

# Scenario-Oriented Information Extraction from Electronic Health Records

Anis Yousefi, Negin Mastouri, Kamran Sartipi

Department of Computing and Software  
McMaster University, Canada  
{yousea2, mastoun, sartipi}@mcmaster.ca

## Summary

Providing a comprehensive set of relevant information at the point of care is crucial for making correct clinical decisions in a timely manner. Retrieval of scenario specific information from an extensive electronic health record (EHR) is a tedious, time consuming and error prone task. In this paper, we propose a model and a technique for extracting relevant clinical information with respect to the most probable diagnostic hypotheses in a clinical scenario. In the proposed technique, we first model the relationship between diseases, symptoms, signs and other clinical information as a graph and apply concept lattice analysis to extract all possible diagnostic hypotheses related to a specific scenario. Next, we identify relevant information items to these hypotheses and search for evidences in the patient's EHR using a weighting mechanism. Finally, we rank the items according to their relevancy to these hypotheses. We have assessed the usefulness of our approach in a clinical setting by modeling a challenging clinical problem as a case study.

## Motivation

- Problem:**
  - due to information overwhelming, it may be difficult for a physician to recall all influencing factors that they need to be investigated in order to make a correct diagnosis in a timely fashion
  - example: physician may forget the effect of a specific medication in causing fever
  - retrieval of scenario specific information from an extensive electronic health record (EHR) is a tedious, time consuming and error prone task
- Solution:**
  - we propose a model and a technique for mining relevant clinical information with respect to the most probable diagnostic hypotheses in a clinical scenario
  - This involves investigation of patient's EHR for evidences that *strengthen* or *weaken* the diagnostic hypotheses

## Definition of Terms

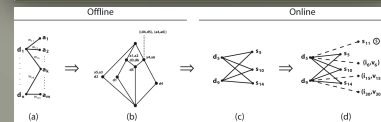
- Clinical Scenario:** a set of clinical observations (symptoms and signs)
- Diagnostic Hypothesis:** an explanation for the cause of observations (having a disease)
- Investigation:** search for evidences which influence the proof of a hypothesis, including:
  - Recognizing the influencing factors (relevant clinical information)
  - Searching EHR to Find evidences (supporting, weakening)
  - Discriminating hypotheses based on evidences
  - Proving a hypothesis
- Relevant Clinical Information:** information from patient's EHR which help in discriminating hypotheses

## Scenario-oriented Approach

The process of mining relevant information and discriminating diagnostic hypotheses involves the following steps:

- Off-line Processing:**
  - Modeling the relationship between the diseases in the domain and their attributes
  - Concept lattice analysis of diseases and their symptoms/signs to find maximal associations
- On-line Processing**
  - Extracting the list of diagnostic hypotheses for a specific patient scenario
  - Discovering the set of relevant information
  - Searching for evidences in patient's EHR
  - Discriminating the hypotheses and determining the 'relevancy' of information

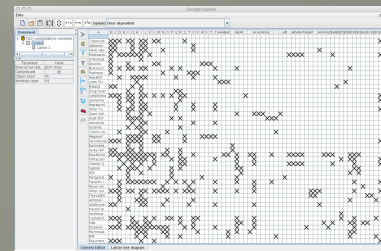
## Overview of the Proposed Approach



- (a)** Generic graph representation of diseases and their attributes (e.g., symptoms, signs, EHR elements, etc.)
  - $W_i$  is a quantity that we assign to an edge to indicate the support of attribute<sub>*i*</sub> in the diagnosis of disease<sub>*i*</sub>
- (b)** Concept lattice representation of a specific disease-attribute graph
- (c)** A maximal association among diseases and symptoms/signs
- (d)** Extending maximal associations in (c) with relevant attributes from EHR

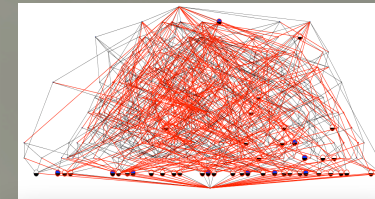
## Context Table

- Captures the relationship between diseases and their symptoms & signs
- We have employed **Concept Explorer** tool to generate and illustrate the concept lattice



## Concept Lattice

- 45 diseases and 64 common symptoms and signs for **Fever of Unknown Origin (FUO) syndrome**
- 499 concepts

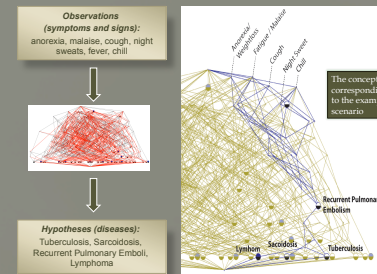


## Example Scenario

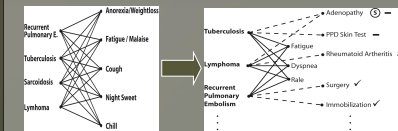
A 68-year-old Spanish female presented with *anorexia, malaise, non productive cough, night sweats, chill and daily fever (temperature, 38.3 C -39.5 C) from 4 days ago*. She recently moved to Canada and spoke English with difficulty and was not cooperative in giving a precise history. She was brought to clinic by her neighbor who was not aware of her past medical history, her medications and exposure or contact with animals or ill people. In her first physical examination, she was diagnosed *community acquired pneumonia* by family physician who prescribed antibiotic medication for her. Over the following weeks her fever persisted. Her medication was then switched to Clarithromycin for treating atypical pneumonia. There was no improvement in her condition. She was referred to specialist for further investigation of Fever of Unknown Origin (FUO).

The high-lighted terms represent symptoms and signs used to extract a specific concept from the generated concept lattice

## Discovering Probable Hypotheses



## Extending Maximal Association with Evidences from EHR



- (a)** Maximal association, extracted from concept lattice
- (b)** Extending maximal association with evidences from EHR
  - S Symptoms/Signs from EHR
  - Attribute not found in EHR
  - \* Attribute found in EHR; The value of attribute in EHR does not match disease-attribute graph;
  - ✓ Attribute found in EHR; The value of attribute in EHR matches disease-attribute graph;

## Pulmonary Embolism (strong)

Degree of relevancy indicates how relevant information is to a hypothesis:

$$W_{ij} = W_{ij1} \times W_{ij2} \times W_{ij3}$$

The likelihood of a hypothesis in a scenario is calculated by aggregating the weights of its supportive evidences

$$w_i = \sum_{j=1}^m w_{ij}$$

There are thresholds to translate quantitative values to qualitative terms: *strong, medium-strong, medium, mild-medium, mild*

Hypothesis	Supporting Evidences	Weakening Evidence	Unfound Relevant Information
Pulmonary Embolism (strong)	Old(strong), Diabetes(strong), Sarcary(strong), Immobilization (strong)	Normal ECG(mild), Normal CRP (medium)	D-Dimer test(strong), Lung perfusion Scan (strong), Pleuritic chest pain(strong)
Lymphoma (medium-strong)	-	-	-
Tuberculosis (medium)	-	-	-
Sarcoidosis (medium)	-	-	-
Sarcoidosis (medium)	-	-	-

## Conclusion

We proposed a solution to the challenges in retrieving relevant information from the patient's EHR. We believe that more intelligent retrieval systems are needed to provide different views of the EHR information for different purposes. We presented a scenario-oriented view as an intelligent extraction of relevant and useful information to serve a more effective diagnosis procedure. The proposed approach is not intended to interfere with the clinicians' diagnostics of patients' health problems, however it assists in extracting relevant information that are necessary and useful for diagnostics and decision making. We assessed the usefulness of our approach in a clinical setting by modeling a challenging diagnostic problem.