CBR

Índice

[Introdução 2](#_Toc139968029)

[Vantagens 2](#_Toc139968030)

[Desvantagens 2](#_Toc139968031)

[Especialmente indicado para 2](#_Toc139968032)

[Ciclo de Aamodt & Plaza 3](#_Toc139968033)

[Retrieve 4](#_Toc139968034)

[Fórmulas 6](#_Toc139968035)

[Reuse 8](#_Toc139968036)

[Revise 8](#_Toc139968037)

[Retain 9](#_Toc139968038)

[Princípios orientadores 9](#_Toc139968039)

[Biblioteca e Casos Iniciais (Seed Cases) 10](#_Toc139968040)

# Introdução

* Baseia-se na representação de casos (ocorrências contextualizadas) que representam recordações (experiências passadas)
* Cada caso é composto por uma descrição e pela respetiva solução
* A resolução de um novo problema consiste em encontrar um caso passado semelhante ao atual e apresentar a solução diretamente ou adaptada ao novo contexto

## Vantagens

* Funcionam de forma mais consentânea com o raciocínio pericial, i.e., baseiam-se em experiência passada
* A aprendizagem é simples: baseia-se na adição de “casos relevantes” à Base de Conhecimento (Case Library)
* A sua implementação evita a fase de extração de conhecimento (pode ser essencialmente baseada em documento que descrevem casos passados)

## Desvantagens

* Dificuldade na adaptação de casos:
  + Normalmente é realizada por regras (outras possibilidades consoante o domínio). Contudo é difícil por natureza e assunto de investigação.

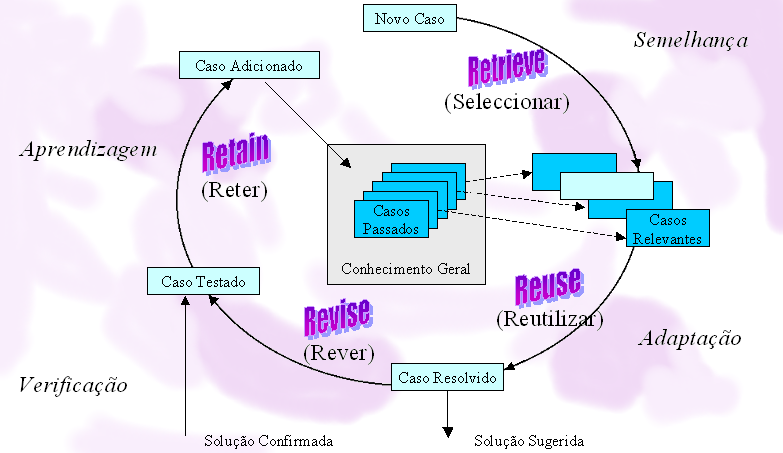
## Especialmente indicado para

* Domínios que mudam com facilidade (dinâmicos)
* Necessidade de aprendizagem constante e automática
* Domínios complexos e/ou mal compreendidos (caso típico: CLAVIER)
* Domínios que dependam claramente de experiências passadas (p.e. aconselhamento jurídico baseado em ações já julgadas)

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

# Ciclo de Aamodt & Plaza



Ciclo de 4 fases:

* Retrieve
  + Procurar o caso ou os casos mais semelhantes ao atual
* Reuse
  + Reutilizar a informação que esse caso contém para resolver o problema atual. A solução pode ser exatamente a mesma ou ter de ser adaptada
* Revise
  + Testar se a solução proposta é válida. Se inválida, proceder à sua correção
* Retain
  + Reter o conhecimento contido no caso atual se contiver algo de diferente; se ensinar uma lição

Terminologia:

* Biblioteca de Casos (Case Library): A estrutura do sistema onde estão armazenados os casos (vulgarmente uma Base de Dados Relacional, em especial nos *shells* comerciais)
* Seed Cases: Casos iniciais carregados no sistema para início da operação

## Retrieve

1. Selecionar um conjunto de casos (possivelmente) relevantes para a solução do problema atual
2. Selecionar o melhor ou melhores de entre eles

**Fase 1**

Vulgarmente os casos contêm pares atributo-valor, p.e. (Tensão, 220V)...

Nesta fase são selecionados casos com contornos (perfil) que parece semelhante ao do caso atual por terem índices iguais.

Contudo, a indexação para CBR difere da indexação vulgar de RDBs:

* RDBs
  + Os índices são de natureza funcional (surgem em consequência do desenho e existem frequentemente em chaves primárias e candidatas)
* CBR
  + Os índices visam a rápida identificação dos casos cujos atributos, contornos, sugerem que sejam relevantes para a resolução do problema atual

Para permitirem predizer a utilidade dos casos da biblioteca, os índices:

* Têm de ser abstratos para funcionarem numa variedade de situações futuras
* Têm de ser suficientemente concretos para impedirem a seleção de casos que não serão uteis na resolução de um problema

**Fase 2**

Habitualmente implementada pelo “algoritmo do vizinho mais próximo” (**nearest neighboor**)muito característico do paradigma CBR.

Calcula uma semelhança global entre o caso atual e cada um dos casos relevantes selecionadas na fase 1.

Baseia-se nos valores dos atributos dos casos em consideração.

Na sua versão mais simples:

* É calculada uma distância entre valores de atributos correspondentes
* Estas distâncias são adicionadas
* A sua soma é transformada numa medida de semelhança

**Exemplo**

Considerem-se dois casos que envolvem dois transformadores.

Atributos: Tensão de saída e Intensidade corrente do secundário

Valores: **Caso1:** A1 = 12V, A2 = 4A **Caso 2:** B1 = 15V e B2 = 2A

Uma imagem com texto, file, diagrama, captura de ecrã

Descrição gerada automaticamente

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

### Fórmulas

Distância Linear para *n* atributos

Uma imagem com Tipo de letra, texto, branco, escrita à mão

Descrição gerada automaticamente

Distância Euclidiana para *n* atributos

Uma imagem com texto, Tipo de letra, escrita à mão, branco

Descrição gerada automaticamente

Uma imagem com texto, Tipo de letra, captura de ecrã, branco

Descrição gerada automaticamente

O valor da semelhança global, S, é normalmente definido como a diferença de uma **distância normalizada**, D, para 1:

Uma imagem com Tipo de letra, símbolo, logótipo, Gráficos

Descrição gerada automaticamente

Uma imagem com texto, Tipo de letra, captura de ecrã, file

Descrição gerada automaticamente

## Reuse

1. Propor a solução de um caso passado como solução para o caso atual (***null adaptation***)

**OU**

1. Adaptar a solução de um caso passado ao caso atual
   1. **Adaptação Transformacional:** são definidos operadores transformacionais {T} – p.e. regras, algoritmos, expressões matemáticas – que guiam a adaptação da solução passada ao caso atual
   2. **Adaptação Derivacional:** um método usado no passado para adaptar uma solução é reutilizado; a cada caso tem de se associada alguma forma de descrição desses métodos

## Revise

1. Avaliação da solução proposta
   1. Por um **professor externo**
      1. Diagnóstico: um técnico ou médico confirma ou não a solução proposta pelo sistema
      2. Planeamento: CLAVIER – a equipa confiram se a carga proposta foi bem sucedida ou não
   2. Por um **processo de simulação**
      1. CHEF – As receitas culinárias propostas pelo SP são aplicadas a um simulador que tenta validá-las
2. Eventual **reparação** dessa solução
   1. Diagnóstico: o diagnostico proposto é modificado pelo utilizador que pode também comunicar um novo ou escolher um já existente, mas não associado ao contexto atual
   2. CLAVIER: se a carga proposta for indicado como “mal sucedida” pelo utilizador (aqui é impossível reparar “completamente” a solução porque não se sabe “como poderia ser bem sucedida”)

## Retain

1. Que conhecimento reter?
2. Quando o reter?
   1. A solução do novo caso é conhecida do sistema, mas alguns atributos têm valores consideravelmente diferentes de outros casos conhecidos
   2. A solução do novo caso é desconhecida do sistema
   3. O novo caso teve uma solução proposta pelo sistema, mas não resolveu o problema
   4. O novo caso tem uma solução que foi obtida por adaptação e funcionou corretamente
3. Quando não reter?
   1. O caso é igual a um que já figura na biblioteca de casos
4. Como o reter?

Há que ter em conta:

* O processo de representação dos casos
* O modo como a aprendizagem é realizada

### Princípios orientadores

Num sistema CBR a aprendizagem pode ser realizada:

* Por inserção de novos casos
* Por atualização (tipificação ou generalização) de casos passados
* Por atualização de fatores de relevância (os coeficientes *w* das fórmulas do nearest-neighboor)
* Por indução de arvores de decisão geradas por ID3 (ou algoritmo semelhante) quando na biblioteca de casos forem realizadas modificações

Um caso deve ser inserido

* Se o seu perfil (índices) for diferente
* Se a solução gerada no Reuse não tiver sido diretamente utilizada:
  + Por ter sofrido modificações manuais ou automáticas na fase de Revise
  + Por ter falhado tendo sido indicado uma solução “completamente nova”
* Um caso pode ser inserido como exemplo “positivo” (i.e. “ a este tipo de caso corresponde a solução X”) ou “negativo” (i.e. “para este tipo de caso a solução Y” não é valida. **Exemplo**: CLAVIER que regista cargas que falharam: estes exemplos são usados para invalidar soluções criadas manualmente)

## Biblioteca e Casos Iniciais (Seed Cases)

Para que um SP-CBR inicie a sua operação tem de dispor de casos iniciais (seed cases). Segundo Watson:

* Os casos devem ser representativos e bem distribuídos
  + **Representativos:** descritos por um conjunto de atributos importantes e de valores típicos
  + **Bem distribuídos:** o conjunto de casos deve cobrir uma boa percentagem das ocorrências que se preveem

Um sistema só é bem aceito pelos seus utilizadores quando consegue resolver, no início da operação, cerca de 80% dos problemas que lhe são postos.

O número de casos iniciais e o número de casos ao longo do tempo de vida varia muito.