A Framework for Data and Mined-Knowledge Interoperability in Clinical Decision Support Systems

Reza Sherafat & Kamran Sartipi

Department of Computing and Software
McMaster University, Canada

Outline

- Introduction to Clinical Decision Support Systems
- Framework for interoperability of data and knowledge
- Semantic data interoperability
- Knowledge interoperability
- Knowledge interpretation
- Tool implementation
- Conclusion
Clinical Decision Support System (CDSS)

A computer program that assists health care professionals in decision making with generating patient-specific, correct, timely, and accurate recommendations, alerts, and reminders.

- Rule-based systems
  - Knowledge-base consists of a set of rules
  - Triggered rules perform actions

- Flow-oriented systems
  - Clinical best practices are modeled as flow charts
Flow-oriented Systems

- The clinical best practice is modeled as a flow chart that defines actions, state transitions, and events
  - Action steps
  - Decision steps
  - Branch steps
  - Synchronization steps
  - Patient state step

- Flows define the sequence of actions to be taken in any particular scenario (e.g. plan for treatment of cancer)

- **Decision nodes** determine the direction of flow by evaluating logical expressions or user input.
Motivations and Goals

- **Motivations**
  - Lack of methodologies or techniques to apply data mining results in CDSS
  - Heterogeneous healthcare environment

- **Goals**
  - Use well-defined methods to access data and knowledge
  - Disseminate and apply knowledge generated by data mining
  - Incorporate data and mined knowledge to be interpreted for clinical purposes
A Data Mining Model Example

- Melanoma Skin Cancer Classifier
  - “Rules for Melanoma Skin Cancer”, Duch et al.

- Dataset for analysis
  - Collected in the Outpatient Center of Dermatology in Rzeszów, Poland
  - Collection contained 290 cases

- Result of data analysis
  - A decision tree
  - Takes 6 input items
  - Produces one output
Input Data

- **Asymmetry** of the skin cancer mark
  - Symmetric spot
  - 1-axial asymmetry
  - 2-axial asymmetry

- **Border** of the skin cancer mark
  - Numerical values 0-8

- **Color** of the skin cancer mark
  - White, Blue, Black, Red, Light brown, Dark brown

- **Diversity** of the skin cancer mark
  - Pigment globules
  - Pigment dots
  - Branched strikes
  - Structureless areas
  - Pigment network

- **C-Blue**
  - Present
  - Absent

---

Reza Sherafat & Kamran Sartipi
Data Mining Model

Total Dermatoscopy Score (TDS) is first calculated from asymmetry, border, color, diversity

\[ TDS = 1.3 \times \text{Asymmetry} + 0.1 \times \text{Border} \]

- Input to the decision tree
  - TDS index
  - C-Blue

- Output of classifier
  - Melanoma Classification
    - Benign Melanoma
    - Blue Melanoma
    - Malignant Melanoma
    - Suspicious Melanoma

- The results as published in literature

R1: IF TDS ≤ 4.85 AND C-BLUE IS absent THEN MELANOMA IS Benign-nevus
R2: IF TDS ≤ 4.85 AND C-BLUE IS present THEN MELANOMA IS Blue-nevus
R3: IF TDS > 5.45 THEN MELANOMA IS Malignant
R4: IF TDS > 4.85 AND TDS < 5.45 THEN MELANOMA IS Suspicious
Environment for Interoperability of Data and Mined Knowledge

- Healthcare database is mined
  - Knowledge is extracted using different mining techniques
  - Knowledge is stored in knowledge-base

- The mined knowledge is encoded using PMML to be made portable

- The patient data resides in databases, e.g. EMR systems
  - Encoded using CDA to be made portable

- Logic Module
  - Accesses both data and knowledge
  - Interprets the knowledge

- Logic modules are executed within the context of a CDSS

As a result the system provides assistance for the healthcare professionals to improve their decision making.
**Data Interoperability**

**Data interoperability**: the ability of heterogeneous information systems to be integrated seamlessly to communicate and collaborate effectively as if they were a unified software system.

**Semantic data interoperability**: the ability of the sender of data to convey the full semantic and context in which each term receives a particular meaning, and the receiver's ability to interpret the message exactly with the meaning that it was encoded with.

- A shared data model is required
Clinical Document Architecture (CDA)

- CDA XML schema
  - Schemas are based on the CDA specification and the HL-7 RIM
  - Usually an schema is used for many purposes, e.g., medications administration, referral, patient history, discharge summary

- CDA instance document
  - Instances of the CDA schema
  - Terms are from standard or local terminology systems

- Data interoperability
  - Data source encodes the data in a CDA instance
  - At the destination the CDA instance document is parsed to extract the data

- CDA validation documents
  - Some constraints are not represented in the CDA schema.
  - Provides additional constraints used for ensuring consistency of the CDA instance document’s data items with the data mining data inputs requirements, numerical ranges, units of measurements, null values, etc.
Data Access

- Patient data is encoded according to the HL-7 data model
  - Data items are encoded in a CDA instance
    - Skin mark’s asymmetry, border, color, diversity, C-Blue
  - Data is encoded as **coded values** and associated with a **code system**
  - **XPath** is used to access the data items
    - An XPath is associated with the semantics of each data item that is its location in the CDA

```
<hl7:ClinicalDocument/hl7:component
<hl7:structuredBody/hl7:component
<hl7:section/hl7:entry[@typeCode='COMP']
<hl7:observation[@moodCode='EVN' and
@classCode='OBS']/hl7:code[@code='1234-3'
and
@codeSystem='2.16.840.1.113883.6.2']/..
<hl7:entryRelationship[@typeCode='COMP']
<hl7:observation[(@classCode='OBS')
and @moodCode='EVN')]}
<hl7:code/attribute::code
```
Data Constraints

- Verification of the data
  - Data items have to conform with the constraints required by the data mining model
  - CDA does not go that far, and CDA schema only specifies the structure of the documents

- Constraints:
  - Numerical data: valid ranges
  - Categorical data: valid values
  - Existence of required fields
  - Code system name
  - Units (e.g., cm vs. m)

- Schematron rules for validation
  - Assertions are XPath expressions that should evaluate as true
Knowledge Interoperability

Ability of a CDSS to incorporate and use the knowledge produced by other sources

- Predictive Model Markup Language (PMML) for encoding the mined knowledge:
  - XML based specification
  - Meta model defines the data structure of the model
  - Different types of data mining models (clustering, classifications, ...)
  - Extendable for model specific constructs
  - Self-describing

- Knowledge dissemination and application
  - PMML documents are used to share, access, exchange, and interpret the mined knowledge

- A logic module will interpret the encoded data mining model for individual patients and provides the results consisting of recommendations, alerts.
PMML encoding

- The data mining result is encoded in a PMML document (off-line)

PMML components:
- **Data Dictionary**: data input/output
- **Transformation Dictionary**: data transformations
- **Data mining model**: data structure of the data mining results, e.g., nodes of the decision tree, or frequent item sets of association rules mining
- **Extensions**: additional information for the output

```
<DataField>
  <Extension extender="CAS" name="Description" value="TDS is low and C-Blue is absent, so the result of classification is Benign-nevus."/>
</Extension>

<DataField>
  <Value displayValue="present" property="valid" value="present"/>
  <Value displayValue="absent" property="valid" value="absent"/>
</DataField>
```

```
<Data mining model="1.13 * Asymmetry + 0.5 * Σ Colors + 0.5 * Σ Diversities">
  <Distribution property="recordCount" value="62"/>
  <Node recordCount="59" score="Blue-nevus">
    <SimplePredicate operator="equal" value="present" field="C-BLUE"/>
    <ScoreDistribution recordCount="59" value="Benign-nevus"/>
  </Node>
</Data mining model>
```
Logic Module

- **Access data items**
  - Using the XPath expressions

- **Verify data**
  - Using the schematron document

- **Access knowledge (data mining results) and apply it**
  - The decision tree classifier

- **Encode results**
  - CDA results document

- Single interface to access the knowledge:

---

*Input (CDA)*

*Output (CDA)*

*Reza Sherafat & Kamran Sartipi*
Interpreting the Knowledge

- Interpret the knowledge
  - Apply the decision tree with patient data

- Provide the results
  - Malignant? Benign? ...

Blue Melanoma
Clinical Guidelines

- Flow-oriented Clinical Decision Support Systems
  - Best practices are represented as flow charts

- Decision nodes determine the direction of flow by evaluating logical expressions or user input

- Enhanced decision nodes interpret data mining models
  - Run the logic modules
  - Determine the flow direction based on the results
Process for Data and Knowledge Interoperability
Standards for Data and Knowledge Interoperability

- **Data**
  - HL-7 Reference Information Model (RIM)
    - A general high level health care data model
  - Clinical Document Architecture (CDA)
    - An XML-based standard for defining structured templates for clinical documents
  - Standard Terminology Systems (UMLS, SNOMED CT, ICD, etc)
    - Standard clinical vocabulary sets

- **Knowledge**
  - Predictive Model Markup Language (PMML)
    - An XML-based standard for representing data mining results
  - Guideline Interchange Format 3 (GLIF3)
    - A clinical guideline definition standard
Implementation

- We used **Protégé** ontology editor tool to define our CDSS.

- We implemented a clinical guideline execution environment as an extension plug-in in **Protégé**.

- An engine in the environment executes GLIF3 guidelines.

- Execution engine reads the model instances from the ontology
  - Traverses the flows
  - Manipulates local variables at each step
  - Parses XML documents using XPath
  - Executes logic modules
  - Interacts with the user

- Logic modules apply data mining models
  - Encapsulate the knowledge representation and interpretation
  - Deployed as web services
Conclusion

- We proposed a framework for data and knowledge interoperability within the health care systems that allows guided clinical decision making.

- We used data mining results as a source of knowledge to enhance clinical decision making.

- We used CDA for data interoperability and PMML for knowledge interoperability within heterogeneous health information systems.

- We developed a prototype tool as a proof of concept that uses clinical guidelines for CDSS and web services for data and knowledge transportation.

- We applied our approach on a case study from the literature.
Publications


A Framework for Data and Mined-Knowledge Interoperability in Clinical Decision Support Systems

Reza Sherafat & Kamran Sartipi

Department of Computing and Software
McMaster University, Canada