprendizado do agente, sugerindo que a abordagem seja promissora. Dessa forma, o presente trabalho propõe métodos de descoberta automática de SF, de forma a facilitar o trabalho de aprendizado e transferência de conhecimento entre tarefas de um determinado domínio.

III. Descoberta de Atributos Salientes

No trabalho de (Chentanez, Barto e Singh, 2004), para que o processo de descoberta de macro-ações seja possível, é necessário que os SFs sejam definidos previamente. Ao invés de colocar a responsabilidade dessa tarefa a cargo do projetista, neste trabalho são propostas duas maneiras de descoberta automática das mesmas.

Dada uma descrição fatorada dos estados do ambiente, tem-se que cada estado é definido por um conjunto de atributos $\mathcal{F} = \{f_1, f_2, \dots, f_n\}$. Dessa forma, o problema de Descoberta de Atributos Salientes consiste em encontrar um subconjuto $\mathcal{F}_{salient} \subset \mathcal{F}$ que contém os SFs relevantes para a classe de problemas em questão.

A hipótese analisada neste trabalho é a de que os atributos presentes em $\mathcal{F}_{salient}$ possuem uma relevância significativa na descrição do problema. Dessa forma, sua remoção da descrição dos estados do ambiente implicaria em uma alteração drástica no valor de aplicação da política ótima para o problema.

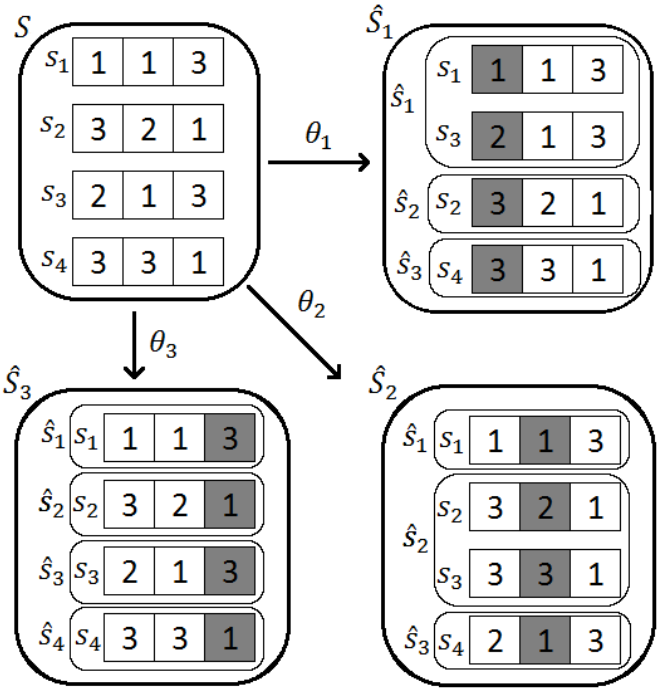
O método proposto neste trabalho para encontrar $\mathcal{F}_{salient}$ consiste em um processo iterativo no qual atributos são removidos da descrição de estados e em seguida avalia-se o impacto dessa remoção no valor da aplicação da política.

O processo de *remoção de um atributo* f_j é comum aos dois métodos propostos e se dá através da função $\theta_j: S \to \hat{S}_j$, que particiona o conjunto S, sendo que cada subconjunto gerado nesse particionamento contém estados que possuem valores idênticos para todos os atributos, desconsiderando-se o valor do atributo removido.

Portanto, dado um conjunto $\mathcal{F} = \{f_1, f_2, \dots, f_n\}$ contendo n atributos, tem-se um total n funções θ_j definidas, cada uma correspondendo à remoção do atributo $j \in \{1, 2, \dots, n\}$.

Para fins de ilustração, na Figura 1 temos um exemplo onde cada estado é definido por três atributos (representados na figura por uma quadrícula), ou seja, $\mathcal{F} = \{f_1, f_2, f_3\}$, sendo que cada atributo pode assumir um valor entre um e três. Como a cardinalidade de $\mathcal{F}$, $|\mathcal{F}| = 3$, é possível definir três funções θ_j, $j = 1, 2, 3$, cada uma responsável pela remoção de um atributo, representado por um sombreamento na quadrícula.

Figura 1 – Aplicação da função $\theta_j: S \to \hat{S}_j$, que realiza o agrupamento dos estados de S, f elementos de $\hat{S}_j$ através da remoção de atributos pela função θ_j.



Ao se aplicar θ_1, gera-se o conjunto $\hat{S}_1$, onde o atributo f_1 foi removido da representação de estados, resultando em uma nova representação onde somente são considerados os valores dos atributos f_2 e f_3. Quando isso ocorre, estados que possuem valores idênticos para f_2 e f_3 são agrupados em um mesmo estado de $\hat{S}_1$.

No exemplo da figura, a remoção do atributo f_1 resultou do agrupamento dos estados s_1 e s_3 em um novo estado $\hat{s}_1$. Isso ocorreu eles possuem os mesmos valores para os atributos restantes, f_2 e f_3. Entretanto, como os estados s_2 e s_4 possuem valores diferentes para os atributos restantes, eles

não puderam ser agrupados. Dessa forma, para cada um deles foi gerado um novo estado unitário na nova representação, sendo que o processo para as demais funções θ ocorre de maneira similar.

A remoção de um atributo implica em um impacto no valor da aplicação da política, pois a descrição de estados utilizada sofreu alteração, implicando em modificações na função-valor. Isso ocorre porque a política que está sendo avaliada pode possuir ações diferentes para cada um dos estados agrupados, e cada estado agrupado pode possuir um valor diferente.

Dessa forma, a remoção de cada um dos atributos possui um impacto no valor de aplicação da política, que pode então ser avaliado. A hipótese investigada neste trabalho é de que os atributos salientes, por sua importância na descrição fatorada de estados, são aqueles cuja remoção da descrição implicará no maior impacto de aplicação da política.

Para o processo de *avaliação*, são propostas duas abordagens, sendo que cada uma delas analisa seletivamente o impacto da remoção de cada um dos atributos, construindo incrementalmente o conjunto $\mathcal{F}_{saliente}$.

A. Avaliação através de abstração de política

A primeira abordagem proposta inicialmente gera uma política definida nos estados presentes em $\hat{S}_j$, gerados por θ_j. Inicialmente agrupam-se os estados em S de acordo com θ_j, sendo então definida uma política $\hat{\pi}_j: \hat{S}_j \times A \rightarrow [0,1]$, levando em conta uma distribuição de probabilidade proporcional ao valor de cada estado agrupado e a respectiva ação, isto é,

$$\hat{\pi}_j(\hat{s}_j, a) = \frac{\sum_{s|\theta_j(s)=\hat{s}_j} \pi(s,a) \cdot C^{\pi}(s)}{\sum_{s|\theta_j(s)=\hat{s}_j} C^{\pi}(s)}.$$

A função $C^{\pi}: S \rightarrow [0,1,2,\dots]$ fornece a quantidade de vezes que um estado $s \in S$ foi visitado durante a aplicação de uma política π. Dessa forma, o valor de $C^{\pi}(s)$, inicialmente nulo para todo $s \in S$, é incrementado de uma unidade a cada vez que o agente atinge estado s durante a aplicação de π.

O impacto $\mathcal{I}_j^{\pi}$ de remover o atributo f_j é dado pelo valor de aplicação da política $\hat{\pi}_j$, isto é, $\mathcal{I}_j^{\pi} = V^{\hat{\pi}_j}$. A definição de $V^{\hat{\pi}_j}$ decorre diretamente da aplicação da política $\hat{\pi}_j$ no problema. Assim, partindo de um estado inicial s_0, o agente segue a ação indicada pela política $\hat{\pi}_j$ para cada estado atingido.

B. Avaliação através de abstração de função-valor

A segunda abordagem proposta considera o impacto $\mathcal{I}_j^{\pi}$ da remoção do atributo f_j para representar a função valor $\hat{V}_j^{\pi}$, onde

$$\mathcal{I}_j^{\pi} = \sum_{s \in S} \left[\hat{V}_j^{\pi}\left(\theta_j(s)\right) - V^{\pi}(s) \right]^2$$

e $\hat{V}_j^{\pi}: \hat{S}_j \rightarrow R$, definida por:

$$\hat{V}_j^{\pi}(\hat{s}_j) = \frac{\sum_{s \in \hat{s}_j} V^{\pi}(s) \cdot C^{\pi}(s)}{\sum_{s \in \hat{s}_j} C^{\pi}(s)}.$$

C. Seleção dos atributos salientes

De posse dos valores de impacto de cada possível remoção de atributo, o processo de seleção dos atributos mais relevantes consiste então em escolher os atributos cuja remoção implicou no maior impacto. Os atributos assim selecionados são então incluídos no conjunto $\mathcal{F}_{saliente}$, similar ao desenvolvido em (Bogdan e Da Silva, 2013).

A limitação do processo de inclusão de atributos em $\mathcal{F}_{saliente}$ pode ser dada pelo estabelecimento prévio da quantidade máxima de atributos para o domínio em questão. Outra forma de limitação possível se daria através da definição de um valor de limiar para o impacto $\mathcal{I}_j^{\pi}$. Nesse caso, quando a variação for menor que o limiar estabelecido, o processo de inclusão para, sendo então gerado o conjunto $\mathcal{F}_{saliente}$. Ainda, pode-se combinar os dois conceitos em uma função que minimiza tanto o impacto $\mathcal{I}_j^{\pi}$, como o tamanho do conjunto $\mathcal{F}_{saliente}$ resultante.

IV. RL como Tecnologia Adaptativa

As técnicas de RL são apresentadas em um conjunto massivo de trabalhos presentes na literatura, possuindo destaque o trabalho (Sutton e Barto, 1998), onde os autores apresentam os fundamentos teóricos e algoritmos relacionados ao controle ótimo adaptativo. Dado que, na aplicação desses algoritmos, o agente busca adequar as funções internas que compreendem o aprendizado adquirido até o momento (na forma de uma função-valor, política ou macro-ação), sua adaptação ao ambiente tende a aumentar à medida em que sua interação com o mesmo lhe permite adquirir conhecimento. Sendo assim, pode-se arguir que as técnicas de Aprendizado por Reforço possuem em sua essência um núcleo de *adaptatividade* do agente em relação ao ambiente à medida em que aprende. Em um trabalho relacionado, onde há um maior destaque para o caráter adaptativo de RL (Sutton, Barto e Williams, 1992), os autores apresentam a idéia de uma forma mais específica, com destaque ao caráter adaptativo do Aprendizado por Reforço.

Essencialmente, as técnicas de Transferência de Conhecimento buscam estender a adaptação do agente a ambientes ainda desconhecidos, através de técnicas que compreendem uma vasta gama de tópicos (Taylor e Stone, 2009; Torrey e Shavlik, 2009; Pan e Yang, 2010). O trabalho apresentado em (Cao *et al.*, 2010) propõe o algoritmo *Adaptive Transfer learning algorithm based on Gaussian Processes*, em que busca-se aumentar o desempenho do agente através de uma análise de similaridade entre as tarefas envolvidas na transferência.

O presente trabalho é baseado em Transferência de Conhecimento através da adaptação de algoritmos de Aprendizado por Reforço aplicados à descoberta automática de macro-ações. Dessa forma, o foco essencial do trabalho é a busca de uma adaptação do agente ao ambiente de aprendizado através da modificação de sua estrutura interna de acordo com a interação que ocorre entre o agente aprendiz e o ambiente em que ele se encontra.

V. Discussão

Apesar da eficiência na aplicabilidade de métodos de aprendizado por reforço, há o problema de lentidão apresentado na solução de problemas mais complexos,

próximos do mundo real, o que impulsionou a investigação de métodos que buscam acelerar o processo de aprendizado.

A identificação de sub-tarefas presentes em uma tarefa a ser resolvida pode permitir que o agente aprenda a resolver cada uma das sub-tarefas de forma independente, o que consiste em um processo relativamente rápido, se comparado à solução da tarefa em si, também possibilitando o aprendizado concorrente de sub-tarefas.

Apesar de resultados positivos e promissores presentes na literatura na utilização de macro-ações para aumento da eficiência de aprendizado, o processo de definição das mesmas representa um empecilho, por ficar a cargo do projetista.

Alternativas de descoberta automática de macro-ações foram propostas na literatura para facilitar o processo de transferência de conhecimento. O presente trabalho se baseia na descoberta através de eventos salientes, que apresentou resultados positivos em relação ao aumento de eficiência de aprendizado.

Apesar desses resultados se apresentarem promissores, a utilização do método baseado em SEs necessita da definição de atributos salientes na descrição fatorada de estados, que, por sua vez, representam sobrecarga de trabalho ao projetista.

A contribuição do presente trabalho é a de propôr dois métodos de descoberta automática de atributos salientes, que por sua vez definem habilidades potenciais. A ideia comum aos dois métodos propostos neste trabalho é a de que os atributos salientes possuem importância significativa na descrição dos estados. A importância individual de cada atributo é medida pelo impacto que sua remoção tem no valor da aplicação de uma política definida no espaço de estados resultante após a remoção do atributo.

Após a avaliação individual de cada atributo, uma ordenação dos mesmos, de forma a selecionar os atributos cuja remoção causa maior impacto no valor de aplicação da política pode ser obtida.

Assim, os atributos selecionados através desse método devem possuir uma importância significativa na descrição dos estados. Esses atributos são então utilizados para compôr o conjunto de atributos salientes para o conjunto de problemas analisados.

A utilização do método proposto tem como resultado esperado um aumento de eficiência na aplicação de métodos de Aprendizado por Reforço para solução de problemas mais complexos através da facilitação da descoberta de macro-ações, cuja utilização já foi demonstrada na Transferência de Conhecimento entre problemas.

Referências

BEIRIGO, R. L. et al. Avaliacao de Politicas Abstratas na Transferencia de Conhecimento em Navegacao Robotica. **Revista de Sistemas e Computa{\c{c}}{\~a}o-RSC,** v. 2, n. 2, 2012.

BERGAMO, Y. P. et al. Accelerating reinforcement learning by reusing abstract policies. VIII Encontro Nacional de Intelig{\^e}ncia Artificial, 2011. p.596-606.

BOGDAN, K. O. M.; DA SILVA, V. F. Forward and Backward Feature Selection in Gradient-Based MDP Algorithms. In: (Ed.). **Advances in Artificial Intelligence**: Springer, 2013. p.383-394.

CAO, B. et al. Adaptive Transfer Learning. AAAI, 2010.

CHENTANEZ, N.; BARTO, A. G.; SINGH, S. P. Intrinsically motivated reinforcement learning. Advances in neural information processing systems, 2004. p.1281-1288.

DA SILVA, V. F.; COSTA, A. H. R. A geometric approach to find nondominated policies to imprecise reward MDPs. In: (Ed.). **Machine Learning and Knowledge Discovery in Databases**: Springer Berlin Heidelberg, 2011. p.439-454.

DA SILVA, V. F.; PEREIRA, F. A.; COSTA, A. H. R. Finding Memoryless Probabilistic Relational Policies for Inter-task Reuse. In: (Ed.). **Advances in Computational Intelligence**: Springer Berlin Heidelberg, 2012. p.107-116.

DA SILVA, V. F.; SELVATICI, A. H.; COSTA, A. H. R. Navigation towards a goal position: from reactive to generalised learned control. **Journal of Physics: Conference Series,** v. 285, p. 012025, 2011.

GULLAPALLI, V.; FRANKLIN, J. A.; BENBRAHIM, H. Acquiring robot skills via reinforcement learning. **Control Systems, IEEE,** v. 14, n. 1, p. 13-24, 1994.

KOGA, M. L. et al. Speeding-up reinforcement learning through abstraction and transfer learning. Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems, 2013. p.119-126.

KONIDARIS, G.; BARRETO, A. S. Skill discovery in continuous reinforcement learning domains using skill chaining. Advances in Neural Information Processing Systems, 2009. p.1015-1023.

KONIDARIS, G.; BARTO, A. G. Building Portable Options: Skill Transfer in Reinforcement Learning. IJCAI, 2007. p.895-900.

KONIDARIS, G.; SCHEIDWASSER, I.; BARTO, A. Transfer in reinforcement learning via shared features. **The Journal of Machine Learning Research**, p. 1333-1371, 2012.

LAZARIC, A. Transfer in Reinforcement Learning: A Framework and a Survey. In: (Ed.). **Reinforcement Learning**: Springer, 2012. p.143-173.

MATOS, T. et al. Stochastic Abstract Policies for Knowledge Transfer in Robotic Navigation Tasks. **Advances in Artificial Intelligence**, p. 454-465, 2011.

MATOS, T. et al. Simultaneous abstract and concrete reinforcement learning. **Proc. of the 9th Symposium on Abstraction, Reformulation and Approximation**, 2011.

PAN, S. J.; YANG, Q. A survey on transfer learning. **Knowledge and Data Engineering, IEEE Transactions on,** v. 22, n. 10, p. 1345-1359, 2010.

PUTERMAN, M. L. **Markov decision processes: discrete stochastic dynamic programming**. Wiley. com, 2009.

SUTTON, R. S.; BARTO, A. G. **Reinforcement learning: An introduction**. Cambridge Univ Press, 1998.

SUTTON, R. S.; BARTO, A. G.; WILLIAMS, R. J. Reinforcement learning is direct adaptive optimal control. **Control Systems, IEEE,** v. 12, n. 2, p. 19-22, 1992.

SUTTON, R. S.; PRECUP, D.; SINGH, S. Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. **Artificial intelligence,** v. 112, n. 1, p. 181-211, 1999.

TAYLOR, M. E.; STONE, P. Transfer learning for reinforcement learning domains: A survey. **The Journal of Machine Learning Research,** v. 10, p. 1633-1685, 2009.

THRUN, S.; SCHWARTZ, A. Finding structure in reinforcement learning. **Advances in neural information processing systems**, p. 385-392, 1995.

TORREY, L.; SHAVLIK, J. Transfer learning. **Handbook of Research on Machine Learning Applications. IGI Global,** v. 3, p. 17-35, 2009.