EXTRAÇÃO DE PREFERÊNCIAS SOBRE COMPORTAMENTOS OBSERVADOS

Valdinei Freire da Silva*, Pedro Lima†, Anna Helena Reali Costa*

*Laboratório de Técnicas Inteligentes
Escola Politécnica da Universidade de São Paulo
São Paulo, SP, Brasil

†Institute for Systems and Robotics
Instituto Superior Técnico - Universidade Técnica de Lisboa
Lisboa, Portugal

Emails: valdinei.silva@poli.usp.br, pal@isr.ist.utl.pt, anna.reali@poli.usp.br

Abstract— Preference Elicitation is a method that helps describing human objectives computationally, for instance, as utility functions, so that an agent can participate in decision tasks representing an individual. The elicitation is made through hypothetical queries to an evaluator, whose answers constraints the possible utility functions. This work introduces the theme of Preference Elicitation over Observed Behaviours, where an evaluator performs evaluations over observed behaviours that an agent executes. The queries (behaviours) are now constrained by the environment dynamics and the observation of the evaluator, presenting some problems: 1) difference between the agent and the evaluator observation; 2) limitation when making queries because of the environment dynamics; and 3) definition of the policy that the agent must execute in a non-deterministic environment in order to get informative evaluations. Those topics are discussed in this paper.

Keywords— Preference Elicitation, Expected Utility Theory.

Resumo— Extração de Preferências (EP) é um método que auxilia na descrição de objetivos em um formato computacional (como, por exemplo, em funções utilidades), para que um agente possa participar nas tomadas de decisões em nome de um indivíduo. A EP é feita através de questões hipotéticas a um avaliador, cujas respostas restringem as funções utilidades possíveis. Esse trabalho introduz o tema de EP sobre Comportamentos Observados, considerando um ambiente com um avaliador que emite avaliações sobre comportamentos demonstrados por um agente. As questões (comportamentos) são agora limitadas pela dinâmica do ambiente e pelo acesso que o avaliador tem do mesmo, originando alguns problemas: 1) diferença na observação do ambiente feita pelo agente e pelo avaliador; 2) limitação das questões que podem ser formuladas devido a dinâmica do ambiente; e 3) definição da política que o agente deve executar em um ambiente não determinista para obter comportamentos que resultam em avaliações informativas. Estes tópicos são discutidos neste artigo.

Keywords— Extração de Preferências, Teoria da Utilidade Esperada.

1 Introdução

A automação de um processo ocorre com a delegação da tomada de decisão a agentes artificiais, que atuam no lugar de um indivíduo. No entanto, em muitos casos, a programação só é possível com a ajuda de um programador, que conhece a arquitetura do agente e o domínio onde o mesmo irá atuar. Para uma real automação é importante eliminar o programador do processo, estabelecendo uma interação direta entre indivíduo e agente na programação deste em prol daquele.

Além do aspecto de interação na programação, também é importante o tipo de informação disponibilizada ao agente, a qual será utilizada para guiar suas decisões. Pode-se considerar o caso de menor autonomia do agente, determinando qual ação deve ser executada para cada situação. No caso de maior autonomia, especifica-se ao agente restrições que devem ser respeitadas para satisfazer a vontade do indivíduo, deixando o agente livre para escolher ações que satisfaçam tais restrições. Um exemplo do último são agentes baseados em utilidade (Russell and Norvig, 1995).

Técnicas de aprendizagem supervisionada podem ser vistas como métodos diretos de programação de política, onde a partir de poucos exemplos de como agir realiza-se generalizações para situações desconhecidas (Mitchell, 1997). Por outro lado, a Extração de Preferências (EP) pode ser vista como um método de programação de agentes baseados em utilidade, onde técnicas baseadas em situações hipotéticas e questões relativas são utilizadas (Dennis, 2003). No entanto, estas técnicas consideram uma compreensão das questões utilizadas, assim como uma linguagem comum entre agente e indivíduo.

Neste trabalho será apresentado o problema de EP sobre Comportamentos Observados (EPCO). Este impõe a restrição de que informações são obtidas apenas sobre situações reais e que devem ser exibidas pelo agente no ambiente. O uso de comportamentos observados apresenta uma maior naturalidade na forma como as questões são colocadas ao indivíduo, mas também apresenta diferentes desafios, que serão discutidos neste artigo.

Na Seção 2 são apresentados alguns dos princípios da Teoria da Utilidade Esperada (TUE), o arcabouço para agentes baseados em utilidade. Na Seção 3 é apresentado o problema de EP do ponto de vista tradicional, enquanto

na Seção 4 é apresentada a versão do problema com comportamentos observados. Finalmente, na Seção 5 apresenta-se algumas conclusões.

2 Teoria da Utilidade Esperada

Considere que se deseja construir um agente *Piloto Automático* para guiar o carro de um passageiro diariamente de sua casa ao trabalho. Uma forma de programar tal agente seria definir para cada situação (cruzamentos) qual decisão tomar (sentido e direção), ou seja, programar o agente com "como satisfazer" o passageiro.

Neste trabalho será explorado outra forma de programação, que é a de programar o agente com "o que é satisfazer" o passageiro e deixar o agente descobrir como fazê-lo. Portanto, deve-se descrever de forma objetiva como o agente deve avaliar cada uma das possíveis ações a tomar.

Uma trajetória pode ser avaliada por vários atributos: tempo do percurso, tamanho do percurso, combustível consumido, qualidade das vias, etc. Deve-se definir qual o compromisso existente entre cada um desses atributos. Por exemplo, o passageiro considera positiva ou negativa uma troca de 20 minutos a mais no percurso pela economia de 1 litro de gasolina?

Além do compromisso entre os atributos, outro aspecto importante é o compromisso com a aleatoriedade. A cada dia o trânsito nas vias de uma cidade é diferente e imprevisível, por exemplo, se ocorre um acidente, o tráfego pode ficar muito mais lento. Então, deve-se definir como o agente considerará opções entre trajetos com pouca aleatoriedade (duração de 60 minutos fixo) e trajetos com mais aleatoriedade (duração de 40 minutos em dias normais, mas mediante acidentes pode durar até mesmo 2 horas).

Se a TUE for utilizada, isso significa definir um valor escalar para cada possível trajetória $\psi \in \Psi$, isto é, a utilidade $u : \Psi \rightarrow \mathbb{R}$ da trajetória; e escolher a política que maximize a utilidade esperada com relação às probabilidades $Pr(\psi|\pi)$ de ocorrência de cada trajetória ψ ao executar a política π, isto é, $V^{\pi} = \sum_{\psi \in \Psi} Pr(\psi|\pi)u(\psi)$. A utilidade esperada mapeia cada política $\pi \in \Pi$ a um valor V^{π} tal que se $V^{\pi'} > V^{\pi''}$ para quaisquer $\pi', \pi'' \in \Pi$ então a política π' é preferível à política π''.

Para que funções utilidades possam representar um indivíduo, entre outras propriedades, ele deve ser coerente, apresentando uma relação de ordem entre todas possíveis trajetórias (Chankong and Haimes, 1983).

3 Extração de Preferências

Um agente artificial não tem um fim em si mesmo, mas representa as expectativas de um indivíduo. A EP consiste em obter junto a um indivíduo suas preferências de modo que possa ser definido um problema de decisão, onde a decisão ótima satisfaça as preferências do indivíduo (Dennis, 2003). Sobre o arcabouço da TUE, a EP consiste em definir uma função utilidade que represente de forma adequada as preferências do indivíduo.

3.1 Questões Hipotéticas

Embora a função utilidade seja ideal do ponto de vista de um problema de decisão, outras formas de representação são mais comuns para um ser humano, por exemplo, a utilização de restrições. No dia-a-dia tal representação é corriqueira. Por exemplo, é fácil observar preferências do tipo: prefiro qualquer outra fruta a limão, prefiro um trajeto que passe pela avenida Brasil, etc.

A representação com restrições apresenta dois problemas. Primeiro, ao estabelecer restrições, nem sempre existirá uma política de ação que obtenha trajetórias que respeite todas restrições. Segundo, se ambientes não determinísticos forem considerados, deve-se determinar restrições baseadas em distribuições de probabilidades para possíveis trajetórias, o que não seria tão natural para um ser humano.

No entanto, a escolha de representação baseada na TUE não impede que as vantagens da representação de restrições sejam utilizadas, isto é, a facilidade de manejo que um ser humano tem com tais representações. A EP considera questões hipotéticas, que ao serem respondidas proporcionam informações sobre as preferências de um indivíduo (avaliador) e sua respectiva função utilidade. Os métodos devem determinar então quais questões devem ser realizadas, buscando principalmente duas propriedades: 1) realizar uma tomada de decisão ótima segundo as preferências do avaliador; e 2) exigir um "baixo" número de interações com o avaliador.

A primeira propriedade não exige uma função utilidade com acurácia em todo o espaço de trajetórias, mas apenas o bastante para garantir a decisão ótima. Se a obtenção de uma decisão quase-ótima for considerada, então um compromisso entre o primeiro e o segundo objetivo deve ser observado. Intuitivamente, quanto mais interações forem realizadas com o avaliador, maior será a acurácia da função utilidade, proporcionando uma melhor tomada de decisão.

Mensurar o segundo objetivo também apresenta dificuldades, pois dependendo do tipo de interação realizada com o avaliador, pode proporcionar um menor ou maior desconforto ao avaliador. Dessa forma, deve-se quantificar melhor o quão exigente é cada uma das interações. Outro ponto sobre o tipo de interação com o avaliador que merece atenção é o psicológico, de quão confiável é a resposta disponibilizada pelo avaliador (Luce and von Winterfeldt, 1994).

O uso de questões hipotéticas, ao invés de situações reais de tomada de decisão, permite minimizar o problema de confiabilidade na resposta do ser humano, além de aumentar a informação disponibilizada em cada interação. Essas questões contemplam trajetórias hipotéticas, que não necessariamente ocorrem no problema de decisão, fazendo uso de distribuições de probabilidades para as trajetórias que não possuem uma política correspondente (Keeney and Raiffa, 1976).

3.2 Tipos de Questões

Tradicionalmente considera-se dois tipos de questões: comparação por pares e questão sobre loterias (Dennis, 2003). A comparação por pares oferece ao programador uma escolha entre duas trajetórias $\psi', \psi'' \in \Psi$. Esta avaliação relativa é fácil de ser respondida quando ambas trajetórias diferenciam muito em valores, uma vez que pessoas podem distinguir facilmente entre opções de forma qualitativa. Note que, apenas valores de ordenação são obtidos, permitindo, por exemplo, que uma ordenação de todos as trajetórias de acordo com a preferência de uma pessoa seja obtida. No entanto, nenhuma informação absoluta sobre valores de trajetórias pode ser inferida. Além disso, se o ambiente for estocástico, a ordenação não é o bastante para se tomar decisões.

A questão sobre loterias envolve avaliações quantitativas. Uma loteria consiste em opções com aleatoriedade $(\alpha'\psi', \alpha''\psi'', \ldots, \alpha^{(n)}\psi^{(n)})$, onde é determinada uma probabilidade de ocorrência $\alpha^{(i)}$ para cada trajetória $\psi^{(i)}$. Na sua forma mais simples, a questão é sobre a preferência entre a trajetória ψ c.p.1 e uma loteria $(\alpha\psi', (1-\alpha)\psi'')$.

A questão sobre loterias permite obter informação sobre a utilidade da trajetória ψ, pois pode-se definir $u(\psi) = \alpha$, onde α é tal que, dadas $\psi^{\perp}$ e $\psi^{\top}$, a pior e a melhor trajetória possível respectivamente, sendo a função utilidade normalizada por $u(\psi^{\perp}) = 0$ e $u(\psi^{\top}) = 1$, a pessoa é indiferente entre a loteria $(\alpha\psi^{\perp}, (1-\alpha)\psi^{\top})$ e a trajetória ψ c.p.1. Tal indiferença não é fácil de ser estabelecida pelo avaliador, mas preferências sobre loterias podem ser utilizadas para estabelecer limiares na função utilidade.

3.3 Estrutura da Função Utilidade

Uma função pode ser representada de várias formas: analítica, curva, tabela, etc. Quando nenhuma informação sobre a função utilidade é conhecida *a priori*, uma opção para descrever a função utilidade é a descrição em tabela. A adoção de uma estrutura paramétrica conhecida para a função utilidade permite reduzir o problema a menos parâmetros para serem determinados.

Ao utilizar um conjunto de atributos Ξ para representar uma trajetória, associa-se a cada trajetória $\psi \in \Psi$ um vetor de atributos $\boldsymbol{\mu}(\psi)$ em um espaço de dimensão $|\Xi|$, isto é, $\boldsymbol{\mu} : \Psi \rightarrow \mathbb{R}^{|\Xi|}$, e diz-se que $\mu_i(\psi)$ é uma medida do i-ésimo atributo do comportamento ψ. O vetor de atributos descreve numericamente uma trajetória de modo a esse vetor ser o bastante para se definir a utilidade de tal trajetória. Na prática, se tivermos duas trajetórias com o mesmo vetor de atributos, mesmo que as trajetórias sejam diferentes sob algum aspecto, as utilidades de tais trajetórias serão iguais. A adoção de atributos, além de ser mais concisa na maioria dos casos, permite o trabalho dentro de um espaço métrico, auxiliando na definição de uma estrutura para a função utilidade.

O avaliador pode ter mais dificuldades em avaliar questões realizadas num espaço n-dimensional, onde vários atributos são considerados e deve-se observar o compromisso entre eles. Ainda, ao adaptar uma superfície neste espaço n-dimensional o número de informações necessárias para se obter uma aproximação satisfatória pode tornar-se muito grande. No entanto, se os atributos apresentam propriedades de independência entre eles, pode-se decompor o problema de obter uma superfície n-dimensional para o problema de obter n curvas bidimensionais e fatores de escala entre as mesmas (Keeney and Raiffa, 1976).

Os atributos $i \in \Xi$ possuem independência aditiva se preferências sobre loterias dos atributos $i \in \Xi$ dependem apenas das respectivas distribuições de probabilidades marginais e não nas distribuições de probabilidades conjuntas. A função utilidade aditiva com n-atributos $u(\mu) = \sum_{i=1}^{n} k_i u_i(\mu_i)$ é apropriada se e somente se a condição de independência aditiva aplica-se aos atributos $i \in \Xi$ (Keeney and Raiffa, 1976).

Para determinar a função $u_i(\cdot)$ faz-se uso de questões hipotéticas, fixando valores para todos os atributos diferentes de i, já que eles são independentes entre si, e varia-se apenas o valor do atributo i. Recorre-se a esse processo para cada um dos atributos.

3.4 Determinando a Função Utilidade

Alguns trabalhos de EP definem a função utilidade impondo restrições geradas após obter uma resposta a uma questão (Boutilier et al., 2005). Por exemplo, se é realizada a questão pela preferência entre a loteria $(\alpha\psi', (1-\alpha)\psi'')$ e a trajetória ψ, sendo preferida a segunda, impõe-se a restrição $\mu(\psi) > \alpha$.

No entanto, se o avaliador não é coerente, deve-se considerar um modelo mais completo. As respostas do avaliador são modeladas por uma distribuição de probabilidades nas possíveis respostas $r \in R_q$ $P(r_t = r | q_t = q, Av)$, onde Av explicita a dependência junto ao avaliador (Chajewska et al., 2000).

Partindo de uma distribuição inicial *a pri-*

ori para um conjunto U de funções utilidades u, pode-se considerar a probabilidade $Pr(u)$ de ocorrência de cada função utilidade u (eventualmente uniforme). Então, dada uma resposta r a uma questão q, pode-se utilizar a seguinte atualização:

$$\Pr(u|q,r) = \frac{\Pr(r|q,u)\Pr(u)}{\sum_{u\in U}\Pr(r|q,u)\Pr(u)}.$$

Boutilier (2003) propõe como tomar decisões com o conhecimento dessas distribuições de probabilidades. Para a avaliação de uma política π pode-se considerar a função valor esperada V_E^π, isto é, a esperança da função utilidade esperada baseada na distribuição $Pr(u)$. Tem-se então $V_E^\pi = \sum_{u\in U} Pr(u)V_u^\pi$, onde V_u^π é a função valor da política π associada à função utilidade u. No entanto estas funções devem ser normalizadas, considerando de forma única, a pior trajetória $\psi_\perp$ e a melhor trajetória $\psi_\top$.

4 EP sobre Comportamento Observados

Um dos pontos principais no qual a EP baseia-se é o uso de questões hipotéticas para obter informações. As questões hipotéticas auxiliam de duas maneiras: 1) permitem escolher pares de trajetórias ou loterias ótimas, isto é, que proporcionem melhoria na acurácia da função utilidade, ou na política ótima que a função utilidade determina; e 2) facilitam a comparação das trajetórias pelo avaliador ao fixar valores de alguns atributos ou ainda utilizando loterias com valores extremos (pior e melhor caso).

No entanto, ao realizar questões e obter respostas de um avaliador, considera-se que o mesmo tenha plena compreensão das questões. Isso envolve não só a questão em si, mas a semântica dos atributos envolvidos. O avaliador deve poder analisar uma trajetória apenas pelos atributos apresentados, exigindo que estes atributos sejam completos e iguais para o avaliador e o agente.

Nem sempre é possível ter conhecimento completo sobre o sistema de avaliação utilizado pelo agente, nem mesmo o avaliador sempre tem acesso completo ao seu próprio sistema de avaliação. A exemplo do que ocorre na área de reconhecimento de padrão (Duda et al., 2000), avaliar exemplos reais, que são interpretados diretamente pelo próprio avaliador, ao invés de uma representação abstrata e com viés, pode ser mais fácil e confortável para o avaliador.

Nesta seção será apresentado o problema de EP sob esta perspectiva, isto é, a da avaliação (respostas) sobre comportamentos observados (questões). Naturalmente este problema coloca as dificuldades apresentadas pela determinação de uma linguagem e atributos bem definidos em outras dimensões: diferentes observações, limitação na escolha das questões e limitação na realização das questões.

4.1 Definição do Problema

O problema EPCO considera um avaliador que deve emitir avaliações sobre trajetórias de comportamentos exibidos por um agente. A Figura 1 mostra um modelo para tal problema. O agente age sobre o ambiente através de ações $a \in \mathcal{A}$, fazendo com que o ambiente transite por estados $s_t \in \mathcal{S}$. O avaliador faz observações o^{Av} do ambiente por períodos $P_i, P_j, \ldots$ e emite avaliações sobre esses períodos. O agente também faz observações o^{Ag} e associa as avaliações recebidas aos comportamentos observados, usando tal informação para construir sua própria função utilidade.

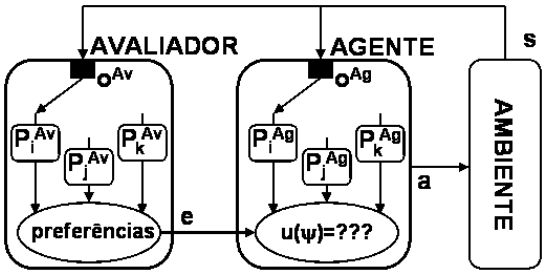

Figura 1: Modelo para EPCO.

Embora nenhuma abstração em forma de atributos seja necessária e o avaliador possa identificar uma situação conforme sua própria percepção, algumas convenções semânticas devem ser feitas para que suas avaliações possam ter significados para o agente. Considere que cada período de avaliação P_i seja composto de n_i trajetórias independentes em termos de avaliação. No exemplo do *Piloto Automático*, cada trajetória poderia ser considerada como o trajeto em um dia, enquanto o período poderia ser considerado uma semana inteira. Isto implica que o trajeto executado na segunda-feira possui uma utilidade independente dos trajetos executados nos outros dias. Dessa forma pode-se redefinir os dois tipos de questões ao trabalhar com comportamentos observados.

Comparação por Pares Após observar dois períodos P_i e P_j cada um com apenas uma trajetória ψ_i e ψ_j respectivamente, o avaliador emite uma resposta entre três opções: `melhor`, `pior` ou `indiferente`. Se a resposta é `melhor`, interpreta-se $u(\psi_i) > u(\psi_j)$. Se a resposta é `pior`, interpreta-se $u(\psi_i) < u(\psi_j)$. Se a resposta é `indiferente`, interpreta-se $u(\psi_i) = u(\psi_j)$.

Questão sobre Loterias Após observar dois períodos P_i e P_j cada um com as trajetórias $\psi_i^1, \psi_i^2, \ldots, \psi_i^{n_i}$ e $\psi_j^1, \psi_j^2, \ldots, \psi_j^{n_j}$ respectivamente, o avaliador emite uma resposta entre três opções: `melhor`, `pior` ou `indiferente`. Se a resposta é `melhor`, interpreta-se $\sum_{k=1}^{n_i} u(\psi_i^k) > \sum_{k=1}^{n_j} u(\psi_j^k)$. Se a resposta é `pior`, interpreta-se $\sum_{k=1}^{n_i} u(\psi_i^k) < \sum_{k=1}^{n_j} u(\psi_j^k)$. Se a resposta é `indiferente`, interpreta-se $\sum_{k=1}^{n_i} u(\psi_i^k) = \sum_{k=1}^{n_j} u(\psi_j^k)$.

4.2 Estrutura da Função Utilidade

O objetivo principal de encontrar uma função utilidade é que o agente possa saber "o que é satisfazer o avaliador" para então definir "como satisfazê-lo". No entanto, nem sempre se pode definir uma estrutura para a função utilidade que represente com acurácia as preferências do avaliador. Por outro lado, não escolher nenhuma estrutura torna o problema muito mais complexo, porém sem atingir uma melhoria correspondente.

Agentes baseados em utilidade possuem uma função utilidade para avaliar o seu próprio desempenho. Algoritmos que maximizam a utilidade esperada às vezes consideram funções utilidades com estruturas pré-definidas para facilitar o encontro de uma política ótima. Esse é o caso, por exemplo, de algoritmos para Processos Markovianos de Decisões (PMD), onde a noção de função utilidade é representada em uma função de custo $c(s_t, a_t)$, que depende apenas do estado e ação atuais (Ross, 1970). Neste caso, a função utilidade apresenta propriedades de independência aditiva e linearidade.

Pode-se limitar a EP a satisfazer condições de otimalidade dependentes da arquitetura do próprio agente, isto é, encontrar uma função utilidade u^* que leve o agente a executar uma política ótima segundo sua arquitetura e as preferências do avaliador (Silva et al., 2007).

4.3 Observabilidade

O fato de agente e avaliador possuirem observações diferentes pode dificultar a interpretação das avaliações do avaliador. Suponha que o agente possua o conjunto de atributos Ξ_{Ag} e uma função estocástica de observação sobre esses atributos $\mu_{Ag} : \Psi \times \Omega \rightarrow \mathbb{R}^{|\Xi_{Ag}|}$, onde Ω representa a dependência sobre uma variável estocástica. Pode-se supor o mesmo para o avaliador, ou seja, o conjunto de atributos Ξ_{Av} e a função μ_{Av}. Então, deve-se explorar que tipo de relação pode haver sobre essas funções de observação.

No caso mais simples, temos que agente e avaliador observam os mesmos atributos e possuem a mesma variável estocástica ($\omega_{Ag} = \omega_{Av} = \omega$). Então, dada uma trajetória ψ, tem-se que $\mu_{Ag}(\psi, \omega) = \mu_{Ag}(\psi, \omega)$. Assim, sempre haverá uma concordância na observação de ambos. Este caso representa um cenário onde ambos utilizam o mesmo sensor e as informações obtidas podem ser utilizadas como na EP tradicional.

Este cenário pode-se complicar se os atributos são os mesmos mas as variáveis estocásticas não o são ($\omega_{Ag} \neq \omega_{Av}$). Este é o caso onde agente e avaliador utilizam o mesmo tipo de sensor, mas cada um possui o seu próprio sensor. Neste caso só se pode garantir igualdade estocástica, isto é, $\mathrm{E}_{\omega_{Ag} \sim \Omega}[\mu_{Ag}(\psi, \omega_{Ag})] = \mathrm{E}_{\omega_{Av} \sim \Omega}[\mu_{Av}(\psi, \omega_{Av})]$. Informações sobre uma situação devem ser obtidas repetidas vezes para serem utilizadas como na EP tradicional, ou considerar um modelo estocástico de respostas integrando tais diferenças.

O cenário mais complicado é quando os atributos são diferentes para agente e avaliador. Nesse caso, nenhuma relação direta pode ser feita com relação a ψ, $\mu_{Ag}(\cdot)$ e $\mu_{Av}(\cdot)$. Alguma relação pode ser estabelecida mediante distribuições fixas na execução dos comportamentos, por exemplo, se a trajetória ψ é sempre obtida quando executada uma política fixa π_ψ. Dessa forma, convém-se estudar conjuntos de políticas que possam apresentar uma maior relação entre as observações do agente e avaliador, permitindo a extração de informações úteis nas avaliações recebidas.

4.4 Trajetórias versus Política

Na EP tradicional um algoritmo determina quais questões serão realizadas e realiza tal questão ao avaliador. As questões são em geral compostas por vetores de atributos, que por sua vez representam trajetórias. Mesmo que as trajetórias limitem-se a trajetórias factíveis, estas trajetórias não precisam ser executadas no ambiente. O fato de executar uma trajetória em um ambiente traz o problema de como executar tal trajetória, já que quando o ambiente utilizado é estocástico, não se pode exibir uma trajetória c.p.1.

Uma solução para esse problema é definir um PMD, onde estende-se o estado do ambiente com as crenças sobre as funções utilidades (Boutilier, 2002) e, ainda, o histórico de atributos até momento. Dessa forma, pode-se sempre escolher a ação que traz mais informação (melhores perguntas) na média. Porém, tal solução apresenta uma alta complexidade computacional. Uma segunda opção é trabalhar com escolhas *off-line* de trajetórias factíveis e alterar *on-line* a política executada conforme a sub-trajetória obtida (Mannor and Shimkin, 2004). Se forem consideradas avaliações sobre loterias, quanto maior o número de trajetórias observadas, maior a probabilidade de demonstrar a loteria planejada.

4.5 Transferência

A limitação de demonstrar uma trajetória para que a mesma possa ser observada, além de limitar as possíveis questões aos ambientes onde as trajetórias estão sendo demonstradas, também exige que o agente tenha acesso a tal ambiente.

Nem sempre o agente tem acesso ao ambiente onde ele será utilizado. Suponha que se deseja utilizar o agente *Piloto Automático* para guiar em Londres, mas que ele deve extrair as preferências do avaliador utilizando as ruas de São Paulo. Como definir uma função utilidade com acurácia adequada em um ambiente, quando pretende-se utilizar o agente em um segundo ambiente?

Se uma medida de distância for determinada, pode-se escolher de forma controlada ambientes onde o agente sofra menos com as dificuldades apresentadas pelo fato de comportamentos serem observados. Por exemplo, escolhendo ambientes deterministas e com igualdade na observação de avaliador e agente, de forma que a função utilidade ainda apresente acurácia no ambiente real.

5 Conclusão e Trabalhos Futuros

O problema EPCO é um arcabouço para programação de agentes baseados em utilidades que além de não requisitar um programador, também é mais natural em termos das considerações sobre o avaliador. No entanto, alguns aspectos teóricos devem ser melhor compreendidos para facilitar a adaptação dos algoritmos já desenvolvidos para o problema de EP tradicional, assim como para estender a teoria já desenvolvida para o mesmo.

Devido a diferença entre as observações do agente e do avaliador, um estudo acurado para definir quando é possível extrair alguma informação deve ser realizado. Isto permitirá a definição de diferentes cenários com respectivas dificuldades que os algoritmos para EPCO terão que abordar.

Algumas técnicas também são necessárias para permitir a adaptação dos diversos algoritmos já existente para o problema de EP tradicional. Primeiro, uma estrutura que permita aprender automaticamente o espaço de trajetórias possíveis e facilite o planejamento de questões é essencial para o desenvolvimento de qualquer algoritmo de EP. Segundo, técnicas para combinar políticas para atingir multiobjetivos podem representar uma boa alternativa com baixo custo computacional para exibir comportamentos desejáveis (Sprague and Ballard, 2003). A criação de tais técnicas podem proporcionar uma facilidade na adaptação dos algoritmos tradicionais de EP, e ainda permitir a criação de novos algoritmos específicos para o problema de EPCO.

Agradecimentos

Este trabalho foi conduzido sob o projeto Multibot CAPES / GRICES (Grant no. 099/03). Valdinei F. Silva também agradece a FAPESP (proc. 02/13678-0) e CAPES (proc. BEX-3388/04-2).

Referências

Boutilier, C. (2002). A pomdp formulation of preference elicitation problems, *Eighteenth national conference on Artificial intelligence*, American Association for Artificial Intelligence, Menlo Park, CA, USA, pp. 239–246.

Boutilier, C. (2003). On the foundations of expected expected utility, *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI-03)*, Acapulco, pp. 285–290.

Boutilier, C., Patrascu, R., Poupart, P. and Schuurmans, D. (2005). Regret-based utility elicitation in constraint-based decision problems, *International Joint Conference on Artificial Intelligence (IJCAI'05)*, pp. 929–934.

Chajewska, U., Koller, D. and Parr, R. (2000). Making rational decisions using adaptive utility elicitation, *AAAI/IAAI*, pp. 363–369.

Chankong, V. and Haimes, Y. Y. (1983). *Multiobjective Decision Making: Theory and Methodology*, North-Holland, New York.

Dennis, B. (2003). A survey of preference elicitation, *Technical report*, Computer Science Department, North Carolina State University.

Duda, R. O., Hart, P. E. and Stork, D. G. (2000). *Pattern Classification (2nd Edition)*, Wiley-Interscience.

Keeney, R. L. and Raiffa, H. (1976). *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*, Wiley, New York.

Luce, R. D. and von Winterfeldt, D. (1994). What common ground exists for descriptive, prescriptive, and normative utility theories, *Management Science* **40**(2): 263–279.

Mannor, S. and Shimkin, N. (2004). A geometric approach to multi-criterion reinforcement learning, *J. Mach. Learn. Res.* **5**: 325–360.

Mitchell, T. M. (1997). *Machine Learning*, WCB/McGraw-Hill, San Francisco, California.

Ross, S. M. (1970). *Applied probability models with optimization applications*, Holden-Day, San Francisco.

Russell, S. and Norvig, P. (1995). *Artificial Intelligence: a Modern Approach*, Prentice-Hall, New Jersey.

Silva, V. F. d., Lima, P. and Costa, A. H. R. (2007). Eliciting preferences over observed behaviours based on relative evaluations, *To be published in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'07)*.

Sprague, N. and Ballard, D. (2003). Multiple-goal reinforcement learning with modular sarsa(0), *International Joint Conference on Artificial Intelligence.*, pp. 1445–1447.