Remembering To Forget

A Competence-Preserving Case Deletion Policy for Case-Based Reasoning Systems

Barry Smyth Mark T. Keane

Introduction

- The utility problem---certain "harmful" knowledge may actually degrade system performance(problem solving efficiency or time)

- The swamping problem---the expense of searching large case-bases for appropriate cases with which to solve the current problem.

Introduction

How to solve the swamping problem?

- Efficient parallel retrieval algorithms

Delaying performance degradation, but they do not eliminate it altogether.

- Information filters applied at different stages in the problem solving process(Markovitch and Scott)
- Selective utilization && Selective retention

Traditional Deletion Policies

- Random Deletion---a random item is removed from the knowledge base once the knowledge-base size exceeds some predefined limit.
- Minton's utility metric

Utility=(ApplicationFreq*AverageSavings)-MatchCost

Traditional Deletion Policy

- **Problems**

- One important difference between pure CBR and Speed-up learning systems.

- The deletion of critical cases can significantly reduce the competence of a CBR system.

Case Competence Categories

- Coverage and Reachability

- Pivotal Cases
- Auxiliary Cases
- Spanning Cases
- Support Case
- Support Group

Definition

Definition 1: Coverage

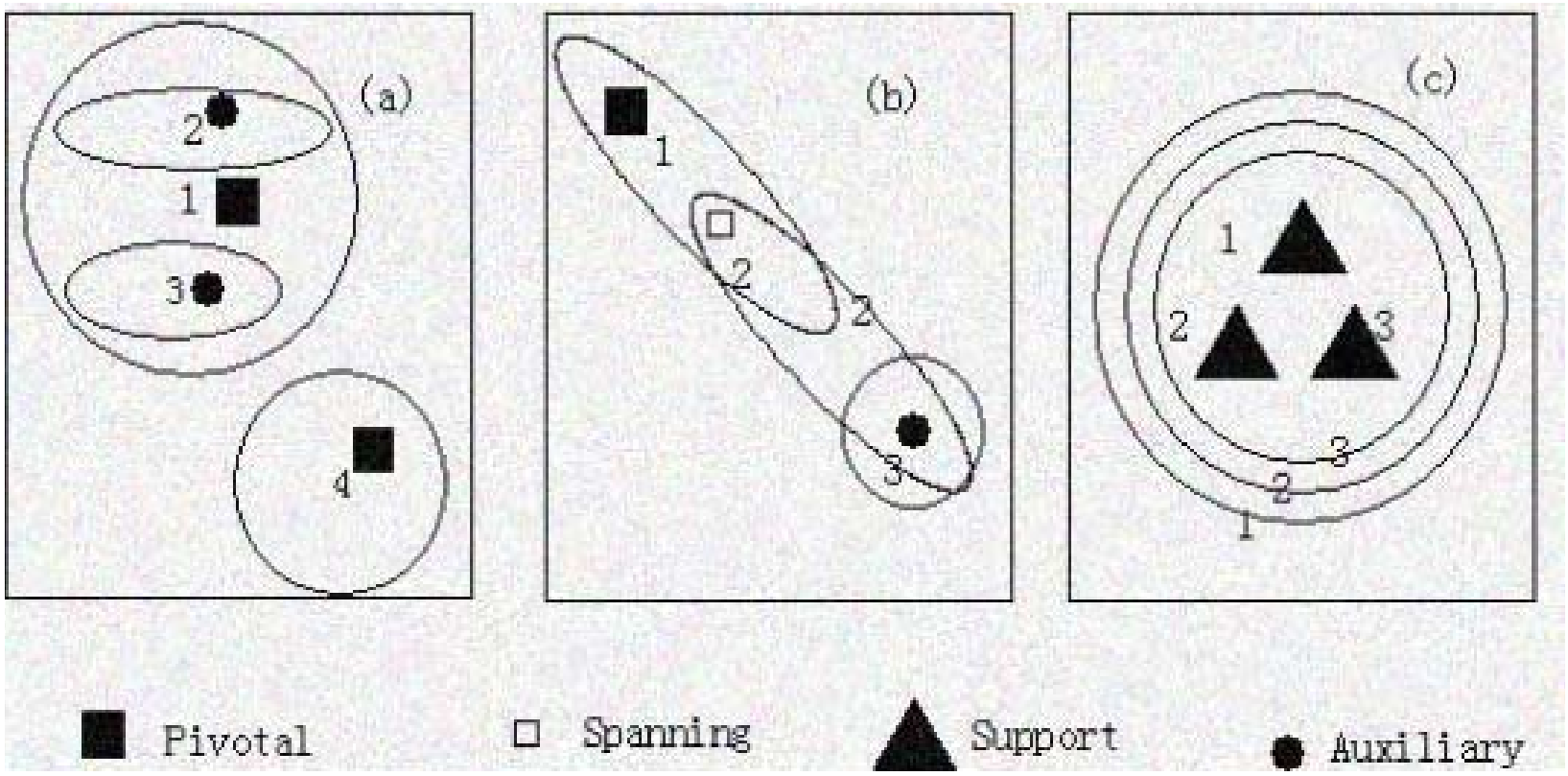
Given a case – base $C = \{c_1,...,c_n\}$, For $c \in C$,

Coverage(c) = $\{c' \in C: \text{Adaptable}(c, c')\}$

Definition2: Reachability

Given a case – base $C = \{c_1,...,c_n\}$, For $c \in C$,

Reachable(c) = $\{c' \in C: \text{Adaptable}(c', c)\}$


(a)
2
1
3
4
(b)
1
2
2
3
(c)
1
2
3
3
2
1
Pivotal
Spanning
Support
Auxiliary

Definition

- Definition 3: Pivotal Cases

Pivot(c) iff Reachable(c)-{c}= Ø

- Definition 4: Auxiliary Case

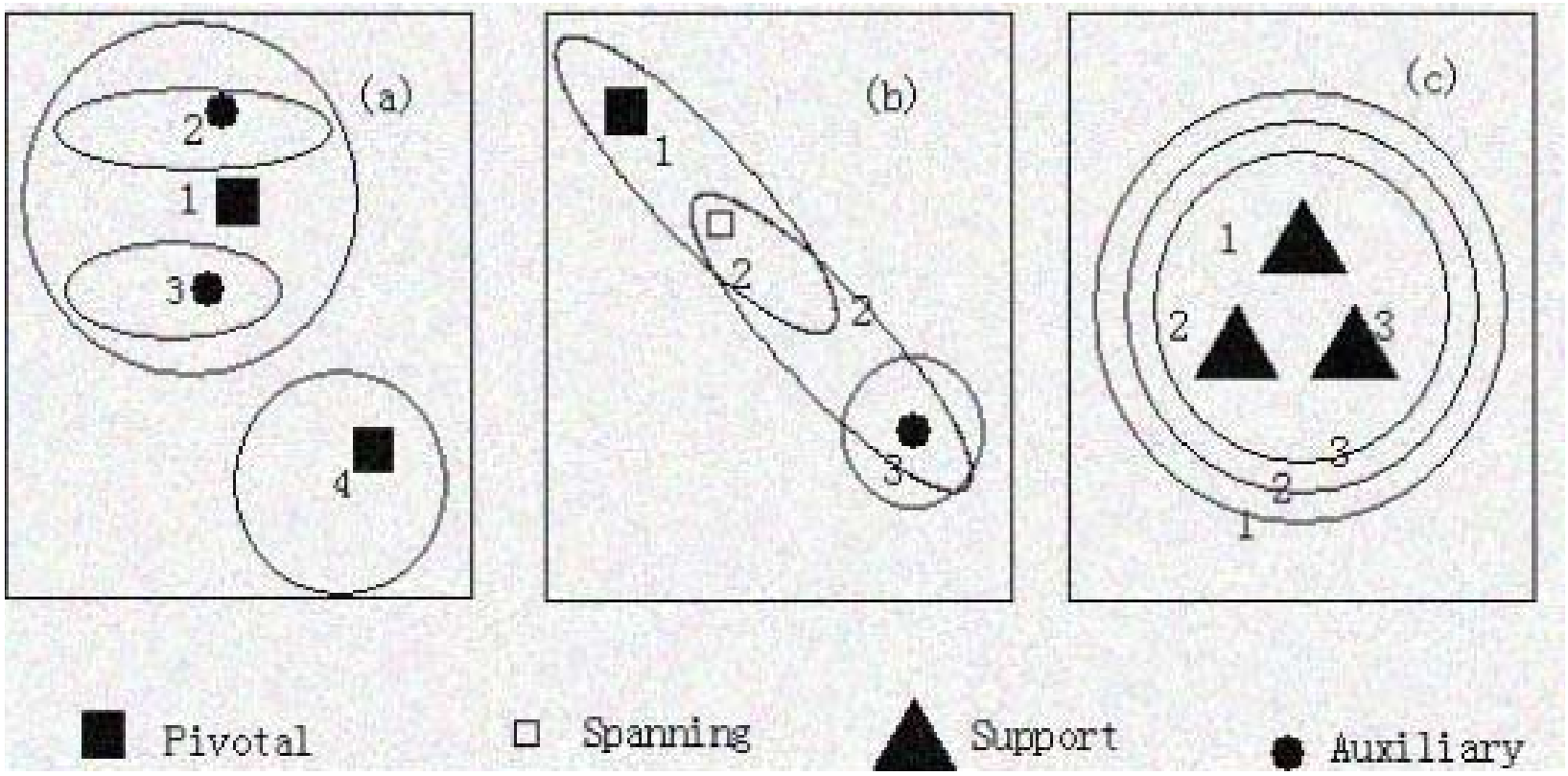
Auxiliary(c) iff ∃c' ∈ Reachable(c)-{c}:

Coverage(c) ⊆Coverage(c')

- Definition 5: Spanning Case

Spanning(c) iff ¬Pivotal(c)∧Coverage(c)

$\cap\ U_{c' \in Reachable(c)-\{c\}} Coverage(c) \neq \emptyset$


(a)
2
1
3
4
(b)
1
2
2
3
(c)
1
2
3
3
2
1
Pivotal
Spanning
Support
Auxiliary

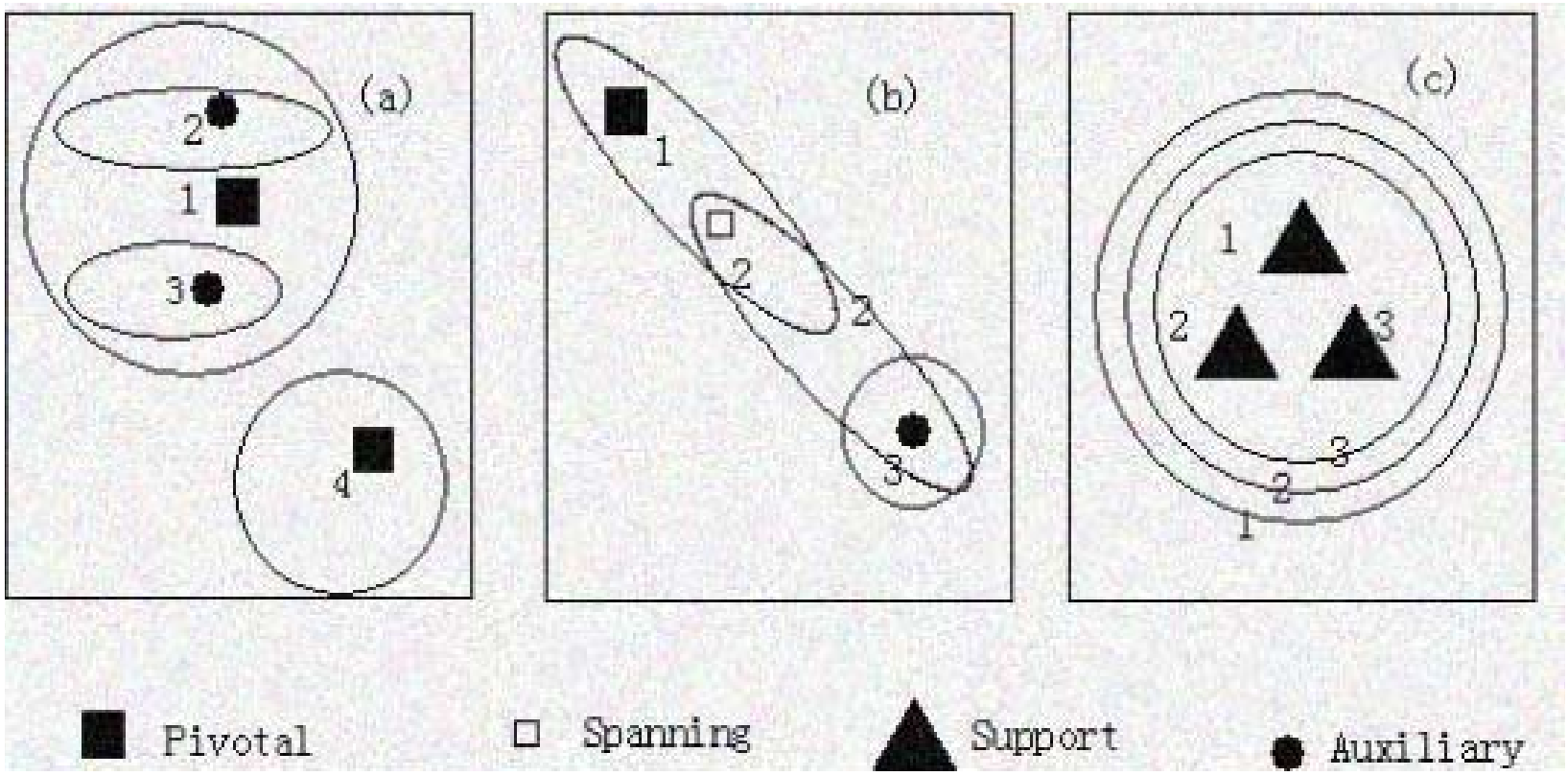
Definition

- Definition 6: Support Case

Support(c) iff $\exists c' \in Reachable(c)-\{c\}$;
$Coverage(c') \subseteq Coverage(c)$

- Definition 7: Support Group

SupportGroup(C') iff $\forall\ c_i,\ c_j \in C'$;
$Coverage(c_i)=Coverage(c_j)$


(a)
2
1
3
4
(b)
1
2
2
3
(c)
1
2
3
3
2
1
Pivotal
Spanning
Support
Auxiliary

Learing Update

```
Learning Update( target , base):
If Pivotal(base) then
   Remove the base from the set of pivotal cases
   Add the base and target as a new support group
   and mark this group as pivotal in origin
Elself Support(base) then
   Add the target to the base's support group
Elself Spanning(base) then
   Make new support group from the base & target
      and mark this group as spanning in origin
Elself Auxiliary(base) then
   Add the target to the set of auxiliary cases
Endlf
```

Deletion Update

```
If Pivotal(case) then
   Remove the case from the set of pivotal cases
ElseIf Spanning(case) then
   Remove the case from the set of spanning cases
ElseIf Support(case) then
   Remove the case from its support group
   If the resulting group is a singleton then
      If the group is pivotal in origin then
         Add this remaining case to the pivots
      ElseIf add this case to the spanning set
         Remove singleton from the support groups
      EndIf
ElseIf  Auxiliary(case) then
   Remove the case from the set of auxiliary cases
EndIf
```

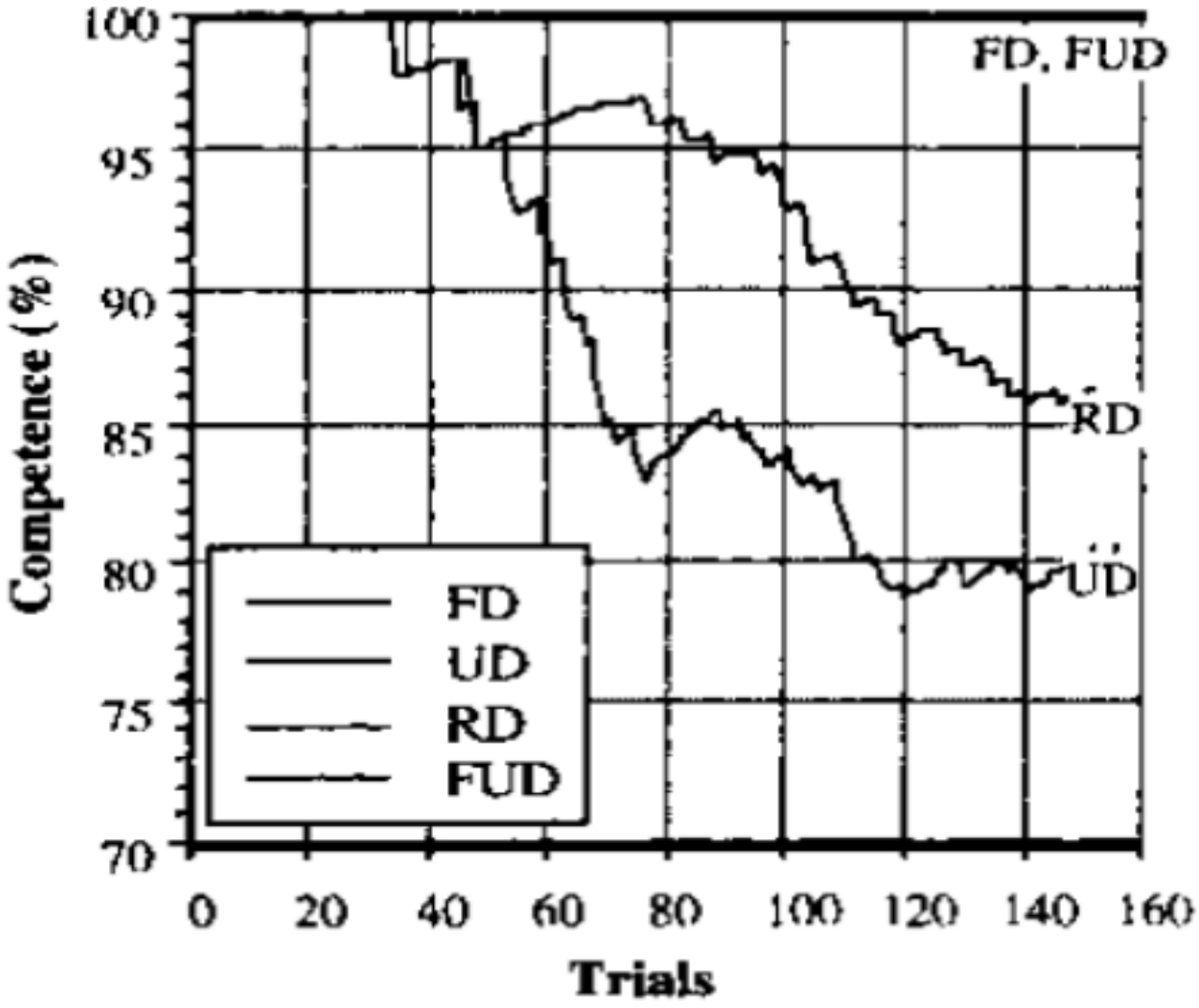
The Footprint Deletion Policy

Delete Case(Cases):
If there are auxiliary cases then
  SelectAuxiliary(AuxiliaryCases)
ElseIf there are Support cases then
  With the largest support group
    SelectSupport(SupportGroup)
ElseIf there are spanning cases then
  SelectSpanning(SpanningCases)
ElseIf there are pivotal cases then
  SelectPivot(PivotalCases)
EndIf

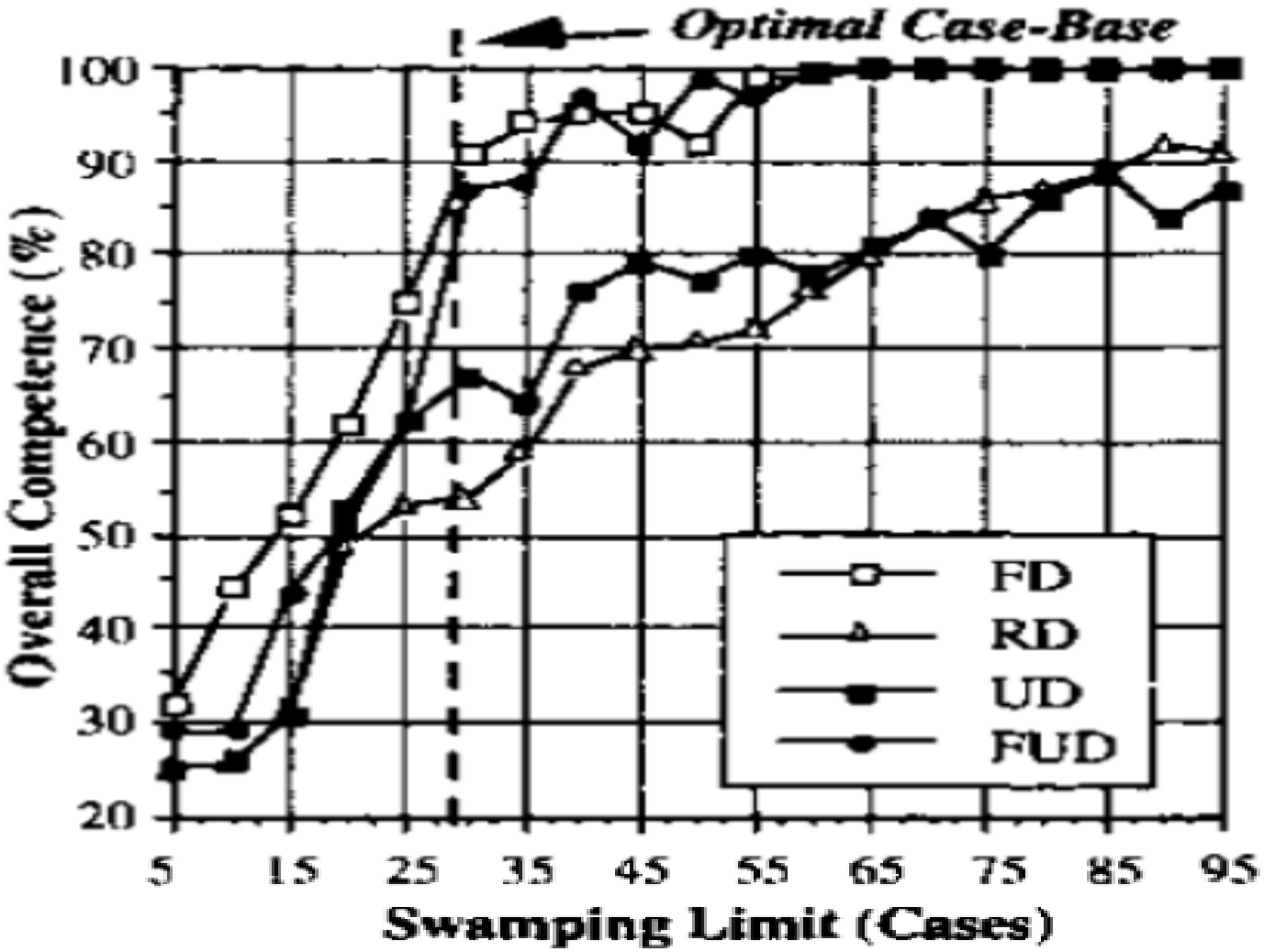
Footprint-utility Deletion

- Combining footprint deletion and utility deletion .
- First, the footprint method is used to select candidates for deletion.
- Second, select the candidate with the lowest utility.

Experiment 1



Experiment 2



Case-Base statistics

Pivotal cases	Auxiliary cases	Support cases	Spanning cases
20	5	25	0