Holmes – Hybrid Ontological & Learning MEdical System Decision Support System

(Atif Khan, John A. Doucette and Robin Cohen)

WATERLOO CHERITON SCHOOL OF COMPUTER SCIENCE

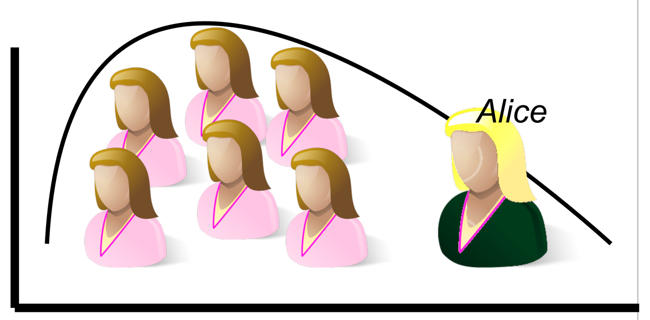
Introduction

Scenario

Treatment Considerations

- who is Alice? (black swan theory)
- missing information
- Alice's medical history
- nature of the prescription
- who is administering the drug?
- knowledge & time constraints

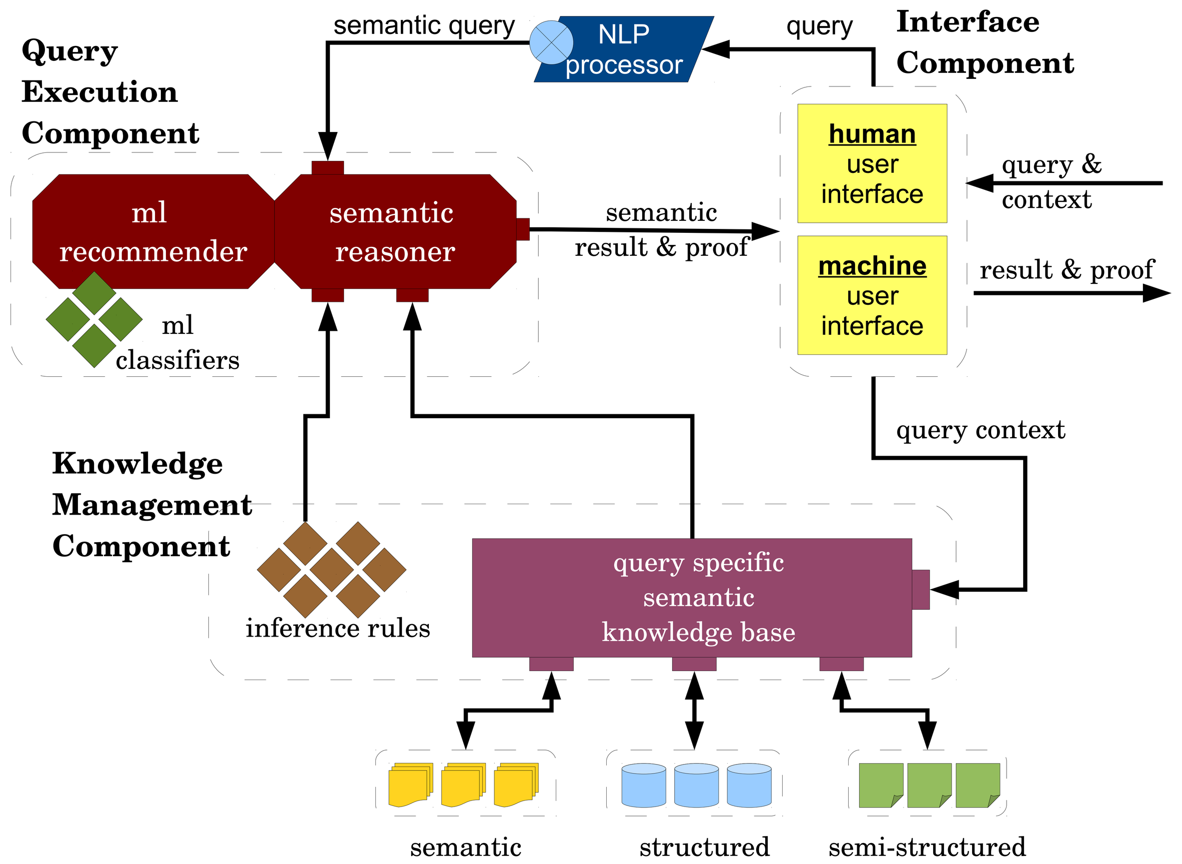
Objectives



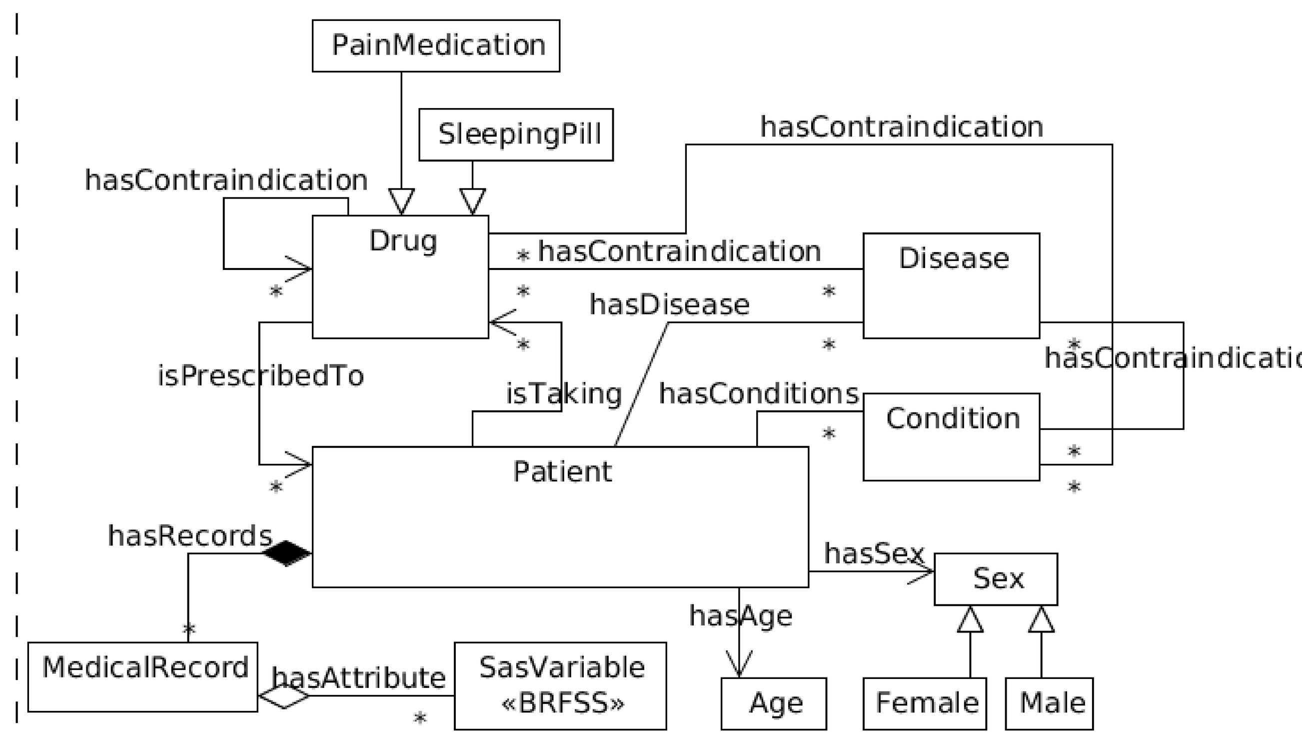
- patient-centric evidence-based medical decision support
- tolerant to noise in patient data → **information challenge**
- automated machine-processable decision making
- operates in constraint environments
- decisions are easy to explain & verify

Architectural Components, Knowledge Representation and Automated Reasoning

Architecture & Components



Semantic Model



Hybrid Decision Making

1. *use a semantic reasoner for knowledge-based reasoning over the structured data using inference rules*
2. *if knowledge is missing then predict missing information using machine learning techniques.*
3. *reevaluate using semantic reasoner*

Algorithm 1 Hybrid Decision Making Algorithm

Let $query$ be the user query, KB be a knowledge base, and $rules$ be set of inference rules.
Let $reasoner$ be the semantic reasoner and $mlrecommender$ be an imputation model.
{First the reasoner attempts to answer the query by itself.}
$reponse[r,p] \leftarrow reasoner.doProof(query, KB, rules)$
{If the reasoner is successful, return the result.}
if $reponse[r,p] \notin \emptyset$ **then**
  return $reponse.r$, $reponse.p$, with confidence 1.
end if
{Otherwise, if the query answer is negative, return an empty result.}
$noresult \leftarrow inspectForFalseModel(proof)$
$unknownresult \leftarrow inspectForCounterModel(proof)$
if $noresult$ and not $unknownresult$ **then**
  return null response, null proof, confidence of 1;
end if
{If the query is presently unanswerable, impute the missing values and answer it.}
if $unknownresult$ **then**
  $predictedValues[] \leftarrow mlrecommender.impute(KB)$
  $KB \leftarrow KB \cup predictedValues[]$
  $reponse[r,p] \leftarrow reasoner.doProof(query, KB, rules)$
  $conf \leftarrow \Pi_{p \in predicedValues[]} p.conf$
  return $reponse.r$, $reponse.p$, $conf$
end if

Proposed Construction

1) Knowledge-based Decision Support System

- structured data representation (**knowledge base**)
- expert knowledge → **inference rules**
  - mimics human thinking
  - heuristic based, evidence based etc.
- reasoning capability (**inference engine**)
- **decisions** are based on rules (axioms) and **can be easily explained**
- decisions are **easy to verify/validate**
- quite **powerful** & robust only when **knowledge-base is complete**

+

2) Learning-based Decision Support System

- learns from raw data and past past examples/cases
  - **patterns** in the clinical data
- utilizes machine learning techniques
- **requires training process** to create inference models
  - training is specific to a line of inquiry
  - training is expensive
- system decisions are often **hard to explain & verify**
- **tolerant to noise** in data
- **effective in finding latent relationships** between data attributes

=

Holmes

- **automates reasoning** using semantic knowledge representation and inference
- **system-made decisions** are easy to verify & explain
- **extremely tolerant to noise** in data
- **suitable for** a diverse set of medical personal & settings

System Evaluation — OMeD Vs. Machine Learning Techniques

Setup

Line of inquiry – sleeping pill prescription

"What sleeping pills can I give to Alice?"

Evaluate criteria

– balanced accuracy = avg(*specificity, sensitivity*)
  – *specificity = tn / (tn + fp)*
  – *sensitivity = tp / (tp + fn)*

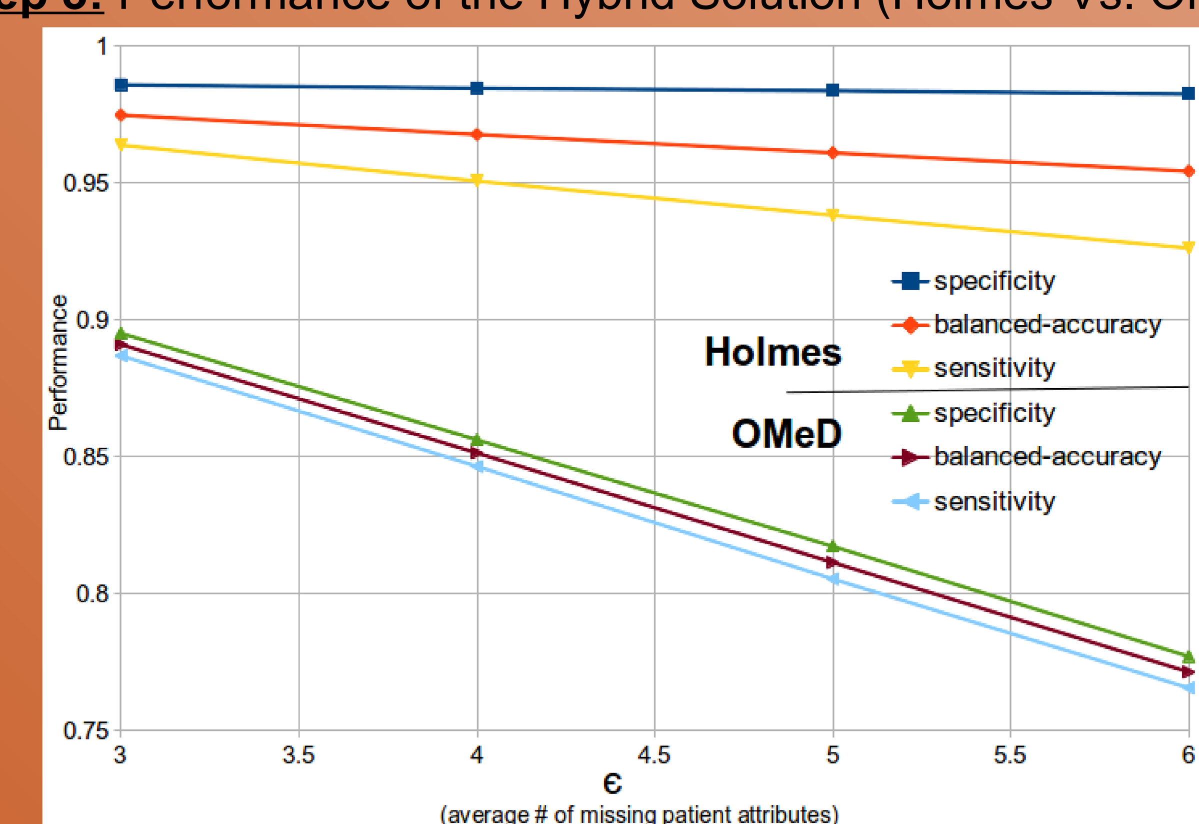
Datasets (real world)

Patient records – CDC Behavioral Risk Factor Surveillance System 2010 dataset

Expert knowledge

– Mayo clinic sleeping pill prescription protocol
– Drug.com drug-to-drug interaction registry

Step 3: Performance of the Hybrid Solution (Holmes Vs. OMeD)

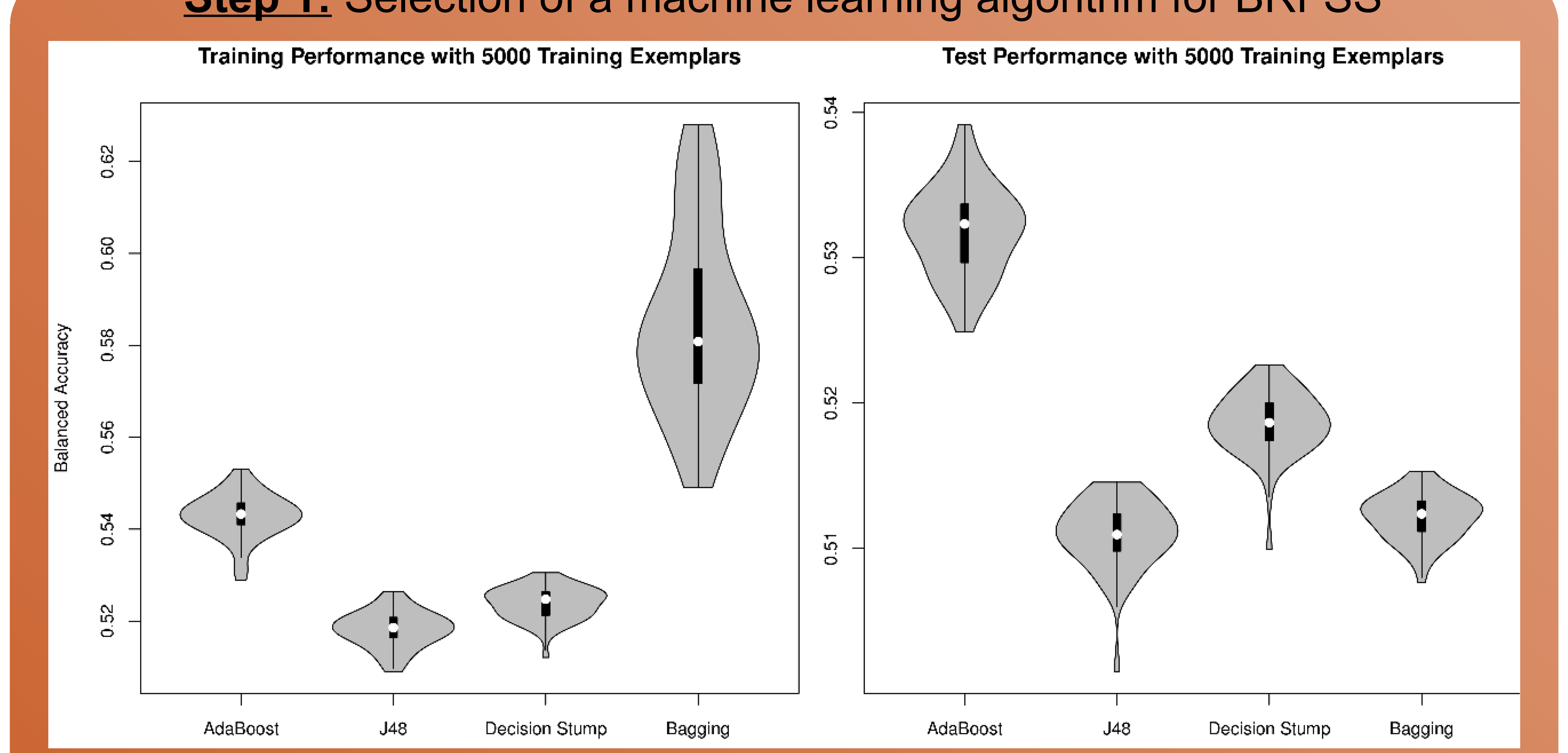


Result: AdaBoost classifier performance is highly tolerant to noise in data

Result Summary

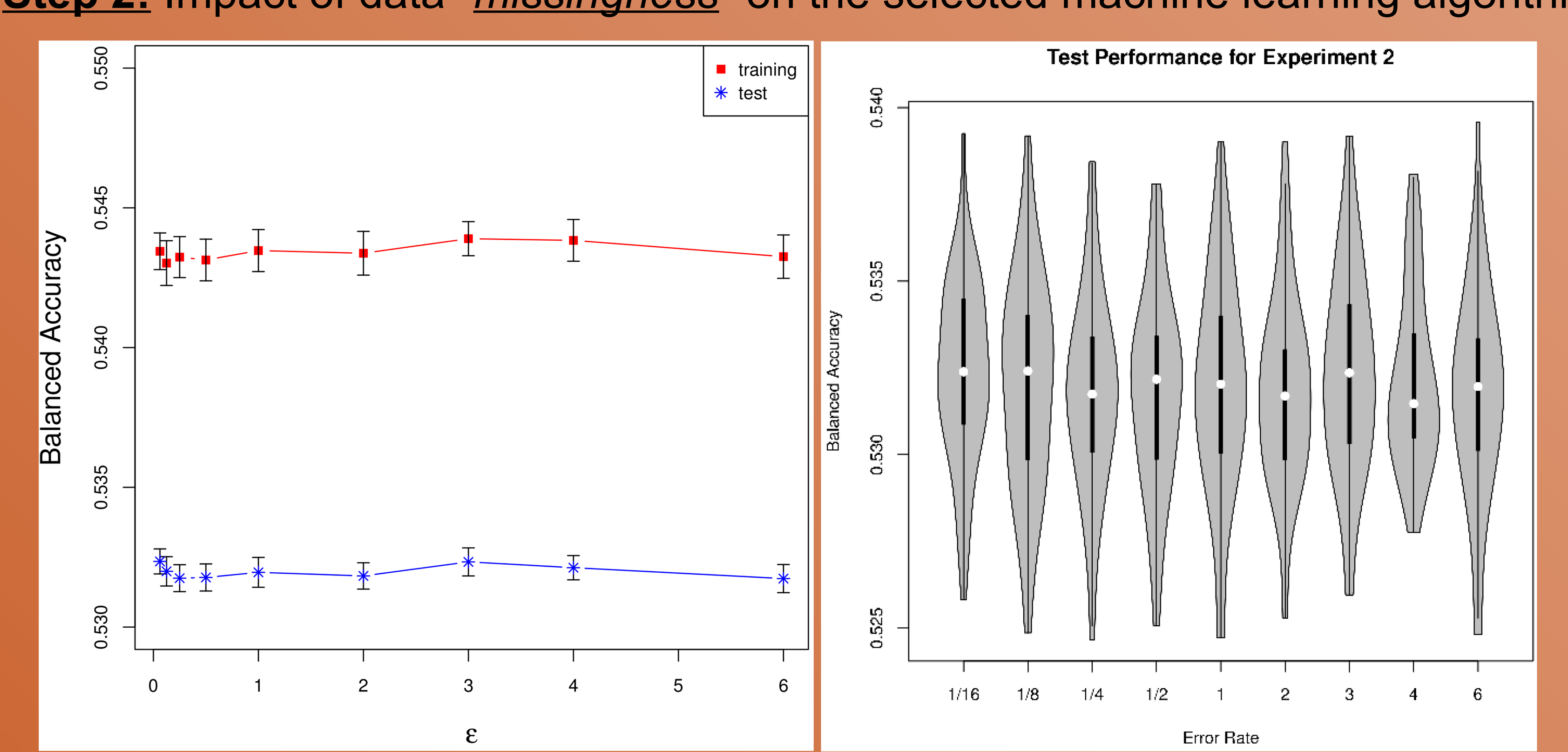
1. Machine learning techniques performed to poorly on their own to be used as the decision support engine.
2. Performance of the knowledge-based solution (OMeD) degrades quite rapidly as data *"missingness"* increases.
3. Ada-Boost based classifier is very resilient to data *"missingness".*
4. The hybrid construction of Holmes is noise resilient and performs better than both OMeD and the best machine learning classifier.

Step 1: Selection of a machine learning algorithm for BRFSS



Result: (a) in general machine learning techniques perform poorly.
(b) AdaBoost classifiers performs the best

Step 2: Impact of data *"missingness"* on the selected machine learning algorithm



Result: AdaBoost classifier performance is highly tolerant to noise in data