Maintenance of case bases: current algorithms after fifty years

J. M. Juarez¹, S. Craw², J. Ricardo Lopez-Delgado¹, M. Campos¹

¹ University of Murcia
² Robert Gordon University
50 years
The Condensed Nearest Neighbor Rule

The purpose of this rule is to utilize the condensed nearest-neighbor decision rule (CNN rule) and to present some case studies to show the potential of the CNN rule. The CNN rule was developed by Peter E. Hart, and it is based on the nearest-neighbor (NN) rule. The NN rule states that an unlabeled instance in a training set is classified into the class of its nearest labeled instance. The CNN rule is a modification of the NN rule, where a subset of the training instances is selected to form a condensed set. The condensed set is then used to classify new instances. The CNN rule is known for its simplicity, efficiency, and effectiveness in various classification tasks. However, the CNN rule has some limitations, such as the need for a good selection of the condensed set and the sensitivity to the choice of the condensed set. Therefore, this paper presents some case studies to show the potential of the CNN rule and to compare its performance with other classification algorithms.

**CNN Algorithm**

**Input:** original case-base C

**Output:** maintained case-base CM

1. CM ← empty
2. c ← first case of C
3. C ← C – {c}
4. CM ← CM U {c}

Repeat

For all c in C

1. C ← C – {c}
2. **if not** CorrecClassifyNN(c, CM)
   1. CM ← CM U {c}
3. **endif**

enfor

Until C without changes

Return CM
• OUTLINE:

1. 50 years map of CBM algorithms
2. Advances last 5 years
3. Conclusions
• **CASE-BASED REASONING (CBR)**
  
  – Solves by retrieving similar problems (cases)
    
    • Retrieve, Adapt, Learn
  
  – ≠ model-based systems
    
    • No structure: atomic case = (problem, solution)
      
      • *Case Base*: Knowledge incremented *dynamically*
  
  – Challenge: number of cases
  
  – Approach: Case-Base **Maintenance** (CBM) alg.

!= Machine Learning: purpose of each case in CBR
COMPETENCE: capacity of each case to solve
- Coverage
- Reachability
- Competence Groups
### 50 years CBM map

#### NN period

<table>
<thead>
<tr>
<th>Year</th>
<th>Algorithm</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1968</td>
<td>CNN</td>
<td>Hart</td>
</tr>
<tr>
<td>1972</td>
<td>RNN</td>
<td>Gates</td>
</tr>
<tr>
<td>1972</td>
<td>ENN</td>
<td>Gates</td>
</tr>
<tr>
<td>1975</td>
<td>SNN</td>
<td>Ritter et al</td>
</tr>
<tr>
<td>1976</td>
<td>RENN</td>
<td>Tomek</td>
</tr>
<tr>
<td>1976</td>
<td>All-KNN</td>
<td>Tomek</td>
</tr>
<tr>
<td>1987</td>
<td>SHRINK</td>
<td>Kibler-Aha</td>
</tr>
<tr>
<td>1991</td>
<td>IB1 IB2 IB3 IB4</td>
<td>Aha-Kibler</td>
</tr>
<tr>
<td>1995</td>
<td>Competence Def.</td>
<td>Smyth-Keane</td>
</tr>
<tr>
<td>1997</td>
<td>Introsp. Learn</td>
<td>Bonzano et al</td>
</tr>
<tr>
<td>1997-2000</td>
<td>DROP 1 2 3 4 5</td>
<td>Wilson-Martinez</td>
</tr>
</tbody>
</table>

#### Competence period

<table>
<thead>
<tr>
<th>Year</th>
<th>Algorithm</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1968</td>
<td>Meta-CB</td>
<td>CumminsDerek</td>
</tr>
<tr>
<td>2011</td>
<td>IBL-DS</td>
<td>Beringer-Hüllermeier</td>
</tr>
<tr>
<td>2007</td>
<td>CTE</td>
<td>Craw et al</td>
</tr>
<tr>
<td>2007</td>
<td>CBE</td>
<td>Delany-Cunningham</td>
</tr>
<tr>
<td>2001-2000</td>
<td>COV RFD RC</td>
<td>McKenna-Smyth</td>
</tr>
<tr>
<td>2001</td>
<td>Maint. Retrieval</td>
<td>Craw-Jarmulak</td>
</tr>
<tr>
<td>2000</td>
<td>KeepItSimple</td>
<td>Yang-Wu</td>
</tr>
<tr>
<td>1999</td>
<td>ICF</td>
<td>Brighton-Mellis</td>
</tr>
<tr>
<td>1998</td>
<td>Competence Model</td>
<td>Smyth-McKenna</td>
</tr>
</tbody>
</table>
• OUTLINE:

1. 50 years map of CBM algorithms
2. Advances last 5 years
3. Conclusions
• Last 5 years
  – Intense research
  – Novel (theoretical) models
  – New problems (non existing)
  – Real life solutions
• Last 5 years: CBM map

Competence period

<table>
<thead>
<tr>
<th>Year</th>
<th>Algorithm</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1968</td>
<td>CNN</td>
<td>Hart</td>
</tr>
<tr>
<td>1972</td>
<td>RNN</td>
<td>Gates</td>
</tr>
<tr>
<td>1975</td>
<td>SNN</td>
<td>Ritter et al</td>
</tr>
<tr>
<td>1976</td>
<td>RENN</td>
<td>Tomek</td>
</tr>
<tr>
<td>1987</td>
<td>SHRINK</td>
<td>Kibler-Aha</td>
</tr>
<tr>
<td>1991</td>
<td>IB1 IB2 IB3 IB4</td>
<td>Aha-Kibler</td>
</tr>
<tr>
<td>1995</td>
<td>Competence Def.</td>
<td>Smyth-Keane</td>
</tr>
<tr>
<td>1997</td>
<td>Intros. Learn</td>
<td>Bonzano et al</td>
</tr>
<tr>
<td>1997-2000</td>
<td>DROP 1 2 3 4 5</td>
<td>Wilson-Martinez</td>
</tr>
</tbody>
</table>

Meta-CB
Cummins-Derek
2011

IBL-DS
Beringer-Hüllermeier
2007

CTE
Craw et al
2007

CBE
Delany-Cunningham
2004

COV
McKenna-Smyth
2001-2000

RC
Maint. Retrieval
2001

IB1 IB2 IB3 IB4
Aha-Kibler

Competence Model
Smyth-McKenna
1998

Competence Improvements

Re-structuring case-bases

Time

Period 2014-2018

NN period

<table>
<thead>
<tr>
<th>Year</th>
<th>Algorithm</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td>Pref-CBM</td>
<td>Abdel-Hüllermeier</td>
</tr>
<tr>
<td>2015</td>
<td>FP-CA</td>
<td>Mathew-Chakraborti</td>
</tr>
<tr>
<td>2014</td>
<td>T-CBM</td>
<td>Lupiani et al</td>
</tr>
<tr>
<td>2016</td>
<td>NEFCS/SSR</td>
<td>Lu et al</td>
</tr>
<tr>
<td>2015</td>
<td>CBNI</td>
<td>Yamamoto et al</td>
</tr>
<tr>
<td>2016</td>
<td>MOE-CBM</td>
<td>Lupiani et al</td>
</tr>
</tbody>
</table>
• Last 5 years: complex problems
  
  – Multi-objective optimization
  
  • Search Space: Case-Base
  
  • Size vs. Accuracy
  
  • Noise vs. Redundancy
  
  • MOEA: NSGA-II
  
  • Optimization:
    
    – Min(no redundant cases)
    
    – Min(dist. Non-redundant)
    
    – Max(accuracy system)
• Last 5 years: improving competence
  – Closure-Competence Model
  • Inadequate Comp. Model
  • Disjoint partitions (consistent)

\[
\text{CompetenceClosure}(G) \iff \forall c, c' \in G, \\
\exists \text{SharedCoveragePath}(c, c') \land \\
\forall c_k \in C - G, \exists c \in G : \text{SharedCoverage}(c_k, c)
\]

Lu-Zhang-Lu Competence Model Proposal 2014

Last 5 years: temporal dimension (real life problems)

- Temporal CBM
- Case: sequence of events
- Adapt CBM approaches

Monitoring Elderly People At Home: falling detection


• Last 5 years: re-structuring case-bases
  
  – Flexible Feature Deletion
  
  • Remove part case
  • Less competence loss
  • High dimensional problems

FFD Leak-Schack, 2015
• Last 5 years: re-structuring case-bases
  – Preference CBR Model
    • Solution Si preferred to Sj
    • Decompose case structure
    • c=(prob,sol,PrefSet)

Pref-CBR approach Abdel-Aziz & Hüllermeier, 2015

Example Pref-CBM approach

• OUTLINE:

1. 50 years map of CBM algorithms
2. Advances last 5 years
3. Conclusions
• Conclusions

  – Deep Impact of Competence Model Smyth-McKenna-Keane

  – New potential CBR applications (not explored in depth):
    • Interpretable reduction of high dimensional problems: e.g. Flexible Feature Deletion [Leake & Schack, 2015]
    • Social network datasets: e.g. Compositional Adaptation [Mathew & Chakraborti, 2017]

  – Applications in the Industry:
    • Monitoring: T-CBM [Lupiani et al 2015]
    • Long term use of intelligent systems: Drift-CBM [Lu et al 2016]

  – Future directions:
    • Fair comparisons (now limited to classical CBM alg.)
Acknowledgements (CBM algorithms authors)

A. Abdel-Aziz  M. K. Haouchine  R. Rowe
D. Aha  T. Kawabe  M. Salamó
M.K. Albert  M.T. Keane  B. Schack
J. Beringer  D. Kibler  A. Smiti
A. Bonzano  R. Knauf  B. Smyth
D. Bridge  Y. Kobayashi  Y. Sukarai
H. Brighton  D. Leake  I. Tomek
S. Chakraborti  R. Lopez de Mantaras  S. Tsuruta
B. Chebel-Morello  S.R. Lowry  D.R. Wilson
S. Craw  J. Lu  N. Wiratunga
L. Cummins  N. Lu  H.B. Woodruff
P. Cunningham  E. Lupiani  J. Wu
Z. Elouedi  R. Marin  Y. Yamamoto
G. Gates  T.R. Martinez  Q. Yang
E. Golobardes  S. Massie  N. Zerhouni
P. Hart  D. Mathew  G. Zhang
E. Hüllermeier  E. McKenna
T.L. Isenhour  C. Mellish
S. J. Delany  J. Palma
J. Jarmulak  G.L. Ritter
THANK YOU FOR YOUR ATTENTION

Jose M. Juarez
jmjuarez@um.es

DOWNLOAD PAPER (OPEN ACCESS):