Experienced Physicians and Automatic Generation of Decision Rules from Clinical Data

William Klement[1], Szymon Wilk[1,2], Martin Michalowski[4], and Ken Farion[3]
[1]Mobile Emergency Triage (MET) Research Group, University of Ottawa, Canada.
[2]Institute of Computer Science, Poznan University of Technology, Poland.
[3]Division of Emergency Medicine, CHEO, Canada.
[4]Adventium Labs, USA.

Introduction

- MET Research deals with engineering clinical decision-making knowledge embedded in Clinical Decision Support Systems.
- Clinical knowledge is acquired from data and from experts.
- Clinical data
  - describe patients with a clinical condition, and
  - the decision specifies a diagnostic outcome.
- Diagnostic decisions are
  - made by Experienced Physicians (EP) and
  - verified to produce the Gold Standard (GS).
- The objective is to acquire unbiased clinical knowledge.

Clinical Diagnostic Domains

- Patient records are instances of a relationship between patient clinical state and a diagnostic decision.
- EPs make good decisions especially for critical patients.
- Discrepancies between GS and EP decisions show that EPs are cautious and over-diagnose healthy patients.
- Their decisions are highly sensitive but are less specific.
- Low specificity is costly, noisy, and unnecessary.

Classical Methods

- Seek relationships between patient attribute values and GS.
- Data is collected by EPs but GS is derived.
- Clinical knowledge is acquired from all records.
- This knowledge is distorted by erroneous characteristics associated with EP decisions.
- EP decisions matching GS provide sound clinical knowledge.

This Study

- Eliminates EP bias (removes mismatches).
- Demonstrates that correct EP diagnosis produce better clinical knowledge.
- Using two real-life clinical domains, we:
  - filter the data by matching EP decisions with the GS,
  - extract clinical knowledge as decision rules, and
  - assess their quality using rule performance metrics.

The Role of EP's Expertise

- Clinical Decision Support Systems (CDSS) provide:
  - information management,
  - alerts to specific events, and
  - patient-specific recommendations.
- Patient-specific CDSS reflect clinician's expertise.
- An experienced physician offers vast knowledge.
- We capture EPs' knowledge by focusing on correct practice.
- We rely on correct EP decisions (EP decision match GS).

Asthma Exacerbation Data

- Is prospectively collected at CHEO.
- Includes patients visiting the ED with Asthma Exacerbation.
- Contains patient history, nursing, physician triage assessment, reassessment after 2 hours, and the verified severity (GS).
- We want to predict the severity as mild or other (critical and moderate).

Abdominal Pain Data

- Also collected in the ED at CHEO.
- Includes patients with Appendicitis who require surgery.
- Most records describe benign cases.
- A definite diagnosis is difficult to obtain during an ED visit (symptoms often resolve without complications).
- The task is to select the correct patient management plan (discharge, observation, or a specialty consult).
- The verified patient outcome is the GS.

Decision Rules

- Given patient records, identify a set of conditions S and relate them to a set of decisions D.
- A decision rule has the form:
  if $s_i, s_j, \cdots, s_w$ then d_v where $s_i, s_j, \cdots, s_w \in S$ and $d_v \in D$
- The contingency table of a decision rule if X then Y

	Y	$\overline{Y}$	
X	a	b	r_1
$\overline{X}$	c	d	r_2
	c_1	c_2	

  - number of rules, conditions $|S|$, and average length $|X|$.
  - $coverage(X, Y) = \frac{a}{c_1}$, $certainty(X, Y) = \frac{a}{r_1}$, and $confirmation(X, Y) = \frac{ad-bc}{ad+bc+2ac}$ (Greco et al 2004)

Experimental Design

- Demonstrate the superiority of knowledge extracted from cases where EP decisions match the GS (EP filtering).
- Improved rules quality and classification performance
  - Phase I: generate decision rules.
  - Phase II: 5 runs of 10-fold cross validation of classification.
- Evaluation:
  - Rule assessment metrics.
  - Sensitivity, Specificity, Accuracy, and Geometric Mean.

Characteristics of Data

		GS Outcome			EP Decisions	
Data	Examples	Positives	Negatives	Ratio	Positives	Negatives
AE_{all}	240	131	109	55%	136	72
AE_{corr}	140	90	50	64%	90	50
AP_{all}	457	48	409	11%	55	402
AP_{corr}	422	34	388	8%	34	388

- *all* includes all patient records.
- *corr* contains patients EP filtered data.
- AE = Asthma Exacerbation, AP = Abdominal Pain.
- APS = Abdominal Pain with under-sampling the majority class.

Phase I Results: Characteristics of Rules

	Both Classes			+ Class			− Class		
Data	Cond.	Rules	Ave. Length	Cond.	Rules	Ave. Length	Cond.	Rules	Ave. Length
AE_{all}	199	50	3.98 $\pm$1.19	92	23	4.00 $\pm$1.00	107	27	3.96 $\pm$1.34
AE_{corr}	13	6	2.17 $\pm$0.98	5	3	1.67 $\pm$1.16	8	3	2.67 $\pm$0.58
AP_{all}	163	38	4.29 $\pm$1.51	99	21	4.71 $\pm$1.65	64	17	3.77 $\pm$1.15
AP_{corr}	83	24	3.46 $\pm$1.14	37	10	3.70 $\pm$1.25	46	14	3.29 $\pm$1.07
APS_{all}	91	25	3.64 $\pm$0.95	51	14	3.64 $\pm$1.08	40	11	3.64 $\pm$0.81
APS_{corr}	19	8	2.38 $\pm$0.52	9	4	2.25 $\pm$0.50	10	4	2.50 $\pm$0.58

- Correct EP decisions produce fewer rules, fewer conditions, and shorter average rule length with a lower standard deviation.

Phase I Results: Performance of Rules

Measure	Data	Both classes	+ Class	− Class
Coverage	AE_{all}	0.055 ±0.044	0.062 ±0.052	0.048 ±0.036
	AE_{corr}	0.534 ±0.311	0.507 ±0.362	0.560 ±0.330
	AP_{all}	0.117 ±0.153	0.068 ±0.051	0.176 ±0.209
	AP_{corr}	0.158 ±0.204	0.176 ±0.130	0.145 ±0.248
	APS_{all}	0.137 ±0.132	0.131 ±0.128	0.145 ±0.143
	APS_{corr}	0.423 ±0.209	0.338 ±0.098	0.507 ±0.271
Certainty	AE_{all}	1.000 ±0.000	1.000 ±0.000	1.000 ±0.000
	AE_{corr}	1.000 ±0.000	1.000 ±0.000	1.000 ±0.000
	AP_{all}	0.955 ±0.157	0.929 ±0.208	0.988 ±0.030
	AP_{corr}	1.000 ±0.000	1.000 ±0.000	1.000 ±0.000
	APS_{all}	0.969 ±0.104	0.964 ±0.134	0.974 ±0.053
	APS_{corr}	1.000 ±0.000	1.000 ±0.000	1.000 ±0.000
Confirmation	AE_{all}	1.000 ±0.000	1.000 ±0.000	1.000 ±0.000
	AE_{corr}	1.000 ±0.000	1.000 ±0.000	1.000 ±0.000
	AP_{all}	0.914 ±0.219	0.953 ±0.165	0.867 ±0.269
	AP_{corr}	1.000 ±0.000	1.000 ±0.000	1.000 ±0.000
	APS_{all}	0.937 ±0.210	0.928 ±0.270	0.949 ±0.104
	APS_{corr}	1.000 ±0.000	1.000 ±0.000	1.000 ±0.000

- Increased rule coverage particularly with under-sampling.
- Rule certainty and confirmation may also be increased.

Phase II Results: Classification Performance

Data	Sensitivity	Specificity	Accuracy	Geometric Mean†
AE_{all}	0.7024 ± 0.1510	0.5930 ± 0.1634	0.7158 ± 0.0865	0.6341 ± 0.1059
AE_{corr}	0.7908 ± 0.1268	0.5070 ± 0.1309	0.6825 ± 0.0803	0.6243 ± 0.1029
AP_{all}	0.4930 ± 0.2231	0.9640 ± 0.0264	0.9252 ± 0.0322	0.6619 ± 0.1952
AP_{corr}	0.5877 ± 0.2236	0.9526 ± 0.0368	0.9305 ± 0.0310	0.7284 ± 0.1696
APS_{all}	0.7913 ± 0.2246	0.7856 ± 0.0525	0.8127 ± 0.0406	0.7777 ± 0.1296
APS_{corr}	0.7470 ± 0.2353	0.8560 ± 0.0642	0.8559 ± 0.0509	0.7909 ± 0.1396

† Entries are averaged over 5 runs of 10-fold cross validation.

- AE: higher sensitivity and lower specificity, accuracy and geometric mean.
- AP: higher sensitivity, accuracy, and geometric mean, but lower specificity.
- PAS: balanced performance everywhere (EP filtering + under-sampling improves the performance).

Conclusions

- Capturing clinical knowledge from examples of correct EP decisions produces better knowledge.
- EP-based filtering reduces the complexity of rules.
- The performance of rules is enhanced with better coverage, higher certainty, and increased confirmation.
- The enhanced quality of clinical knowledge improves the classification performance with increased sensitivity.

Future Work

- Assess the impact of our filtering on other classification models (decision trees, Bayesian classification, ensemble methods).
- Explore its potentials on critical patients.
- Investigate the integrity of the GS and class labels.
- How can we exploit relationships between GS, labels, and EP decisions in clinical data?
- What are the benefits? Are there any?

Questions?
Thank you.