Case-Based Reasoning
Watson chapters 1 - 4

Bill Cheetham, Kai Goebel

Slides modified from Dr. Ian Watson

Contents
1. What is Case-Based Reasoning?
2. How does CBR work?
3. Advantages / Disadvantages
4. Who Uses CBR?
   1. case study - Lockheed
   2. lists at web sites
   3. GE
What is CBR?

1. A case-based reasoner solves new problems by using or adapting solutions that were used to solve old problems
2. offers a reasoning paradigm that is similar to the way many people routinely solve problems

What Is CBR?

1. How will you get home?
2. Generate a path
3. Or remember the way
4. Path to new location?
5. Remember close path
6. Adapt it
What Is CBR?

What is 12 x 12?
144
What is 12 x 13?
12 x 12 + 12
156

In order to have a phone in every house 1/10 of the entire US population would need to be phone operators

Create a phone book with people and their phone numbers
Who uses CBR?

Lawyers
1. find previous ruling that applies to case
2. show that it applies to current case

Real Estate Appraiser
1. find similar comparable houses
2. estimate value of target based on value of comparable

Quotes
1. "We know nothing of what will happen in future, but by the analogy of experience." -- Abraham Lincoln
2. "Study the past if you would divine the future." -- Confucius
3. "If at first you don't succeed, you are running about average." -- Bill Cosby
The Case-Based Cycle

1. Problem

2. RETRIEVE

3. Similar Cases

PRIOR CASES

CASE-BASE
The Case-Based Cycle

1. **Problem**
2. **Retrieve**
   - **Similar Cases**
3. **Revise**
   - **Solution**
History of CBR

CBR can trace its roots to the field of psychology and theories about how human memory works

1. “Episodic Memory” [Tulving 1972] provides a method for storing and recalling large chunks of related information such as events, scenes, occurrences, and stories

2. “schema” [Rumelhart 1977] reasoning is the process of applying chunks of information to new situations
History of CBR- scripts

Roger Schank’s group invented scripts at Yale natural language lab during the mid-seventies

“scripts represent generalizations about actions that should take place in stereotypical situations”

Restaurant sequence of events:
1. enter restaurant,
2. be seated by hostess,
3. obtain menu from waiter,
4. order drinks,
5. waiter leaves,
6. waiter returns with drinks,
7. order food,
8. wait for waiter,
9. waiter returns with food,
10. eat food,
11. waiter returns with bill,
12. pay bill & leave

History of CBR - MOPs

Schank’s lab began research based on the notion that for remembering and reasoning tasks both general knowledge structures, like scripts, and specific instances are crucial to understanding.

Memory Organization Packets (MOPs) integrate general knowledge with experiences.
History of CBR - TOPs

Thematic Organization Packets (TOPs) categorize situations by the plans of the participants rather than the details of the situation.

Schank’s daughter was diving in the ocean looking for sand dollars. He pointed out where a group of them were, yet she continued to dive elsewhere. He asked why, and she told him that the water was shallower where she was diving. This reminded him of the old joke about a drunk searching for his lost ring under the lamppost where the light was better.

History of CBR - Software

1. The 80’s
   1. the original CBR programs CASEY, CHEF, JULIA were written in LISP
   1. research tools in public domain

2. The 90’s
   1. the development of commercial CBR tools, mostly in C

3. Today
   1. commercial Web-based Java tools
   1. specific application (customer self-service)
What is a Case?

1. several features describing a problem
2. plus an outcome or a solution
3. cases can be very rich
   1. text, numbers, symbols, plans, multimedia
4. cases are not usually distilled knowledge
5. cases are records of real events
6. and are excellent for justifying decisions

2 types of case features

- **Patient Ref #: 1024**
- **Patient Name: John Doe**
- **Address: 12 Elm Street**
- **Next of Kin: Jane Doe**
- **Photo:**
- **Age:** 53
- **Sex:** Male
- **Weight:** 225 lbs
- **Height:** 5’11”
- **Blood Type:** A neg.

**unindexed features**
Not predictive & not used for retrieval, they provide background information to users

**indexed features**
Predictive and used for retrieval
What is a Case-Base?

1. A case-base is a set of cases.
2. Case-bases are usually just flat files or relational databases.

A robust case-base, containing a representative and well distributed set of cases, is the foundation for a good CBR system. [Kriegasman & Barletta, 1993].

How Does Retrieval Work?

1. Imagine a decision with two factors that influence it.
2. Should you grant a person a loan?
   - Net monthly income
   - Monthly loan repayment
How Does CBR Work?

- these factors can be used as axes for a graph

How Does CBR Work?

- a previous loan can be plotted against these axes
How Does CBR Work?

1. and more good loans

How Does CBR Work?

1. plus some bad loans
Lazy Learning

1. past cases (loans) may tend to form clusters, but you don’t need to find them.

How Does CBR Work?

1. a new loan prospect can be plotted on the graph.
How Does CBR Work?

1. and the distance to its nearest neighbors calculated

How Does CBR Work?

1. the best matching past case is the closest

1. this suggests a precedent

1. the loan should be successful
How Does CBR Work?

1. over time the prediction can be validated

...net monthly income

It was a good loan

How Does CBR Work?

1. the system is learning to differentiate good and bad loans better
How Does CBR Work?

1. As more cases are acquired its performance improves.

Retrieval Issues

1. Do all indexed features have the same weight?
2. Is the similarity linearly proportional to the distance a case is from the new problem?
3. What distance measure should be used (city block, line of sight, …)
4. Uniformity of solution space
How to weight the features

1. First normalize the attributes
   1. find min. and max.
   1. set min = 0, max = 1

1. Real estate appraiser example
   1. living-area between 2,000 and 3,500 sq. ft.
   1. 2,000 = 0
   1. 3,500 = 1
How to weight the features

1. Ask an expert
2. Look for trends in data (plot, regression)
3. Use leave-one-out testing
   1. Select an item from the case base where you know the solution
   2. Run your CBR system on the case
   3. Determine error = difference in solution suggested by CBR and actual solution
   4. Update weights to minimize this error (GA?)

Adaptation

In many situations the case returned is not the exact solution needed
1. Many techniques can be used
   1. Rule-based system
   2. Regression
   3. Heuristic search
   4. Other
1. The best technique depends on the application
Adaptation

Problem Space

the adapted solution can be created by many methods, such as, interpolation or extrapolation

Solution Space

adapted solution

input problem description

the adapted solution can be created by many methods, such as, interpolation or extrapolation

Rule Based Adaptation

Example from real estate appraiser

1 retrieved a house selling at $100,000 that is exactly like the one being appraised except it has a 2 car garage and target has 1 car garage
1 the value of an extra garage is $4,000
1 adjusted value is $100,000 + $4,000
How do you learn Adaptations

Analyze the Data
1. Look for trends in data (plot, regression)
2. Compare similar items
   1. Find two data items that are identical except for one difference
   2. The solution difference is the value of the change in the attribute
   3. Evaluate rules by testing different ones

Exercise
1. Estimate height using students as the case-base
2. 5 students come to front and be the case base
3. Predict the height of a 6th student
Summary

1. In real life the problem space is \( N \) dimensional
2. New features can be added if they become relevant
3. Feature vectors can be weighted to reflect their relative importance
4. Tolerant of noise & missing data
5. \( k \)-Nearest Neighbor Retrieval

Advantages of CBR

1. CBR is intuitive - it’s how we think
2. No knowledge elicitation to create rules or methods
3. This makes development easier
4. Systems learn by acquiring new cases through use
5. This makes maintenance easy
6. Justification through precedent
When to Apply CBR?

1. when a domain model is difficult or impossible to elicit
2. when the system will require constant maintenance
3. when records of previously successful solutions exist

CBR with Confidence

Confidence based on:
1. Number of cases matching
2. Similarity of matching cases to new problem
3. Similarity of matching cases to each other
CBR with Confidence

Blue dots - data points
Red line - actual line to be predicted
Yellow line - prediction by a neural network

CBR Systems Learn

1. decision making is dynamic
1. CBR systems learn by acquiring new cases
   1. no addition of new rules
   1. no retraining of neural networks
   1. no re-evolving new populations with new genomes
   1. no re-induction of rules from data
Case-Base Issues

1. How many cases are needed
2. How to remove overlapping cases
3. How to efficiently search
   1. create abstractions from cases
   2. multiple case bases
4. What features to use for indexing
5. How to weight the features

Disadvantages of CBR

1. Can take large storage space for all the cases
2. Can take large processing time to find similar cases in case-base
3. Cases may need to be created by hand
4. Adaptation may be difficult
5. Needs case-base, case selection algorithm, and possibly case-adaptation algorithm
Disadvantages of CBR

1. if you require the *best* solution or the *optimum* solution - CBR may not be for you
2. CBR systems generally give *good* or *reasonable* solutions
3. this is because the retrieved case often requires adaptation

CBR vs regression

1. regression summarizes data while CBR retains all data points
2. regression is good for simple relationships
3. hard for regression to learn strange shapes
CBR vs Rule-Based Systems

1. CBR offers a cost-effective solution to the ‘knowledge acquisition bottleneck’ problem
2. CBR systems can learn from experience and so can be self-maintaining
3. Rule-based systems are better when it is hard to gather case data

CBR vs Rule Based System

1. rule-based systems justify decisions by showing a rule trace
2. decision grant loan because
   rule 24 -> rule 61 -> rule 43 -> rule 202
3. rule traces are opaque & can be confusing to users
CBR vs NN

1. Neural nets cannot justify their decisions
2. Users have to trust the computer is always correct
3. Neural nets cannot take advantage of domain knowledge

Applications of CBR

- **Classification:** “The patient’s ear problems are like this prototypical case of otitis media”
- **Compiling solutions:** “Patient N’s heart symptoms can be explained in the same way as previous patient D’s”
- **Assessing values:** My house is like the one that sold down the street for $250,000 but has a better view”
- **Justifying with precedents:** “This Missouri case should be decided just like Roe v. Wade where the court held that a state’s limitations on abortion are illegal”
- **Evaluating options:** “If we attack Cuban/Russian missile installations, it would be just like Pearl Harbor”
Lockheed - CLAVER

1. Lockheed makes aircraft parts from composite materials.
2. These materials are made from layers of carbon fibers that are formed into a single component by curing in a large oven, called an autoclave.

Lockheed

1. PROBLEM - how to optimize the loading of an autoclave for curing composite materials.
2. Different materials need different heating & cooling procedures.
3. Materials interact with each other in the autoclave.
4. Mistakes are VERY costly.
Lockheed

1. 2 experienced operators relied on plans of previously successful layouts
2. New layouts were adapted from old
3. If successful they were added to a library
4. They wanted to develop a decision support tool to assist experts and to retain expertise as a corporate asset

Lockheed

1. Lockheed had NO model of the autoclave (a rule-based tool failed)
2. The manufacturers could not provide one
3. Layouts did not repeat
4. Materials were constantly changing
5. Designs constantly change
6. Elements interact
7. CBR was used
Soft Computing: Case-Based Reasoning

**Lockheed**

- Parts list retrieval
- Most similar case adaptation
- Known layouts
- Adapted case learning

Their system was implemented in 1990.

CLAVIER started with 20 successful layouts.

CLAVIER now has hundreds of successful layouts.

It retrieves a successful layout or adapts one 90% of the time.

It acts as a corporate memory.
CBR Web sites

**www.ai-cbr.org**
- CBR research, places, people, papers, conferences summaries

**www.cbr-web.org**
- German perspective on CBR
Soft Computing: Case-Based Reasoning

Applying CBR

This section provides information about existing Case-Based Reasoning systems, applications, and tools.

Projects & Applications
Research projects and practical applications using CBR technology

Tools
CaseBases, tools, and programming libraries

Related:
Professional CBR consulting

Soft Computing: Case-Based Reasoning

HomeValueCheck

Tell Me More

FURTHER DETAILS ON OUR VALUATION PROCESS

HomeValueCheck Analysis

- Sales Information
- Subject Information
- Market Comparable
- Comps
- Sales
- Adjustments
- Regression
- Average
- Value

HomeValueCheck processes up-to-the-minute information from national and regional data banks to offer the most recent sales prices of comparable homes in your neighborhood. Reliable predictions of value such as appraised figures, sales of comparable homes, and various other features are prime considerations in the above model. This data is then analyzed using sophisticated independent valuation techniques to provide a price estimate value for your home.

IMPEKRT COMPARABLE ANALYSIS
1. The physical characteristics of your property are compared to the characteristics of each sales comparable property.
2. Selling prices of the sales comparable properties are adjusted to reflect differences in characteristics between the sample property and your property.

For example, if your home has three bedrooms and a sales sample property...