Improving Recognition Algorithm Evaluation through Explicit Decision-Making

Richard Zanibbi*, Dorothea Blostein+, and James R. Cordy+

April 29, 2008
*Department of Computer Science, Rochester Institute of Technology
+School of Computing, Queen’s University
Pattern Recognition Tasks as Decision Problems

Problem: Pattern Classification
From a set of classes, decide the correct class for an input
• e.g. OCR, Fingerprint recognition, Face recognition

Structure Model
A set of structures (graphs) for describing the composition of an input
• May be used for pattern classification
• e.g. Line and line intersection model for digits (e.g. for a “4”)

Problem: Structural Pattern Recognition
Given a structure model, decide the correct structure for an input
• Model instance decided through a series of decisions that locate (segment), classify, and relate (parse) input regions
• e.g. Speech recognition, Molecular Scene Analysis, Table Recognition
Table Recognition

Table Detection
locating table regions

Table Structure Recognition
determining table composition

### Table 1. Document Recognition Journals

<table>
<thead>
<tr>
<th>Journal</th>
<th>Full Name</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPAMI</td>
<td>IEEE Transactions on Pattern Analysis and Machine Intelligence</td>
<td>monthly, IEEE</td>
</tr>
<tr>
<td>IJDAR</td>
<td>International Journal on Document Analysis and Recognition</td>
<td>quarterly, Springer-Verlag</td>
</tr>
<tr>
<td>PR</td>
<td>Pattern Recognition</td>
<td>monthly, Elsevier</td>
</tr>
<tr>
<td>IPRAI</td>
<td>International Journal on Pattern Recognition and Artificial Intelligence</td>
<td>eight times/year, World Scientific</td>
</tr>
</tbody>
</table>

Source: from a listing of pattern recognition journals provided online at [http://www.ph.usb.br/pt/Info](http://www.ph.usb.br/pt/Info)

---

Reuse
Compression
Information Retrieval

... (TPAMI, IEEE ..., monthly, IEEE)
(IJDAR, Intern..., quarterly, Springer-Verlag)
...
Overview

• Decision-based specification: collecting information about recognition decisions
  • Allow us to address the question: how do decisions interact and affect one another?

• Using decision histories to observe new performance metrics (historical recall and precision)

• Example: comparison and combination of two published methods for recognizing table structure using conventional and historical metrics
Decision-Based Specification of Recognition Algorithms

The Recognition Strategy Language (RSL)
Implementing Recognition Algorithms

Conventional Approach
Algorithms, interpretations, supporting infrastructure in language of choice (C/C++, Lisp, Prolog, Ocaml, Haskell….)

“What and how” defined using one representation

Decision-Based Approach (What vs. How)
Use a very high-level “scripting” language
• to organize decisions around the structure of interpretations
• to define inputs/outputs for decisions, decision parameters
• to record and update interpretations from decision results

Concrete decision functions: any language (w. transl.)
RSL: a functional decision-based specification language
• Interpretations represented as directed attributed graphs
1. Translate RSL to TXL (Using TXL)

2. Pass Input Graph (text file) to Program

3. Output (text files):
   - Accepted interpretations (graphs)
   - Log containing the decision history
Example RSL Program

model regions
    *REGION Word Cell Header Entry Image
end regions

model relations
    *contains adj_right close_to
end model

recognition parameters
    sMaxHorDistance  5.0  %mms
    sMaxColSep       20.0 %mms
    sMaxHorSep       25.0 %mms
end parameters
strategy main  %% initial interpretation, params as input
   relate { Word } regions with { hor_adj } using
      selectHorizAdjRegions(sMaxHorDistance)

   segment { Word } regions into { Cell } using
      segmentHorizAdjRegions()
      observing { hor_adj } relations

   classify { Cell } regions as { Header, Entry } using
      labelColumnHeaderAndEntries()

   %% ... other operations/strategy function calls

   accept interpretations

end strategy
Input (Word Bounding Boxes)

End (Month)
August

Decision 1
relate {Word} regions with {hor_adj} using
selectHorizAdjRegions(sMaxHorDistance)

Decision 2
segment {Word} regions into {
segmentHorizAdjRegions()
observing {hor_adj} relat}
11

: h {hor_adj} using
: (sMaxHorDistance)

segment {Word} regions into {Cell} using
segmentHorizAdjRegions()
observing {hor_adj} relations

classify {Cell} regions
labelColumnHeaders

**Decision 2**

**Parameters**
- sMaxHorDistance
- aMaxColSep
- aMaxRowSep

**Interpretation**

**Transform**

**Observe (Word, hor_adj)**

**Decision 2 Interpreter**

segmentHorizAdjRegions

Alternatives (0 or more of):
1. (Word-1)
2. (Word-2)
3. (Word-3)
4. (Word-1, Word-2)
5. (Word-1, Word-3)
6. (Word-2, Word-3)
7. (Word-1, Word-2, Word-3)

**Decision Record**
- (Word-1, Word-2, Word-3)

**Observe (Cell)**

**Decision 2 Labeling**

**Parameters**
- sMaxHorDistance
- aMaxColSep
- aMaxRowSep

**Interpretation**

**Transform**

1. Cell
2. Cell

**Label Column Headers**

Alternatives (0 or more of):
1. (Cell-1, Header)
2. (Cell-1, Entry)
3. (Cell-2, Header)
4. (Cell-2, Entry)
Decision 2

segment {Word} regions into {Cell} using segmentHorizAdjRegions()
observing {hor_adj} relations

d-3)

Decision 3

classify {Cell} regions as {Header, Entry} using labelColumnHeadersAndEntries()

interpretation

Parameters
sMaxHorDistance
aMaxColSep
aMaxRowSep

Transformation

Interpretation

end (Month)
August

alt. (0 or more of):
1. (Cell-1, Header)
2. (Cell-1, Entry)
3. (Cell-2, Header)
4. (Cell-2, Entry)

Decision Record

(Word-1, Word-2),
(Word-3)
Decision Histories & Evaluation
New Metrics from Decision Histories: Historical Recall and Precision

Accepted Hypotheses (A)

Precision: \( \frac{|C|}{|A|} \)

Correct (C)

Recall: \( \frac{|C|}{|T|} \)

Recognition Targets (T)

False Negatives (F)

Recognition Targets: Correct Hypotheses
<table>
<thead>
<tr>
<th>Hypothesis History</th>
<th>0 (no cells)</th>
<th>1 (all words -&gt; cells)</th>
<th>2 (merge adj. cells)</th>
<th>3 (split at wide word gaps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accepted Cells</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Borrowed Books</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Book</td>
<td>Returned</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAS '04</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICDAR '03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICPR '02</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rejected Cells</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recall</td>
<td>4/8 (50.0%)</td>
<td>2/8 (25.0%)</td>
<td>8/8 (100.0%)</td>
<td></td>
</tr>
<tr>
<td>Precision</td>
<td>4/12 (33.3%)</td>
<td>2/5 (40.0%)</td>
<td>8/8 (100.0%)</td>
<td></td>
</tr>
<tr>
<td>Historical Recall</td>
<td>4/8 (50.0%)</td>
<td>6/8 (66.7%)</td>
<td>8/8 (100.0%)</td>
<td></td>
</tr>
<tr>
<td>Historical Precision</td>
<td>4/12 (33.3%)</td>
<td>6/17 (35.3%)</td>
<td>8/19 (42.1%)</td>
<td></td>
</tr>
</tbody>
</table>
Using Decision Histories: Frames of Reference for Evaluation

Goal-based Evaluation ("Normal")
Analyze decision outcomes relative to ground truth
  • recall, precision of decision given goal structures

Alternative-based Evaluation
Analyze decision outcomes relative to available alternatives in the decision space
  • recall, precision of decision given valid alternatives
  • recall, precision of alternatives given goal structures
  • probability of randomly selecting a valid alternative?

  • Do poor decisions result from poor alternative selection, or a poor set of alternatives (from previous decisions)?
Example: Comparing and Combining Table Structure Recognition Algorithms

Algorithms:
Handley (2001)
Hu et al. (2001)
<table>
<thead>
<tr>
<th>Information Used</th>
<th>(Handley, 2001)</th>
<th>(Hu et al., 2001)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Geometric</td>
<td>Geometric, Lexical</td>
</tr>
<tr>
<td>Recognition Strategy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Cells repeatedly merged/split</td>
<td>• Hierarchical word clustering: columns</td>
</tr>
<tr>
<td></td>
<td>• Fully ruled tables: special case</td>
<td>• Word lexical types affect column, row structure</td>
</tr>
<tr>
<td></td>
<td>• Lines, cell projection gaps provide separators</td>
<td>• Cells produced after columns and rows</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Added text line detection</td>
</tr>
<tr>
<td>RSL Program</td>
<td>~540 lines</td>
<td>~240 lines</td>
</tr>
<tr>
<td>Types</td>
<td>15 Region, 6 Relation</td>
<td>22 Region, 2 Relation</td>
</tr>
<tr>
<td>Rec. Parameters</td>
<td>26</td>
<td>8</td>
</tr>
<tr>
<td>Ext. Decisions (TXL)</td>
<td>~5000 lines</td>
<td>~3000 lines</td>
</tr>
</tbody>
</table>

Input
- Words BBs, Lines
- Word BBs

Output
- Cells, Separators
- Cells, Rows, Columns

Index Structure
- Boxhead, Stub, Body

Recognition Strategy
- Fully ruled tables: special case
- Lines, cell projection gaps provide separators
- Added text line detection
Combining Recognition Algorithms

Classifier Combination

Classifier combinations often out-perform single classifiers in complex domains (e.g. OCR)
- Combination techniques include majority vote, boosting (AdaBoost (Freunde & Schapire)), learning combinations, game-based approaches
- Key idea: integrating matches to multiple models more informative than a single model (“different views”)

Structural Recognition Alg. Combinations

How to combine systems that include segmentation, classification, and relating objects?
- Idea: use a decision-based representation (e.g. RSL) to organize decision functions, to allow easy re-sequencing, alteration (e.g. through parameters), and combined.
Table 49.—Average values for bulk density, grain density, and total pore space of gray dacite from the lateral-blast deposits and of pumice lapilli from pyroclastic-flow deposits of Mount St. Helens

<table>
<thead>
<tr>
<th>Type of deposit</th>
<th>Bulk density (g/cm³)</th>
<th>Grain density (g/cm³)</th>
<th>Total pore space (percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lateral blast, May 18</td>
<td>2.166</td>
<td>262</td>
<td>2.52</td>
</tr>
<tr>
<td>Pyroclastic flow, May 18</td>
<td>0.74</td>
<td>8</td>
<td>2.55</td>
</tr>
<tr>
<td>May 25</td>
<td>0.95</td>
<td>2</td>
<td>2.53</td>
</tr>
<tr>
<td>June 12</td>
<td>1.08</td>
<td>10</td>
<td>2.53</td>
</tr>
<tr>
<td>July 22</td>
<td>0.88</td>
<td>11</td>
<td>2.55</td>
</tr>
<tr>
<td>August 7</td>
<td>1.02</td>
<td>12</td>
<td>2.61</td>
</tr>
<tr>
<td>October 16–18</td>
<td>1.12</td>
<td>12</td>
<td>2.65</td>
</tr>
</tbody>
</table>

1 Number of determinations.
2 Data from Hoblitt and others (this volume).
3 Grain density (Dg) not determined; total pore space calculated using Dg=2.60.
<table>
<thead>
<tr>
<th>Type of deposit</th>
<th>Bulk density</th>
<th>Grain density</th>
<th>Total pore space (percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (g/cm³)</td>
<td>Mean (g/cm³)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No. 1</td>
<td>No. 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>Lateral blast, May 18--</td>
<td>21.66</td>
<td>2.52</td>
<td>3</td>
</tr>
<tr>
<td>Pyroclastic flow,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>May 18------</td>
<td>.74</td>
<td>2.55</td>
<td>3</td>
</tr>
<tr>
<td>May 25------</td>
<td>.95</td>
<td>2.55</td>
<td>3</td>
</tr>
<tr>
<td>June 12------</td>
<td>1.08</td>
<td>2.53</td>
<td>3</td>
</tr>
<tr>
<td>July 22------</td>
<td>.88</td>
<td>2.55</td>
<td>1</td>
</tr>
<tr>
<td>August 7-----</td>
<td>1.02</td>
<td>2.61</td>
<td>3</td>
</tr>
<tr>
<td>October 16-18</td>
<td>1.12</td>
<td>2.65</td>
<td>5</td>
</tr>
</tbody>
</table>

**Handley**
R: 25/52 = 48%
P: 25/38 = 66%

<table>
<thead>
<tr>
<th>Type of deposit</th>
<th>Bulk density</th>
<th>Grain density</th>
<th>Total pore space (percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (g/cm³)</td>
<td>Mean (g/cm³)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No. 1</td>
<td>No. 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>Lateral blast, May 18--</td>
<td>21.66</td>
<td>2.52</td>
<td>3</td>
</tr>
<tr>
<td>Pyroclastic flow,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>May 18------</td>
<td>.74</td>
<td>2.55</td>
<td>3</td>
</tr>
<tr>
<td>May 25------</td>
<td>.95</td>
<td>2.55</td>
<td>3</td>
</tr>
<tr>
<td>June 12------</td>
<td>1.08</td>
<td>2.53</td>
<td>3</td>
</tr>
<tr>
<td>July 22------</td>
<td>.88</td>
<td>2.55</td>
<td>1</td>
</tr>
<tr>
<td>August 7-----</td>
<td>1.02</td>
<td>2.61</td>
<td>3</td>
</tr>
<tr>
<td>October 16-18</td>
<td>1.12</td>
<td>2.65</td>
<td>5</td>
</tr>
</tbody>
</table>

**Hu et al.**
R: 45/52 = 87%
P: 45/52 = 87%
(RSL) Combination of Handley and Hu et al. Algorithms

<table>
<thead>
<tr>
<th>Type of deposit</th>
<th>Bulk density</th>
<th>Grain density</th>
<th>Total pore space (percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (g/cm³)</td>
<td>Mean (g/cm³)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No. 1</td>
<td>No. 1</td>
<td></td>
</tr>
<tr>
<td>Lateral blast,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>May 18---------</td>
<td>2.166</td>
<td>2.52</td>
<td>3</td>
</tr>
<tr>
<td>Pyroclastic flow,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>May 18---------</td>
<td>.74</td>
<td>2.55</td>
<td>3</td>
</tr>
<tr>
<td>May 25---------</td>
<td>.95</td>
<td>(3)</td>
<td>0</td>
</tr>
<tr>
<td>June 12--------</td>
<td>1.08</td>
<td>2.53</td>
<td>3</td>
</tr>
<tr>
<td>July 22--------</td>
<td>.88</td>
<td>2.55</td>
<td>1</td>
</tr>
<tr>
<td>August 7-------</td>
<td>1.02</td>
<td>2.61</td>
<td>3</td>
</tr>
<tr>
<td>October 16-18</td>
<td>1.12</td>
<td>2.65</td>
<td>5</td>
</tr>
</tbody>
</table>

Recall: 51/52 = 98%
Precision: 51/53 = 96%
Cell Detection Results (Hu et al., 2001)
RSL Re-implementation on Table ‘a038’ (UW-III)

<table>
<thead>
<tr>
<th>Header</th>
<th>Cells</th>
<th>(Dec. 86)</th>
</tr>
</thead>
</table>

![Graph showing percentage vs. decision number](image-url)
Cell Detection Results (Handley, 2001)
RSL Re-implementation on Table ‘a038’ (UW-III)

0: Input (words and lines)

1: Classify words as cells

16: Merge ‘horizontally close’ cells

35: Merge cells sharing column, row assignments. **Nearly 50% of correct cells rejected; new correct cells also detected**

47: Two cells merged producing column header

51: Merge header cells bounded by two horizontal lines

83: Merge cells sharing line and white space separators

*Inference times shown are those affecting cells*
Combined Cell Detection Results

Run Hu, stop after header cells identified

Handley mod.:

- Keep passed cells at first step (all words labelled as cells)
- Weak merge decisions removed (32, 59, 86, along with supporting decisions)
Future Directions

Decision-Based Specification in RSL

- Support feature computations (currently defined inside decision functions)
- Support automated collection and summary of intermediate interpretations, evaluation metrics

Evaluation

- Other metrics for decision histories (fickleness?)
- Reporting historical metrics over large test sets (mean & variance, other?)

Combination and Optimization

- Extend RSL to support full/partially automated construction of RSL programs from existing ones
- Tuning of parameters, e.g. Graph Transformer Networks (LeCun et al.)
- Explore useful operations based on performance data for integrating decision-based specs (genetic algs?)
Thank you.

Thanks to Joshua Zimler and Ben Steele for help with figures.

Support:

RIT Department of Computer Science
Table Structure Model for (Handley, 2001) RSL Implementation

- **Region type contains**
- **Region type classified as**
- **Region relation (tail) is (relation) to (head)**
Dependency Graph for (Hu et al., 2001) 
RSL Implementation

Region/Relation/Parameter (head) depends on (tail)
RIT Document and Pattern Recognition Lab (DPRL)

Goals:
1. Improve theory and tools for constructing and evaluating pattern recognition systems
2. Apply these to problems in document recognition and pen-based computing

Members:
- Richard Zanibbi
- Kurt Kluever (Master’s student)
- New members welcome!

http://www.cs.rit.edu/~rlaz/dprl.html
Current Directions:

1. Theory and Tools:
   • Tools for recognition module integration and evaluation, such as the Recognition Strategy Language (*Zanibbi et al.*)
   • Game-theoretic models of recognition problems and systems (e.g. for classifier combination)
   • Machine learning algorithms for system optimization

2. Applications:
   • Pen and image-based math entry (lab maintains open-source Freehand Formula Entry System (*Smithies, Novins, Arvo, Zanibbi et al.*))
   • Optical character recognition (OCR)
   • Image and text-based document retrieval
   • “CAPTCHAs” (for distinguishing humans from 'bots')
   • Table recognition, etc.
Table Recognition: Surveys


