

Autonomous Agents and Multiagent Systems

2nd Semester - 2015/2016

Agents in Wolf Pack



Group 17 - <Taguspark>

*<André Lima>* - <70572>

*<Samuel Gomes>* - <76415>

*<André Fonseca>* - <84977>

# Abstract

**A solução de problemas multiagentes, principalmente quando envolvem aprendizagem por reforço tem enorme potencial mas são de difícil implementação**. Este projeto visa criar uma solução a um problema multiagente do género de perseguição. Implementámos 3 tipos de soluções: reativas, deliberativas e de aprendizagem. **Os resultados foram satisfatórios principalmente com a solução deliberativa e com a aprendizagem**. Com estas técnicas comprovamos como é possível resolver de forma rápida e eficiente este problema multiagente.

**Keywords**: *multiagentes, arquiteturas reativas, arquiteturas deliberativas, arquitetura de aprendizagem, Q-learning modular*

# Index

1. Introduction 4

2. The Scenario 4

3. Agent Architectures / Algorithms 4

3.1 Reactive Architecture 4

3.2 Deliberative Architecture 4

3.3 Hybrid Architecture 5

3.4 Coordination, Cooperation, Negotiation 5

3.5 Emotions Component 5

3.6 Learning Component 5

4. Comparative Study 5

5. Conclusions 6

6. References 6

# Introduction

In this section you should describe the purpose of the work, the problem that is being addressed, the approach and structure of the report.

**Notes:**

* The Introduction *is not* a copy of the abstract;
* Reports should not exceed 20 pages, including the cover, abstract, index and references, according to the format shown in this document.
* All references used in the development of the project must be clearly marked and placed in the References section.

No âmbito da Unidade Curricular de *Agentes Autónomos e Sistemas Multi-Agente*, o grupo 17 escolheu realizar o desafio denominado por *Wolf Pack*. **Este desafio consistiu na elaboração de três sistemas: (1) um sistema reativo, (2) um sistema deliberativo e (3) um sistema de aprendizagem.**

Um agente reativo é um agente que, tomando em conta apenas as suas percepções do ambiente, reage ao que acontece à sua volta. Neste projeto, o nosso agente reativo possui um número variável no seu estado interno**. Este número define o seu campo de visão e é usado como se definisse o alcance dos olhos do lobo.**

Um agente deliberativo possui um estado interno, que memoriza certas propriedades do ambiente que o rodeia e permite guardar estados do ambiente dentro da memória do agente. Os nossos agentes deliberativos possuem um plano de pathfinding e dois vetores. Cada vetor no seu estado interno é referente às suas comunicações com outros lobos.

**A parte mais importante deste desafio é um problema de *Q-learning* modular**. Este sistema de *Q-learning* trata cada lobo como um *Independent Learner* e cada lobo tem um módulo em que toma por referência um parceiro e a presa. Esta implementação permitiu dar a volta ao problema do gigante número de estados que um problema de *Q-learning* monolítico criaria. Assim conseguimos segmentar um problema grande cheio de estados e melhorá-lo aproveitando-nos de indireção.

**O desafio do *Wolf Pack* descreve o problema no qual quatro lobos têm de caçar uma presa, encurralando-a entre eles**. Esta presa pode ter vários movimentos e o lobo também. Os vários diferentes tipos de lobos vão nos permitir observar comportamentos emergentes, de cooperação e de alcateia em relação á caçada da presa.

**Neste documento estão descrições detalhadas das arquiteturas e implementações dos três tipos de agentes, bem como do cenário e também dos resultados das implementações, as suas conclusões e comparações com base em valores parametrizáveis.**

# The Scenario

**Este projeto é composto por 2 tipos de agentes num total de 5 agentes: 4 agentes do tipo lobo e um agente do tipo ovelha**. **Ambos os agentes partilham as mesmas ações: mover para cima, mover para baixo, mover para a esquerda, mover para a direita**. Só é possível mover para tais posições se estiverem desocupadas. Os agentes com aprendizagem têm uma quinta ação que é ficarem parados.

**A ovelha é extremamente simples, capaz apenas de se mover de forma aleatória pelo mapa.** Acrescentámos também uma opção para uma versão reativa da ovelha capaz de fugir aos lobos próximos. Nesta situação a ovelha tem como sensor um campo de visão capaz de distância d capaz de identificar os lobos. E faz como que o contrário do que os lobos fariam ao vê-la.

**Os lobos são muito mais complexos**. Na sua versão reativa têm apenas uma variável associada, o seu *field of view* (não confundir com memória visto que não guarda informação apresentada ao longo do tempo. A única coisa que o lobo “memoriza” é o seu alcance máximo de visão). Com ela, o lobo pode identificar e saber (naquele momento) onde estão os outros lobos e a ovelha.

Quando deliberativos, os lobos têm uma ação extra (o uivar) que transmite uma mensagem aos outros lobos indicando o lado pelo qual planeiam flanquear a ovelha, comunicando assim o seu plano e os planos de outros lobos com quem já comunicaram. Para poderem saber o plano de outros lobos e a quem já enviaram mensagem, têm agora memória sobre estes vários valores.

Por último, na sua versão com aprendizagem, os lobos perdem novamente a capacidade de enviar mensagens mas têm agora conhecimento dos múltiplos *Q-values* possíveis para cada par “outro lobo-presa” e, além disso regem-se por receber recompensas no ambiente à sua volta.

# Agent Architectures / Algorithms

In this section you should describe the several architectures and algorithms according to what is required in the project’s description, *e.g.*, reactive, deliberative / BDI, hybrid, etc. The text should present a conceptual description mentioning aspects of implementation only if necessary for the understanding of the concepts. Some of the aspects to describe in each subsection are:

## Reactive Architecture

* The *<perception> \* → <actuator>* rules developed for each type of agent;
* The description of the arbitration mechanism of rules, *i.e.*, how they are ordered;
* The justification for any internal state variable that was used.

Para o problema proposto, os nossos agentes, os lobos, possuem um estado interno. Um campo de visão que é, no máximo e em cada dimensão, metade do tamanho do mundo nessa dimensão. Este campo de visão é feito como um quadrado de dimensão d x d à volta do agente. Usando este estado interno, os agentes conseguem detetar se a presa está dentro do seu campo de visão. Usámos um número que representa d como estado interno, esse número foi usado como estado interno porque é uma característica dos agentes.

Em relação às regras perceção -> ação, estas foram desenvolvidas igualmente para todos os lobos.

Neste caso, cada lobo verifica se a presa está no campo de visão, e tenta ir na sua direção. No entanto em caso de empate entre distâncias em x e y é escolhida a direção aleatoriamente

Em último caso, quando algum lobo não consegue ver a presa, apenas escolhe aleatoriamente uma de quatro ações: andar em frente, para trás, para os lados.

## Deliberative Architecture

No que diz respeito a esta parte do projeto, chegámos à conclusão que uma arquitetura BDI seria demasiado complexa para o tipo de agentes que considerámos.

Assim não seguimos a estrutura BDI dado que a implementação desta técnica se revelou opcional.

A nossa implementação resume-se a três módulos:

* **A comunicação entre os agentes** de modo a obterem um consenso sobre o seu subobjetivo;
* Procura de caminhos: **recorremos ao algoritmo A\*** tal como na aula de laboratório 3, para procurar caminhos mais curtos até cada uma das posições adjacentes à ovelha (norte, sul, este e oeste). **Assim, usámos como heurística para atualizar os custos, a distância euclidiana entre pontos dado que mapeamos a solução ótima ao caminho mais curto.** Os nós da procura são implementados como uma lista com 3 campos que representam o f, o g e uma posição na grelha (representada por uma lista com coordenadas);
* Estado wander: quando um lobo não tem informação necessária para poder cumprir o seu objetivo, este entra num estado que consiste em vaguear pelo mundo em zigzag de modo a voltar a obter informação o mais rapidamente possível. Esse movimento é gerado a partir de operações locais que usam os limites do mapa para poderem mudar de direção quando conveniente. Esta operação têm em conta o *field of view* do agente, dado que quanto maior o *fov* do agente **menor a frequência com que precisa de passar pelo centro do mapa** (dado que adquire mais dados sobre o mundo a cada momento).

**O agente deliberativo utiliza todos estes módulos de forma a minimizar o tempo demorado a encontrar a presa**. A prioridade para o agente neste sistema é formar um plano em conjunto com os outros lobos. Isto é, obter uma posição adjacente à ovelha para a qual ele e só ele é responsável de ir e depois manter-se nessa posição. Se o agente não tem plano, assim que ele vê a ovelha, este é formado, e assim que possível tenta comunica-lo com os outros lobos. **Se não consegue encontrar a ovelha, usa o movimento definido no estado *wander***. Por último, se vê outros lobos e não lhes enviou mensagem sobre o seu objetivo atual, realiza esta tarefa nesse “tick”.

## Coordination, Cooperation, Negotiation

**No sistema deliberativo, os lobos são capazes de comunicação de forma a coordenarem entre eles um plano de ataque. Cada lobo tem conhecimento sobre o flanco pelo qual pretende atacar a presa, os flancos escolhidos pelos outros lobos atacar e a informação sobre se os outros lobos ouviram ou não a sua mensagem.**

Quando um lobo encontra pela primeira vez a ovelha, este decide encurralá-la pelo lado que lhe está mais próximo. No entanto, se tem conhecimento que outro lobo já escolheu esse plano de ataque, decide por o segundo melhor, terceiro ou até mesmo quarto, dependendo dos planos já usados.

**Assim que o lobo encontra um outro lobo, se ainda não lhe enviou a mensagem do seu plano de ataque, então envia essa mensagem para ter a certeza que o outro lobo sabe**. Enviar a informação é algo penoso ao lobo pelo que ele tem de “parar e uivar” gastando um turno nessa ação. Decora também que a mensagem foi enviada àquele colega para não a enviar novamente sem necessidade.

**Existem situações em que ocorrem conflitos e dois lobos escolhem atacar o mesmo flanco.** Para resolver esse impasse, quando um lobo recebe uma mensagem que indica que já alguém escolheu o seu plano**, o lobo desiste da sua ideia em prol da do colega e tem de escolher um flanco novo dos ainda disponíveis**. Naturalmente, isso significa que terá novamente de avisar todos os outros lobos do seu plano novamente.

Esta última ação por parte dos lobos é bastante interessante visto que leva a um novo comportamento emergente em que, quando próximos de apanhar a ovelha, os lobos trocam algumas mensagens entre si até terem planeado completamente quem ataca por que lado.

## Learning Component

If learning techniques were used in the project, detail the aspects on which learning was incorporated, what learning algorithms were used and how they were implemented in the specific case of the project.

Devido ao aumento exponencial do espaço de estados de” Q-learning” quando aumenta o número de agentes, implementámos uma técnica diferente: “Q-learning modular”.

**Note:** This section is especially important for the project “*Wolf Pack”*.

# Comparative Study

In this section you should present the comparative study made between the several approaches explored in the creation of agents. In particular, the study should be able to provide conclusions about which of the approaches explored contributed more to the "success" of the agents.

Some of the aspects to describe this section are:

* The approaches that will be compared in the tests;
* The metrics used to determine the "success" of each agent / team;
* The tests that were designed;
* The results obtained in each studied condition (charts, tables, etc.);
* A statistical analysis of the results;
* Conclusions (limitations / advantages of each approach, etc.).

**Notes:**

* Any conclusions must be substantiated by the results of the experiments;
* This section applies to all projects, *i.e.*, even if the requested comparative study does not refer to the several architectures used, the general indications described above remain valid, only the object of the comparative study changes.

Para podermos qualificar as várias implementações, conduzimos uma bateria de testes que nos permitiu obter um melhor entendimento sobre o funcionamento prático dessas implementações.

Assim para a arquitetura reativa considerámos os seguintes parâmetros:

- O tempo que o sistema demora a encontrar uma solução;

…

E como consequência, obtivemos os seguintes resultados:

[IMAGEM]

A análise da imagem permite concluir que pilas

Para a arquitetura deliberativa foram considerados os seguintes parâmetros:

…

Obtivemos para este caso, os seguintes resultados:

[IMAGENS]

Assim conclui-se que pilas

Para a arquitetura com aprendizagem foram considerados parâmetros mais específicos:

…

[IMAGENS]

Pelo que podemos afirmar que pilas

Assim, pela análise dos resultados é de notar que para problemas de menor dimensão, a solução reativa teve bons resultados, perdendo eficiência para problemas com uma complexidade mais elevada.

Por outro lado podemos observar que os outros sistemas são mais eficientes na resolução de problemas de elevada complexidade, notando que a estrutura com aprendizagem consegue obter mais eficiência devido ao facto de após um dado número de episódios se adaptar ao ambiente onde se encontra inserida.

# Conclusions

This section should summarize the proposed solution to the problem presented in the Introduction. In particular, you should present an interpretation of the results described in the Comparative Study section.

**Note:** It is also expected in this section to describe how the work presented could be extended in order to motivate future work.

# References

This section should list all the references considered relevant to the understanding of the work presented. Here are some examples of reference formatting for several types of publication:

**For a book:**

1. D. Patterson and J. Hennessy, *Computer Organization and Design: The Hardware/Software Interface*, San Francisco, CA, USA: Morgan Kaufmann Publishers, 1997.

**For a journal article:**

1. B. Kernighan and S. Lin., An Efficient Heuristic Procedure for Partitioning Graphs. *The Bell System Technical Journal*, Vol. 49, No. 2, pp. 291-307, February 1970.

**For a paper included in the proceedings of a conference:**

1. F. Fallah, S. Devadas, and K. Keutzer. OCCOM: Efficient Computation of Observability-Based Code Coverage Metrics for Functional Simulation. In *Proceedings of the Design Automation Conference*, pp. 152–157, June 1998.

**For a dissertation:**

1. D. Cheng. *Power Estimation of Digital CMOS Circuits and the Application to Logic Synthesis for Low Power*. PhD thesis, University of California at Santa Barbara, December 1995.

**For a technical report:**

1. E. Sentovich. *SIS: A System for Sequential Circuit Synthesis.* University of California, Berkeley, April 1992.

**For a webpage:**

1. *Instituto Superior Técnico, Official webpage*. Retrieved from: <http://tecnico.ulisboa.pt/>. Last accessed April 2016.