

# Các hệ thống dựa trên mẫu cho trước

---

TS. NGUYỄN ĐÌNH HÓA

# Giới thiệu chung

---

Case-based reasoning (CBR)

Case: là mô tả chi tiết về một hoặc nhiều vấn đề cần giải quyết, kèm theo một mô tả chi tiết về giải pháp cho các vấn đề đó

Case: hàm chứa tri thức từ kinh nghiệm thực tiễn

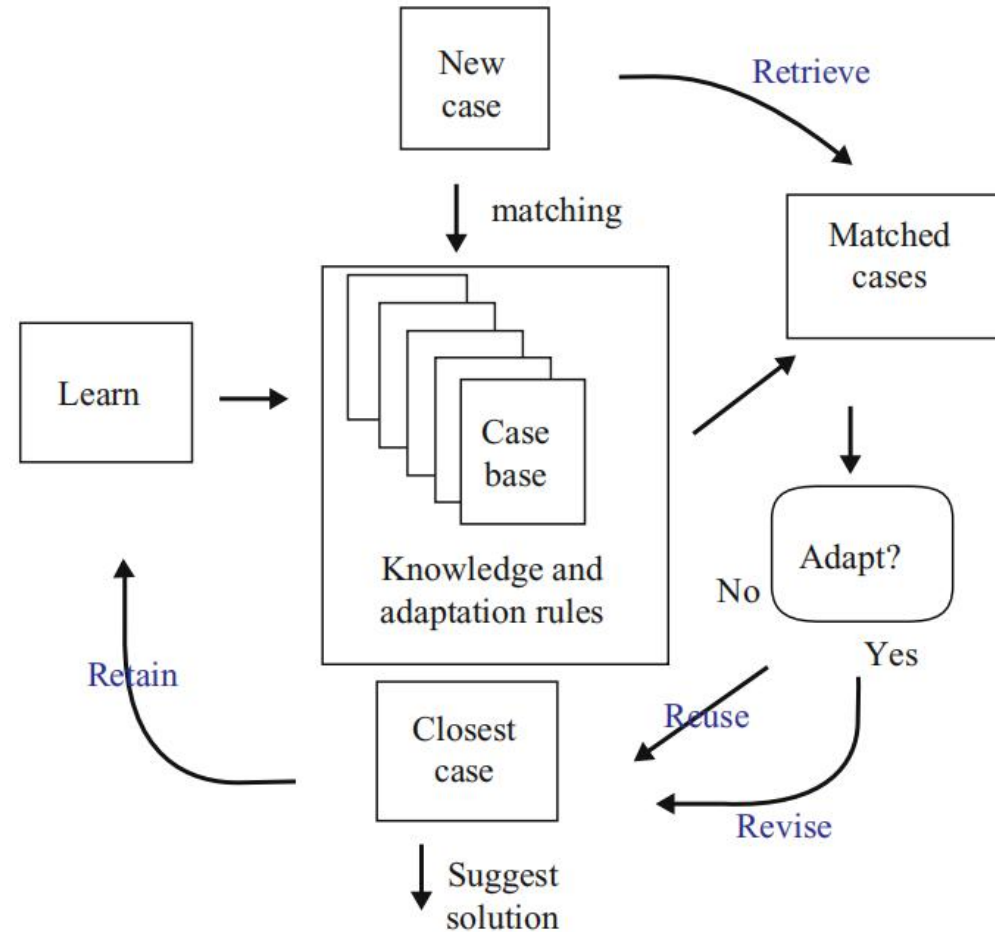
Trong hệ thống CBR: thông tin được chứa đựng trong các cases (mẫu) thay vì biểu diễn theo luật

Khi gặp một vấn đề mới: hệ thống tìm kiếm các mẫu đã có xem có mẫu nào trùng với vấn đề hiện tại. Nếu tìm thấy một mẫu trùng, hệ thống sẽ giải quyết vấn đề bằng giải pháp đã có.

CBR hoạt động giống con người xử lý vấn đề theo kinh nghiệm.

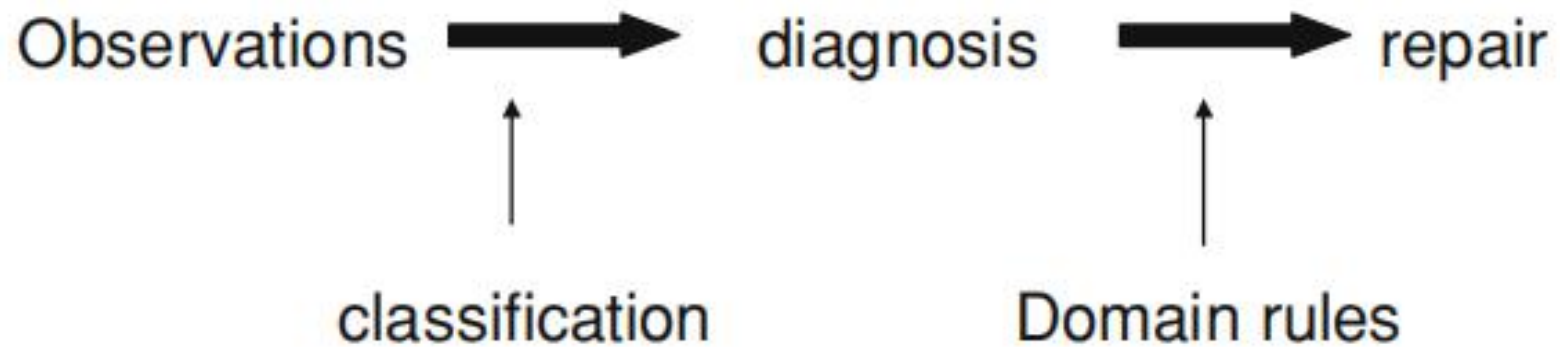
# Quy trình hệ thống

---



# Trình tự vận hành của CBR

---



# CBR – một vài ứng dụng

---

Help desks: lưu trữ các tình huống/kịch bản có sẵn để trả lời khách hàng

Tư vấn luật: lưu trữ các tình huống về luật pháp đã xảy ra trong quá khứ để tư vấn khách hàng

Tư vấn sức khỏe: lưu trữ các hiện tượng bệnh lý đã xảy ra trước đây để tư vấn bệnh nhân

Tư vấn hồng học thiết bị

...

# CBR – quá trình chẩn đoán

---

Chẩn đoán về mặt kỹ thuật:

- Quan sát hiện tượng, đo đạc thông số
- Tìm kiếm nguyên nhân hỏng hóc và đưa ra phương án sửa chữa

Chẩn đoán trong CBR:

- Mô tả hiện tượng
- Mô tả hỏng hóc và nguyên nhân hỏng hóc
- Mô tả phương án sửa chữa

Các thông tin cần lưu trữ trong CBR:

- Toàn bộ các trường hợp đã xảy ra trong quá khứ
- Tìm trường hợp nào giống nhất với vấn đề cần giải quyết, sử dụng nó để xử lý vấn đề đang gặp phải

# CBR – các bước thực hiện

---

Thu thập thông tin quan trọng mô tả về vấn đề mới xảy ra trong hệ thống

Tìm các trường hợp trong quá khứ giống nhất với thông tin vừa thu thập được

Lựa chọn trường hợp giống nhất, phù hợp nhất với vấn đề hiện tại

Tìm cách áp dụng giải pháp của trường hợp quá khứ đó để giải quyết vấn đề hiện tại

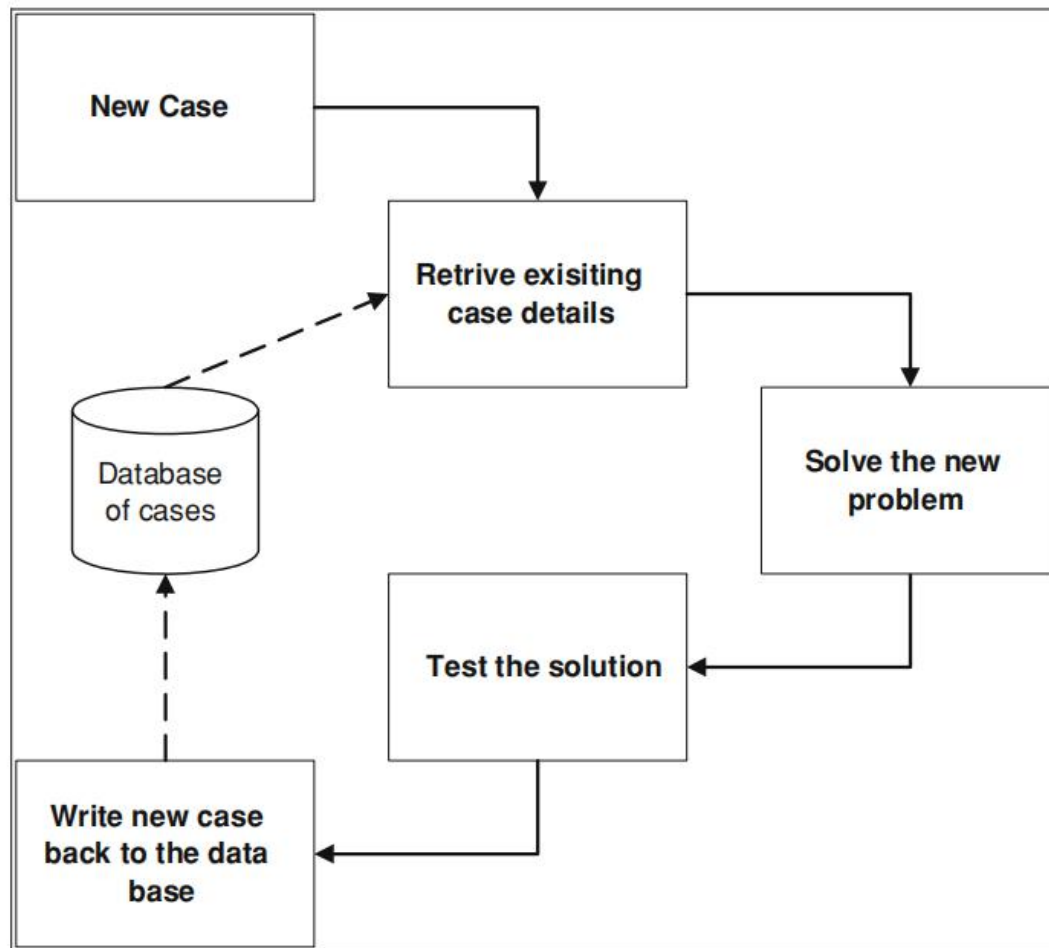
Đánh giá kết quả của giải pháp cho vấn đề hiện tại, lưu vấn đề mới này vào bộ dữ liệu

Nếu không tìm được trường hợp khả thi trong quá khứ, cần tìm giải pháp cho vấn đề hiện tại, rồi lưu vấn đề mới cùng giải pháp của nó vào trong CSDL

Nếu có nhiều trường hợp giống với vấn đề hiện tại:

- Phân tích để loại bỏ các thông tin còn mập mờ về vấn đề hiện tại
- Có thể tồn tại nhiều giải pháp cho vấn đề hiện tại

# CBR – các bước thực hiện





# CBR – ví dụ

---

Sửa lỗi máy in: giả sử trong CSDL lưu hai trường hợp đã xảy ra trong quá khứ

C A S E  1	<b>Problem (Symptoms)</b> <ul style="list-style-type: none"><li>• <i>Problem:</i> Document won't print</li><li>• <i>Printer type:</i> Laser</li><li>• <i>Ink or toner empty:</i> No</li><li>• <i>Paper empty:</i> No</li><li>• <i>State of lights:</i> On</li><li>• <i>Printer ready:</i> Yes</li></ul>
	<b>Solution</b> <ul style="list-style-type: none"><li>• <i>Diagnosis:</i> Paper Jam</li><li>• <i>Repair:</i> Open printer and remove blockage</li></ul>

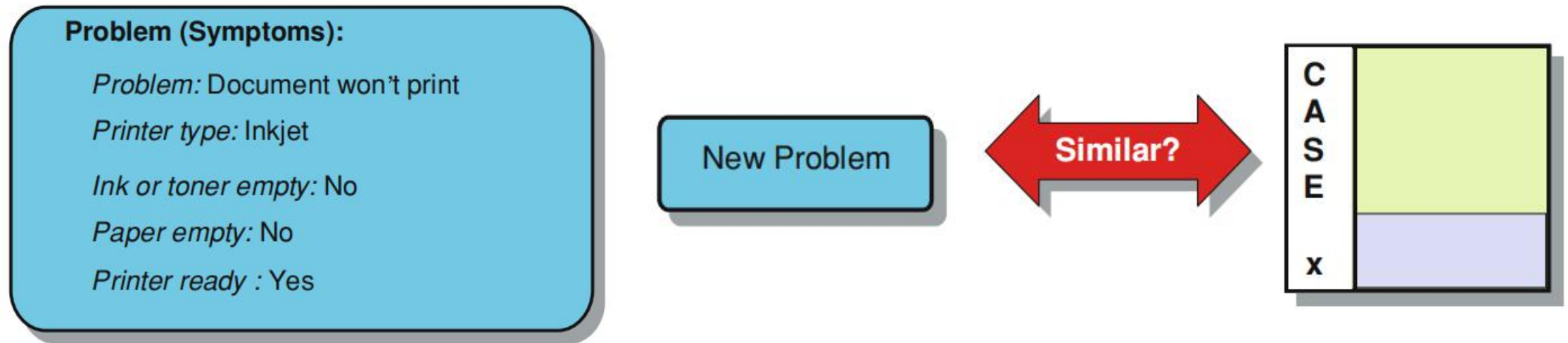
C A S E  2	<b>Problem (Symptoms)</b> <ul style="list-style-type: none"><li>• <i>Problem:</i> Document won't print</li><li>• <i>Printer type:</i> Inkjet</li><li>• <i>Ink or toner empty:</i> No</li><li>• <i>Paper empty:</i> Yes</li><li>• <i>State of lights:</i> On</li><li>• <i>Printer ready:</i> Yes</li></ul>
	<b>Solution</b> <ul style="list-style-type: none"><li>• <i>Diagnosis:</i> No paper</li><li>• <i>Repair:</i> Add paper to printer</li></ul>

# CBR – ví dụ

---

Khi có vấn đề mới về lỗi máy in: cần tìm giải pháp từ các trường hợp đã có trong CSDL

Các thông tin được thu thập về vấn đề hiện tại như sau:



# CBR – ví dụ

## Problem (Symptoms):

*Problem:* Document won't print

*Printer type:* Inkjet

*Ink or toner empty:* No

*Paper empty:* No

*Printer ready :* Yes

C A S E  1	<b>Problem (Symptoms)</b> <ul style="list-style-type: none"><li>• <i>Problem:</i> Document won't print</li><li>• <i>Printer type:</i> Laser</li><li>• <i>Ink or toner empty:</i> No</li><li>• <i>Paper empty:</i> No</li><li>• <i>State of lights:</i> On</li><li>• <i>Printer ready:</i> Yes</li></ul>
	<b>Solution</b> <ul style="list-style-type: none"><li>• <i>Diagnosis:</i> Paper Jam</li><li>• <i>Repair:</i> Open printer and remove blockage</li></ul>
C A S E  2	<b>Problem (Symptoms)</b> <ul style="list-style-type: none"><li>• <i>Problem:</i> Document won't print</li><li>• <i>Printer type:</i> Inkjet</li><li>• <i>Ink or toner empty:</i> No</li><li>• <i>Paper empty:</i> Yes</li><li>• <i>State of lights:</i> On</li><li>• <i>Printer ready:</i> Yes</li></ul>
	<b>Solution</b> <ul style="list-style-type: none"><li>• <i>Diagnosis:</i> No paper</li><li>• <i>Repair:</i> Add paper to printer</li></ul>

# CBR – ví dụ

---

So sánh vấn đề hiện tại với các trường hợp đã lưu trữ:

- Loại máy in: Inkjet -> Case 2
- Máy in sẵn sàng: Yes -> Case 1 và Case 2

Các thuộc tính khác nhau có ý nghĩa khác nhau trong quá trình so sánh, do vậy cần phải có trọng số so sánh trên các thuộc tính.

Ví dụ về tầm quan trọng của các thuộc tính:

- Các thuộc tính quan trọng: Tên bệnh? Có giấy không? Máy in có sẵn sàng không?,...
- Các thuộc tính không quan trọng: Loại máy in, thông số nguồn điện,....

# CBR – so sánh theo trọng số

---

Độ tương đồng được tính theo quy tắc như sau:

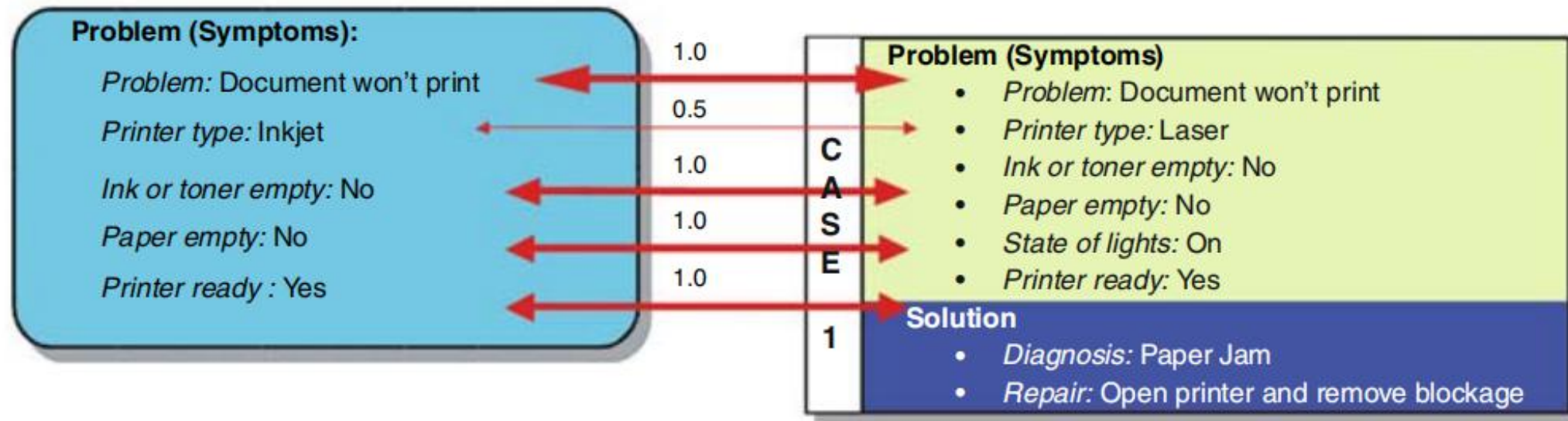
- Thuộc tính quan trọng nhất có trọng số là 6,
- Các thuộc tính không quan trọng có trọng số giảm dần
- Thuộc tính không quan trọng nhất có trọng số là 1

$$S(Problem, Case_i) = \frac{\sum_{j=1}^m w_j \cdot d(problem_j, case_{i,j})}{\sum_{j=1}^m w_j}$$

Trong đó,  $w_j$  là trọng số của thuộc tính thứ  $j$ ,  $m$  là số thuộc tính,  $d(problem_j, case_{i,j})$  là độ tương đồng giữa thuộc tính thứ  $j$  của Problem với thuộc tính thứ  $j$  của  $case_i$ ,  $d(.)$  lấy giá trị trong khoảng  $[0, 1]$ .

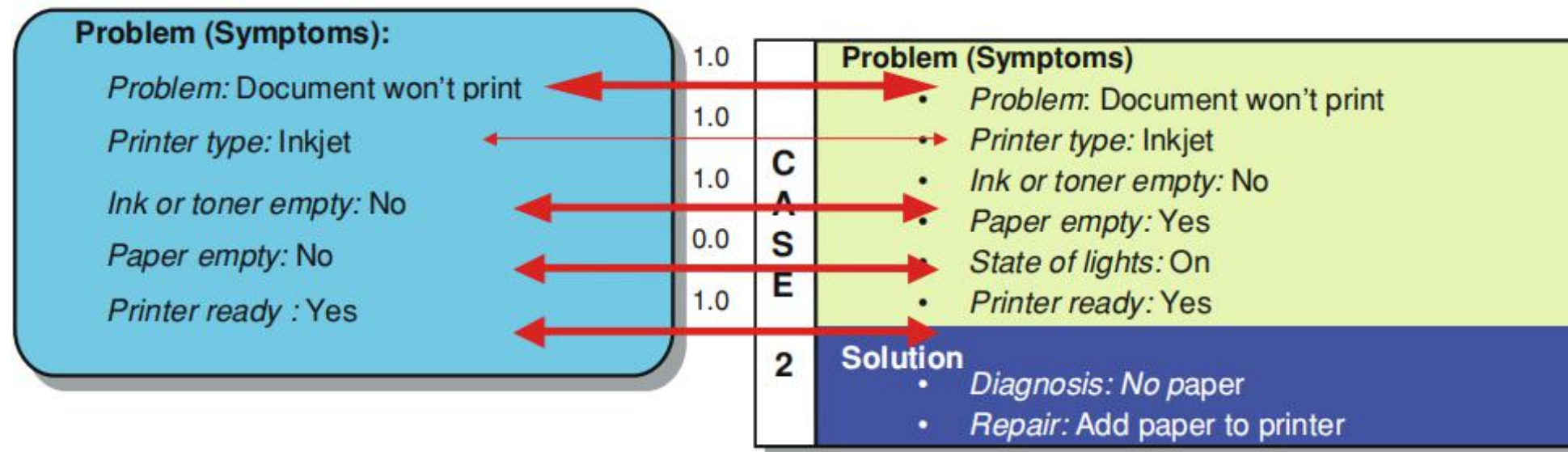
# CBR – so sánh theo trọng số

$$S(\text{Problem, Case 1}) = (6*1 + 1*0.5 + 6*1 + 6*1 + 6*1)/25 = \mathbf{0.98}$$



# CBR – so sánh theo trọng số

$$S(\text{Problem, Case 2}) = (6*1 + 1*1 + 6*1 + 6*0 + 6*1)/25 = \mathbf{0.76}$$





# CBR – ví dụ

---

Trong trường hợp chẩn đoán không ra kết quả như ý, vấn đề hồng học cần phải được xem xét và xử lý bởi kỹ thuật viên chuyên nghiệp

Sau khi xử lý xong vấn đề mới, trường hợp mới này lại được lưu trong CSDL để tham khảo trong các lần tiếp theo



# CBR – ví dụ

---

Giả sử sau khi kiểm tra kỹ máy in, kỹ thuật viên phát hiện trạng thái “No power”, và giải pháp được đưa ra là “Check electrical supply”.

Hiện tượng xảy ra trên máy in được bổ sung thêm là “Status light: Off”.

Case mới được thêm vào hệ thống sẽ là

C A S E  3	<b>Problem (Symptoms)</b> <ul style="list-style-type: none"><li>• <i>Problem:</i> Printer does not do anything</li><li>• <i>Printer type:</i> Inkjet</li><li>• <i>Ink or toner empty:</i> No</li><li>• <i>Paper empty:</i> No</li><li>• <i>State of lights:</i> Off</li><li>• <i>Printer ready:</i> Unknown</li></ul>
	<b>Solution</b> <ul style="list-style-type: none"><li>• <i>Diagnosis:</i> No power</li><li>• <i>Repair:</i> Check electrical supply</li></ul>

# CBR – ưu và nhược điểm

---

## Ưu điểm:

- Giảm lượng kiến thức cần phải có khi xử lý vấn đề mới xảy ra. Hệ thống chỉ đơn thuần tìm kiếm các trường hợp giống với vấn đề hiện tại, không cần triển khai các luật suy diễn.
- Hệ thống tự cập nhật các trường hợp mới, không cần điều chỉnh tri thức bên trong hệ thống.

## Nhược điểm:

- Số mẫu quá ít, hoặc không đa dạng sẽ ảnh hưởng đến hiệu quả giải quyết vấn đề
- Quá trình tìm kiếm các trường hợp cũ cần được thực hiện một cách cẩn thận
- Việc xác định trọng số cho các thuộc tính đóng vai trò quan trọng
- Khó diễn đạt được tri thức cho người dùng.

## CBR nên được dùng khi nào?

- Các vấn đề gặp phải khó được mổ xẻ phân tích, khó áp dụng tri thức dựa trên luật để giải quyết.
- Các nguyên lý xử lý vấn đề đó khó có thể diễn đạt bằng tri thức cụ thể, chỉ có thể dựa trên kinh nghiệm của các lần gặp trước đó. Không phát triển được thành các luật của tri thức.

# So sánh giữa RBR và CBR

Yếu tố so sánh	RBR	CBR
Nội dung tri thức: có thể diễn đạt bằng các đơn vị tri thức nhỏ không?	Tri thức được biểu diễn bằng các luật thường nhỏ, độc lập, rõ ràng	Các Case biểu diễn khối tri thức lớn, mối liên hệ giữa vấn đề cần giải quyết và giải pháp thường không rõ ràng
Lĩnh vực tri thức: dễ hiểu và dễ lưu trữ lại không?	Dễ hiểu. Các luật, hoặc heuristic, thường được nhận thức và lưu lại được	Lĩnh vực của tri thức thường khó hiểu, tuy nhiên vẫn có thể tìm kiếm các giải pháp cho vấn đề cần giải quyết.
Tri thức có cần khớp với vấn đề cần giải quyết không?	Để giải quyết vấn đề, các luật được sử dụng phải khớp đúng với tình huống thực tế của vấn đề	Các Cases vẫn có thể được sử dụng kể cả khi chỉ khớp một phần với vấn đề cần giải quyết

# CBR – Ví dụ về dự đoán giá nhà

---

## Bài toán:

Một người muốn bán căn hộ trong một thành phố. Anh ta nên xác định giá nhà thế nào?

1. Bắt đầu bằng giá tiền mà anh ta đã mua trước đây, cộng với sự mất giá đồng tiền trong các năm vừa qua.
2. Dự đoán giá nhà trên thị trường để xác định giá tiền cho căn hộ mà anh ta đang sống.

# CBR – dự đoán giá nhà

---

Dựa trên quy tắc chung: trong khu vực mà anh ta sinh sống, giá nhà theo mặt bằng chung là bao nhiêu, ví dụ 30 tr / m<sup>2</sup>

Dựa trên các trường hợp cụ thể:

- Căn hộ bên cạnh cùng diện tích, nhưng khác cách bố trí với căn nhà anh ta đang ở, vừa được bán với giá 2 tỷ.
- Căn hộ tầng dưới nhỏ hơn một chút so với căn anh đang ở, nhưng nhiều cửa sổ hơn, vừa được bán với giá 2.2 tỷ
- Căn hộ tầng trên rộng hơn căn anh đang ở, và có nhiều phòng vệ sinh hơn, được bán với giá 3 tỷ
- Căn hộ của tòa nhà bên cạnh, cùng tầng với căn của anh, và cùng diện tích, được bán với giá 1.7 tỷ.
- ...

# CBR – dự đoán giá nhà

Một số yếu tố cần xem xét khi xác định giá nhà:

<b>Yếu tố</b>	<b>Căn 1</b>	<b>Căn 2</b>	<b>Căn 3</b>	<b>Căn 4</b>	<b>Căn cần bán</b>
Diện tích (m <sup>2</sup> )	70	85	90	110	95
Số phòng ngủ	2	3	3	3	3
Số phòng vệ sinh	1	2	2	3	2
Số cửa sổ	2	3	3	4	4
Số mặt thoáng	1	1	2	2	2
Diện tích bếp (m <sup>2</sup> )	10	12	15	20	16
Diện tích phòng khách (m <sup>2</sup> )	15	15	15	20	17
Hướng cửa chính	Đông	Tây	Tây - Nam	Bắc	Đông
Số tầng	10	15	12	11	11
<b>Giá bán (Tỷ đồng)</b>	<b>2.1</b>	<b>2.2</b>	<b>2.5</b>	<b>3.1</b>	<b>???</b>

# CBR – dự đoán giá nhà

---

Sử dụng công thức tính khoảng cách theo trọng số

Các trường hợp đã xảy ra:  $n$  trường hợp, mỗi căn hộ có  $m$  thông tin cần so sánh (tổng diện tích, số phòng ngủ, số cửa sổ, số nhà vệ sinh, hướng cửa sổ, số tầng,...)

$$X_i = (X_i^1, \dots, X_i^m)$$

Giá nhà của mỗi căn hộ đã được bán được ước tính dựa trên công thức trọng số như sau:

$$Y_i = \beta_0 + \beta_1 X_i^1 + \dots + \beta_m X_i^m + \varepsilon_i$$

Từ đó, mỗi căn hộ đã bán sẽ được biểu diễn bởi một bộ vector:

$$(X_i^1, \dots, X_i^m, Y_i) \text{ for } i = 1, \dots, n$$

Căn nhà mà anh muốn bán có thông tin như sau:  $X_{n+1} = (X_{n+1}^1, \dots, X_{n+1}^m)$

# CBR – dự đoán giá nhà

---

Công thức để tính giá nhà có thể được áp dụng dựa trên độ tương đồng như sau:

$$\hat{Y}_{n+1} = \frac{\sum_{i \leq n} S(X_i, X_{n+1}) \cdot Y_i}{\sum_{i \leq n} S(X_i, X_{n+1})} + \varepsilon_i$$

Trong đó  $\varepsilon_i \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$