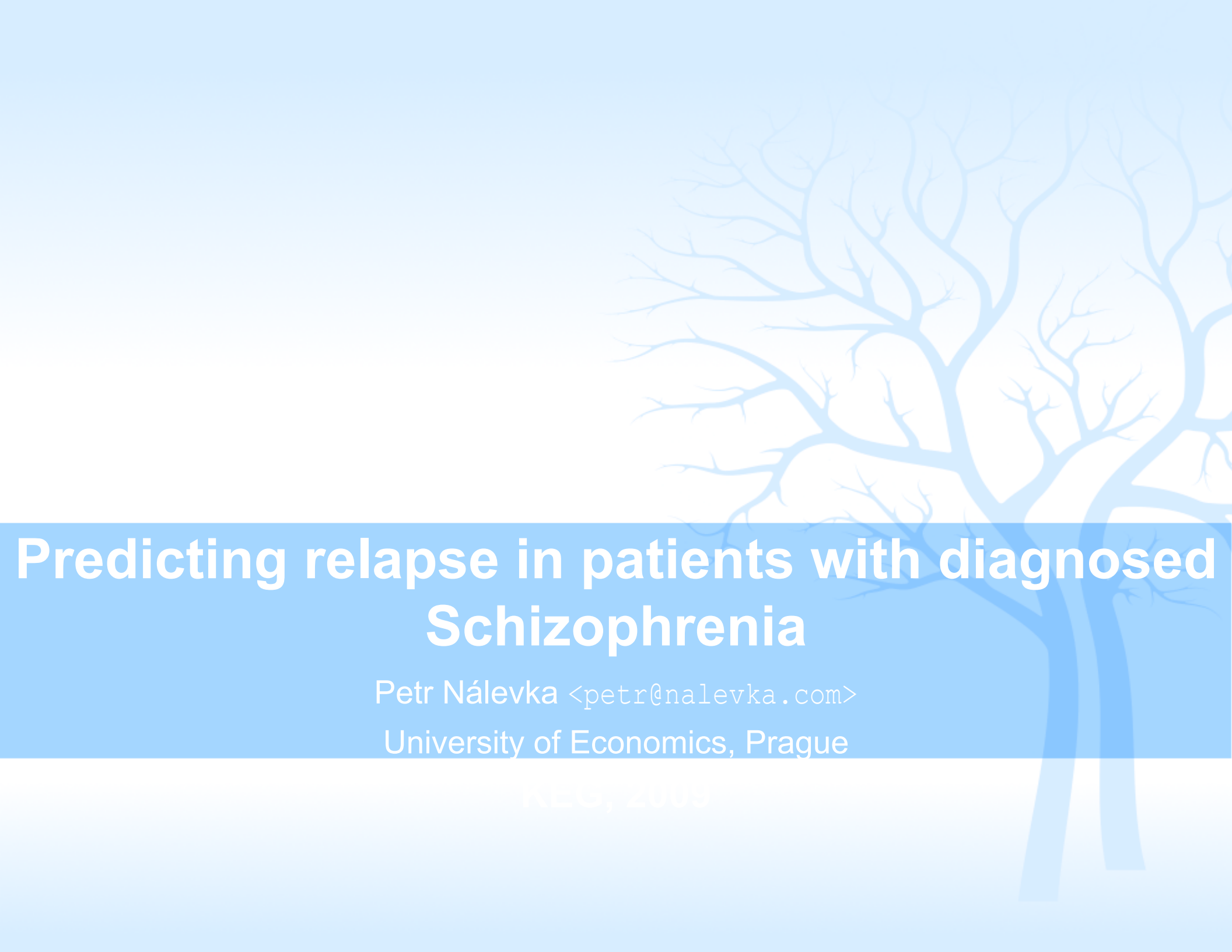
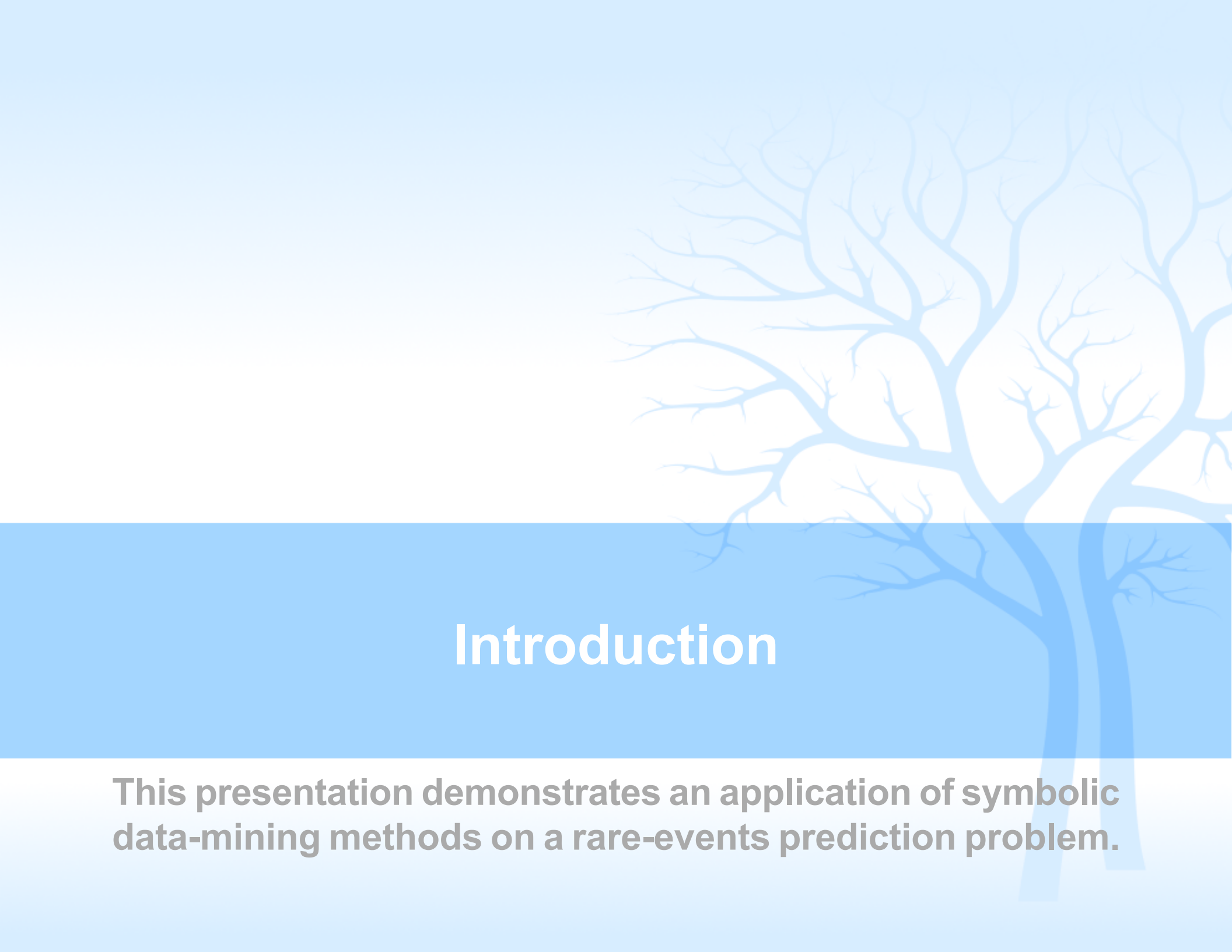


Predicting relapse in patients with diagnosed Schizophrenia

Petr Nálevka <petr@nalevka.com>

University of Economics, Prague

KEG, 2009



Introduction

This presentation demonstrates an application of symbolic data-mining methods on a rare-events prediction problem.

Schizophrenia

Schizophrenia is a persistent, often chronic and usually serious mental disorder affecting a variety of aspects of behavior, thinking, and emotions. It may be accompanied with perception distortions affecting all five senses, including sight, hearing, taste, smell and touch, but most commonly manifest as auditory hallucinations, paranoid or bizarre delusions, or disorganized speech and thinking with significant social or occupational dysfunction.

Relapse of Schizophrenia

- 50% - 80% incidence of relapse without treatment
- 40% incidence of relapse with anti-psychotic medication
- Non-psychotic prodromal symptoms
  - irritation, sleep problems, tense, fear, anxiety, quietness or withdrawal...
- Psychotic prodromal symptoms
  - Similar to schizophrenia symptoms but milder
- Timely detection of prodromal symptoms — relapse prediction and prevention

ITAREPS

Information Technology Aided Relapse Prevention in Schizophrenia

ITAREPS

Aim — reducing the number of illness exacerbations and subsequent psychiatric hospitalizations, improving and speed up communication between the patient and his/her outpatient psychiatrist. As a result, the program can improve quality of life of patients with schizophrenia.

How it Works

- Patient, carer, psychiatrist
- 10 questions each week
- Targeting prodromal symptoms, evaluating state 0-4
- Intuitively designed threshold mechanism — alerts
- Medication

Data

- Patient's personal data, questionnaire answers, alert history, hospitalization history, relapse history and medication usage
- Patient's data
  - Demographic and diagnostic information
  - Study branch, sex, age, marital status, education, diagnosis, family diagnosis, information about carer and relation to carer, other diagnostics measures such as CGI, GAF, Hayward, symptoms (positive, negative, cognitive disorganization, mood), medication information (dosage, usage, frequency) etc...
- Questionnaire
  - Carer/Patient, date, arrival, answers
- Hospitalization and Relapse
  - Patient, date, duration, reason

Rare Events

Knowledge and Prediction

Objectives of this Research

- Employ machine learning and data-mining methods to better understand the data
- More sophisticated alternative to the simple threshold method
- More accurate and timely prediction to improve overall efficiency of the program

Symbolic Method

- Not only predict but understand
- Unique data — may bring new knowledge about Schizophrenia
- Understandable, explained to doctors

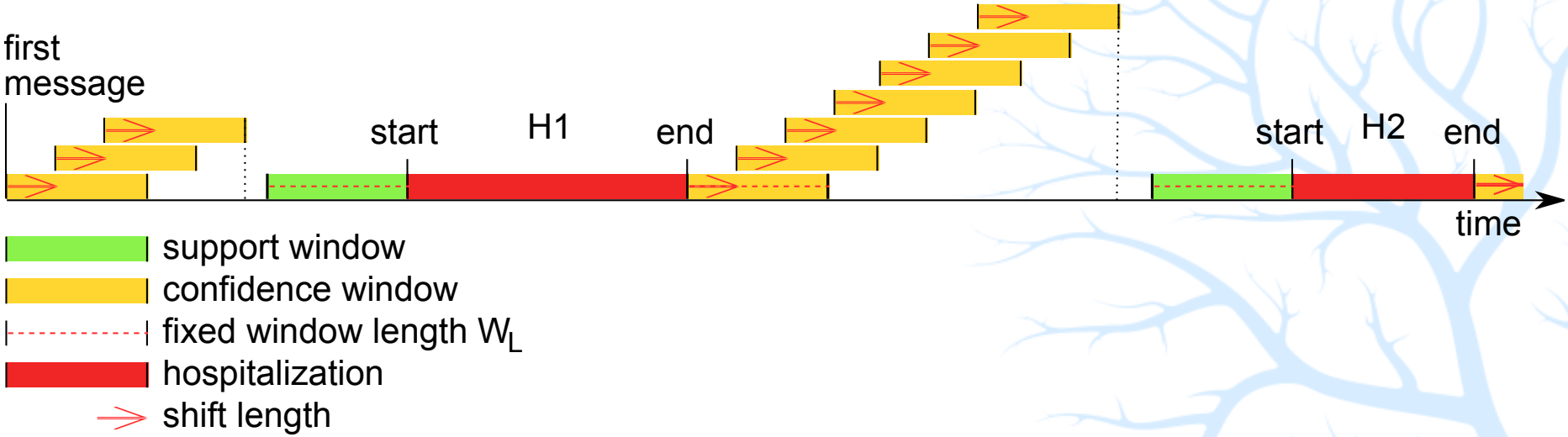
Specific Data-mining Task

- Temporal domain
  - Distinguishing between time periods preceding re-lapse and others
- Rare events
  - Relapse is rare - (57 positive events)
- Difficulty
  - Only few examples in data support the target class
  - Temporal nature of events may be difficult to describe
  - Different related temporal events may be difficult to match

Rare events

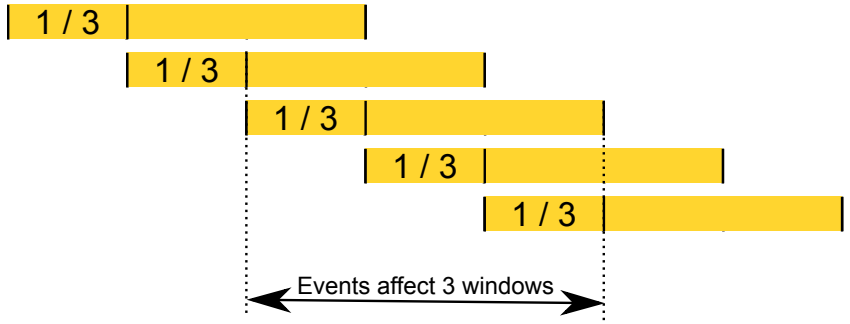
- e.g. prediction of natural disasters, fraudulent transactions in financial institutions or all sort of different system failures.
- ICT systems failures - status events in time
- Critical factors
  - questionnaire answers ability to describe patient's state
  - answer consistency between patients
  - influence of the disease on the answers
  - cooperation, misunderstandings etc...

Temporal Windows

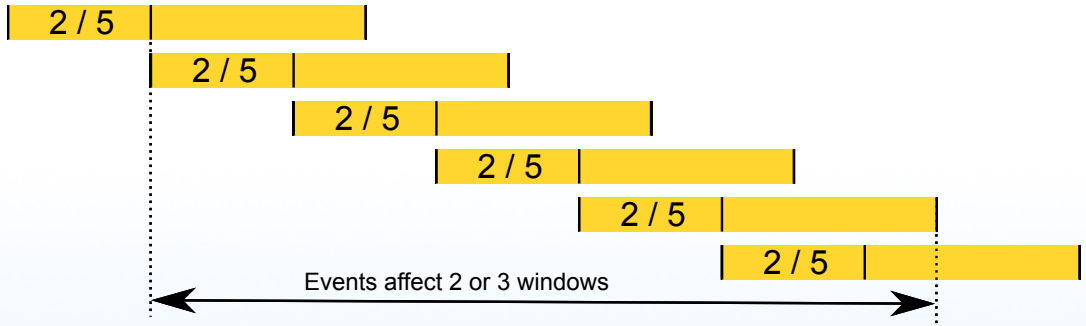


- Split events into fixed size temporal windows
- Separate windows into positive and negative

Negative Window Overlap



1 / shift length **has to be an integer to make most events influence the same number of windows**



Mining on Windows

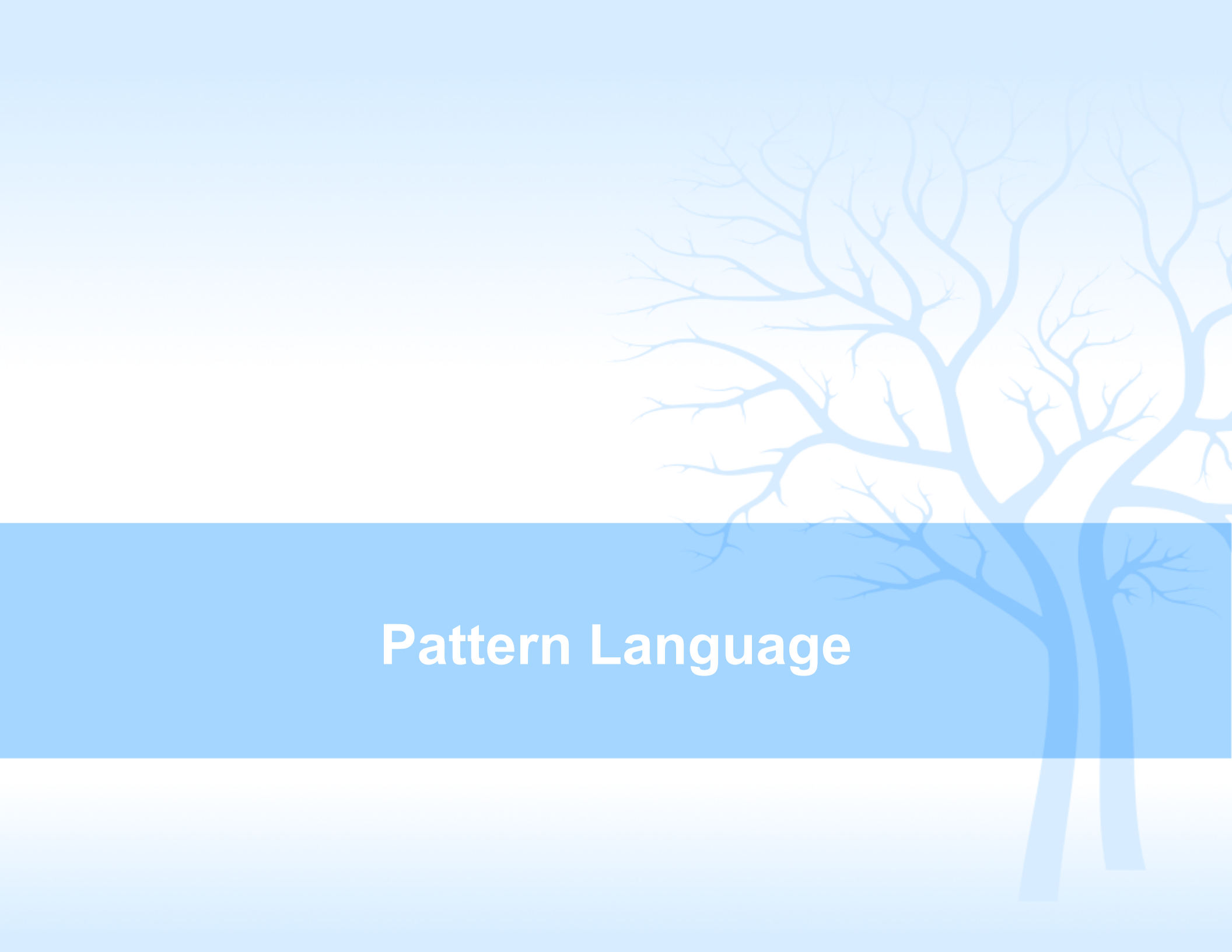
- Generate patterns describing about windows and evaluate them
- Success in describing positive windows
- Evaluation - one measure is not sufficient
  - Literature: modified support and confidence, precision and recall

Sensitivity and Specificity

		Condition	
		Positive	Negative
Test	Positive	TP	FP
	Negative	FN	TN
		Sensitivity - TP / (TP + FN)	Specificity - TN / (FP + TN)

Sensitivity and Specificity

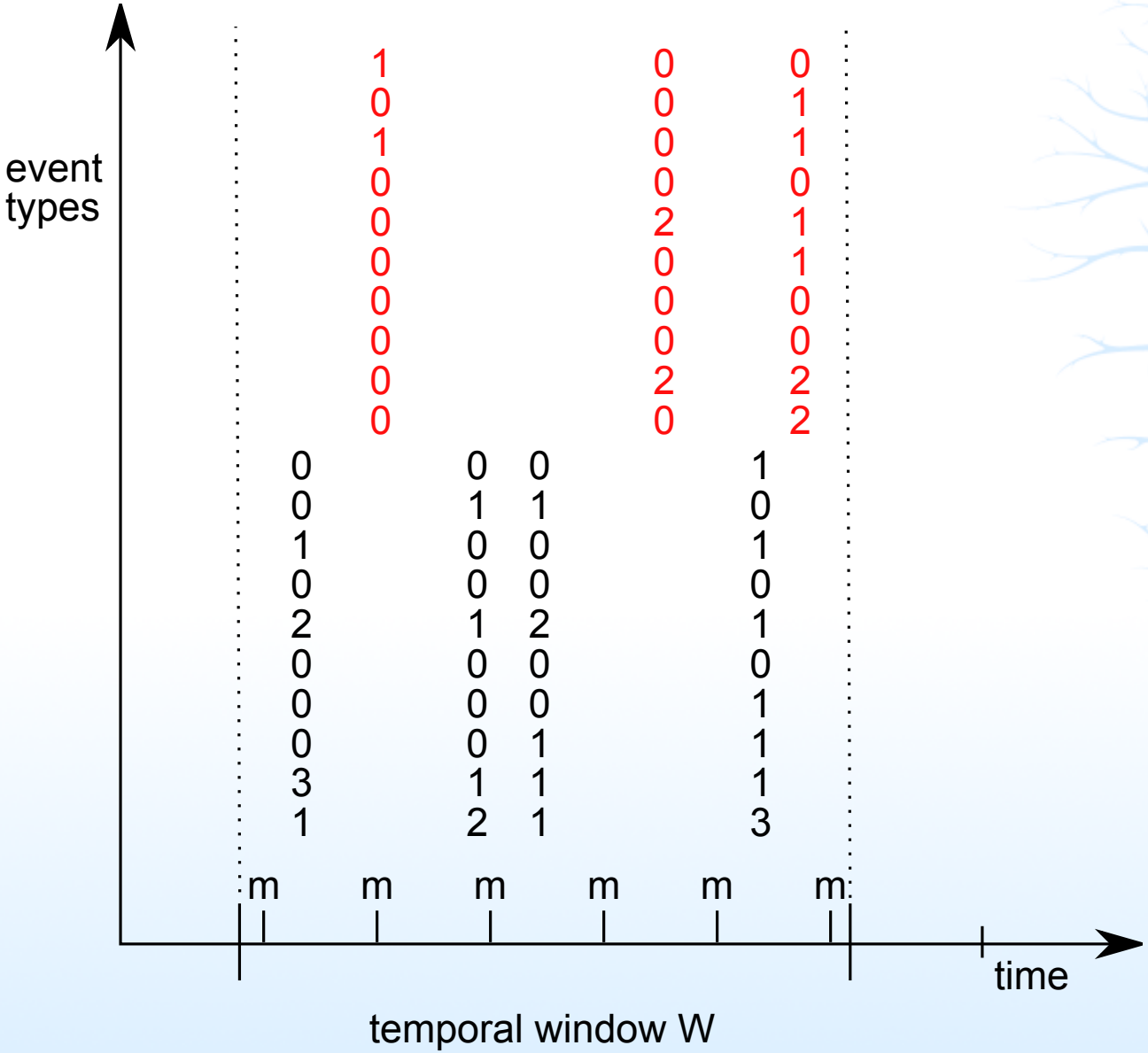
- Fights rare events problem - "Only few examples in data support the target class"
- Straightforward interpretation
  - Sensitivity - coverage of diagnosis
  - Specificity - error of diagnosis
  - Even in each window - Sensitivity = 1, Specificity = 0
- Popular in medicine domain (understandable to doctors)


Pattern Language

Pattern Language

- Interpretability, descriptive power, performance
- Literature: event sets, event grammar (wildcard '*', next '.' and '|' any order)
- Specifically designed pattern language

Windows Events



Window Development

- Characterizing window as a whole
  - Per window: average, standard deviation, trend, co-operation
  - Per question: average, standard deviation, trend
- No issues with event matching
- Still persists event position - trend
- Discretization into enumerations - zero, low, moderate, high, extreme...

Specific Context Operators

Standalone assertions e. g. `average(high)` may be wrapped by context operators.

Example

- more than 70% of answers have `trend(worsening)`
- window has `cooperation(good)`
- question number 4 has `average(extreme)`

Conjunction and Disjunction

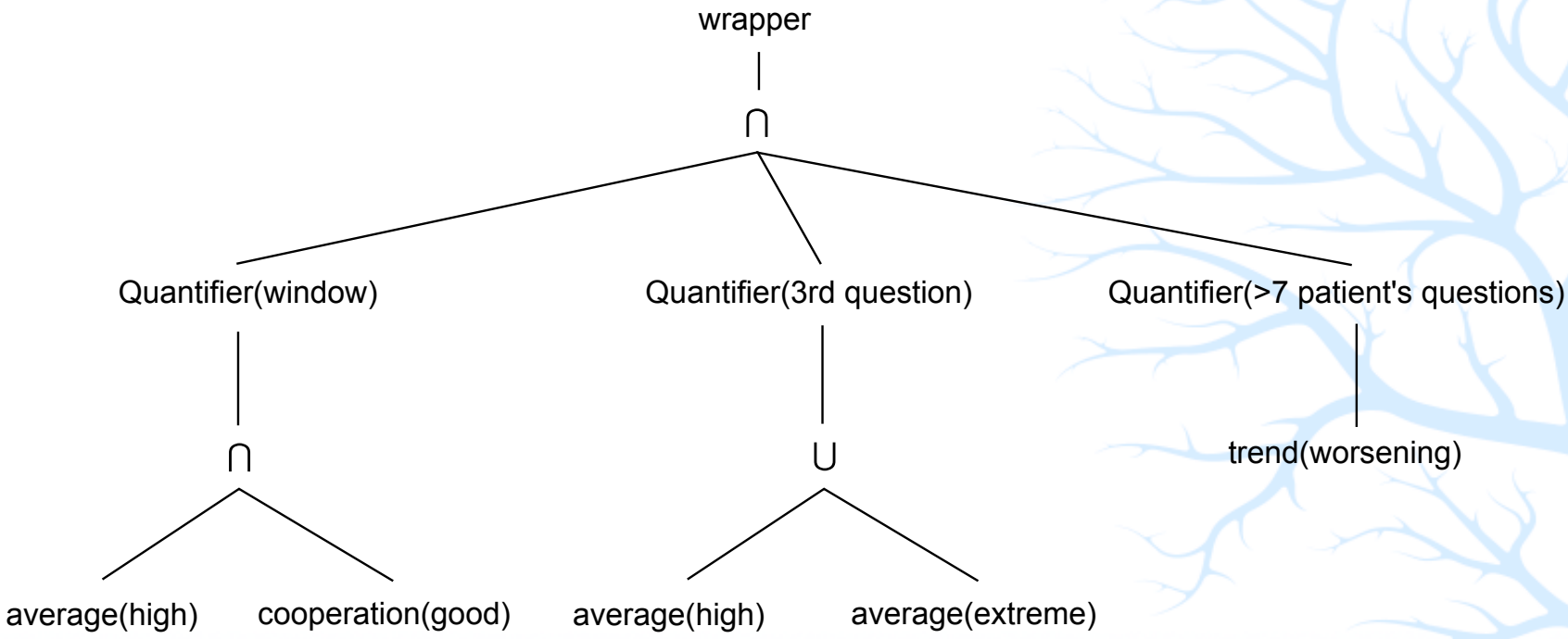
Combine primitive assertions into composite assertions through *Conjunction*

Making primitive assertions softer through *Disjunction*

Enumeration Parameters

- Nominal / ordinal
- Min / max join

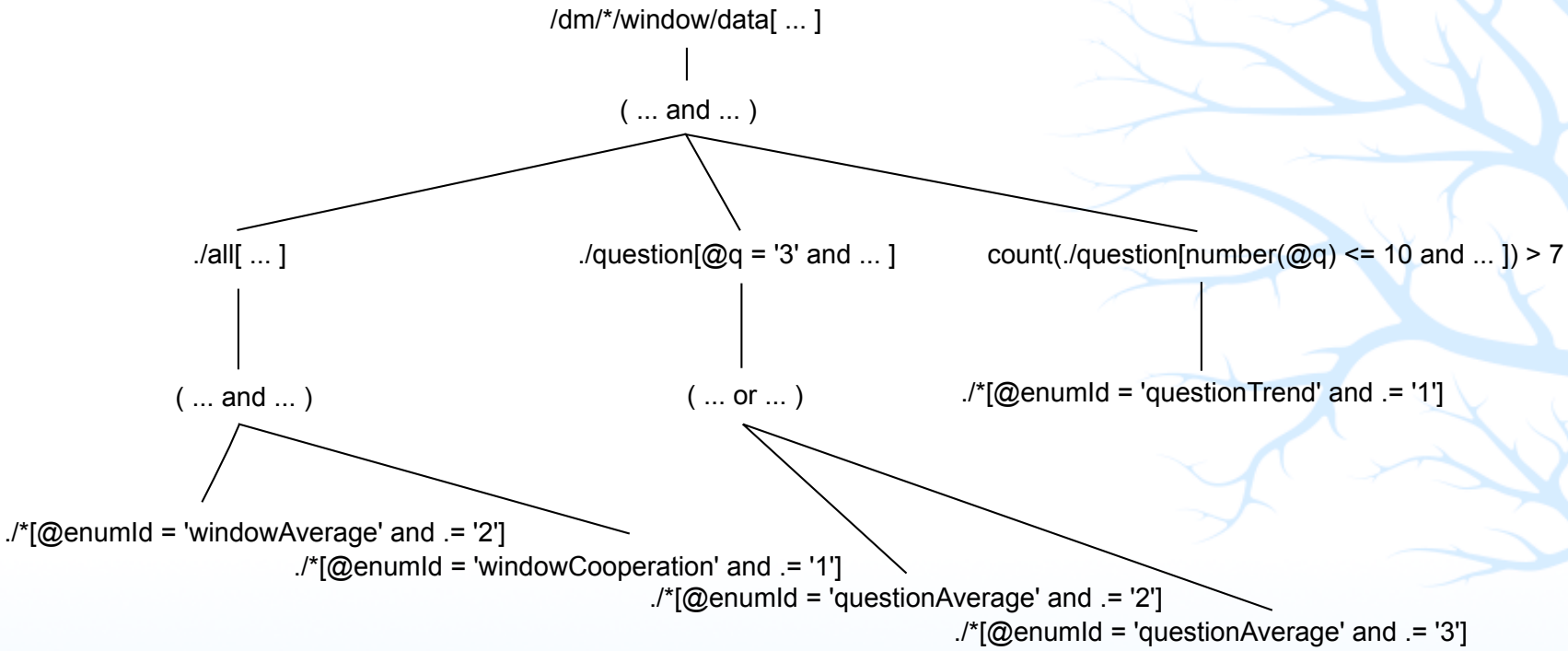
Conjunction and Disjunction



Implementation: Java object tree

Evaluating Assertions

XPath Query



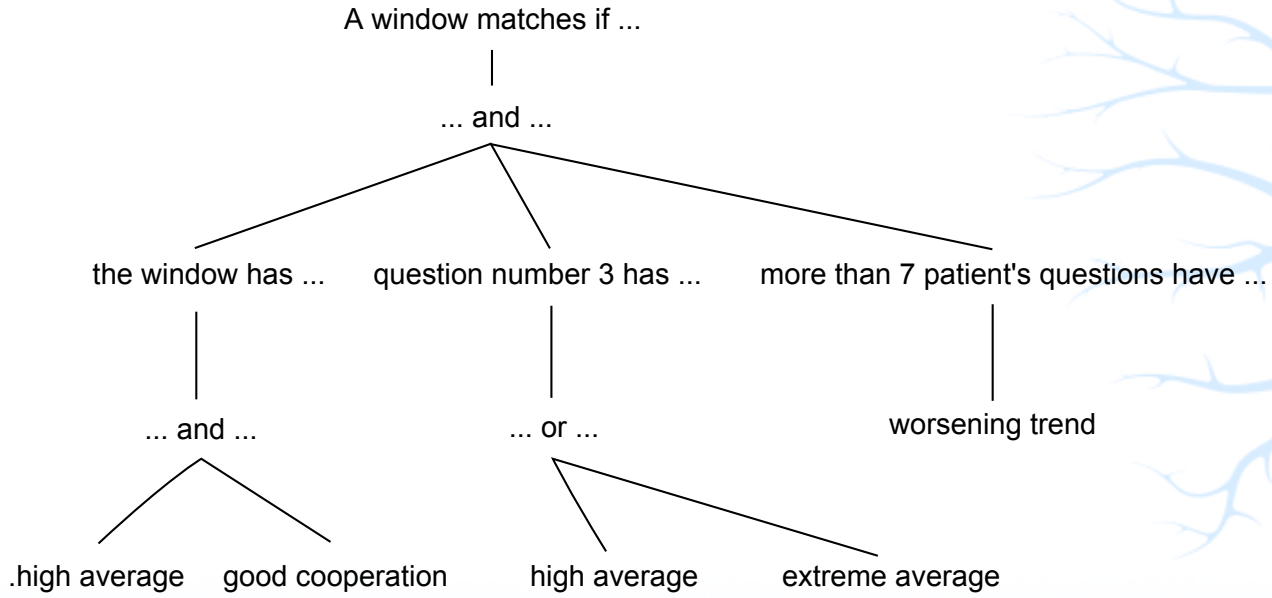

```
/dm/*/window/data[( ./all[( /*[@enumId = 'windowAver-
age' and .= '2'] and ./*[@enumId = 'windowCooperation'
and .= '1' )] ] ./question[@q = '3' and ( ./*[@enumId
= 'questionAverage' and .= '2'] or ./*[@enumId = 'ques-
tionAver- age' and .= '3'] )] and count(./question[num-
```

Evaluating Assertions

```
ber(@q <= 10 and ./*[@enumId = 'ques- tionTrend' and .=
'1'])]) )]
```

Explaining Assertion

Human understandable explanations.

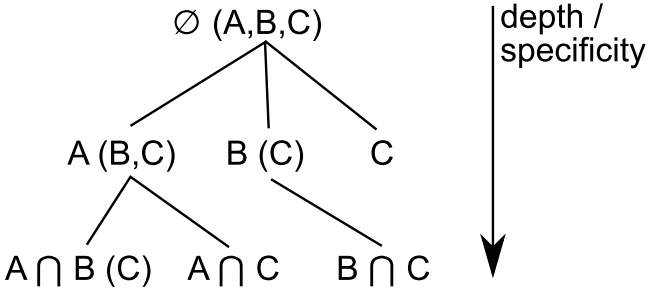


A window matches if window has high average and good cooperation and question 3 has high average or extreme average and more than 7 patient's questions have worsening trend.

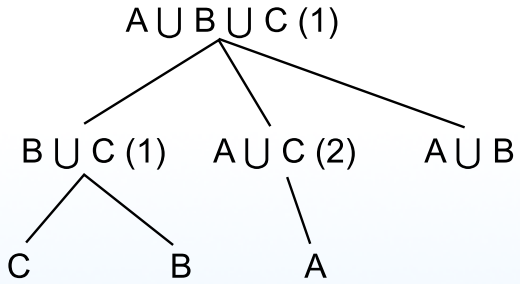
Search

Generating Patterns

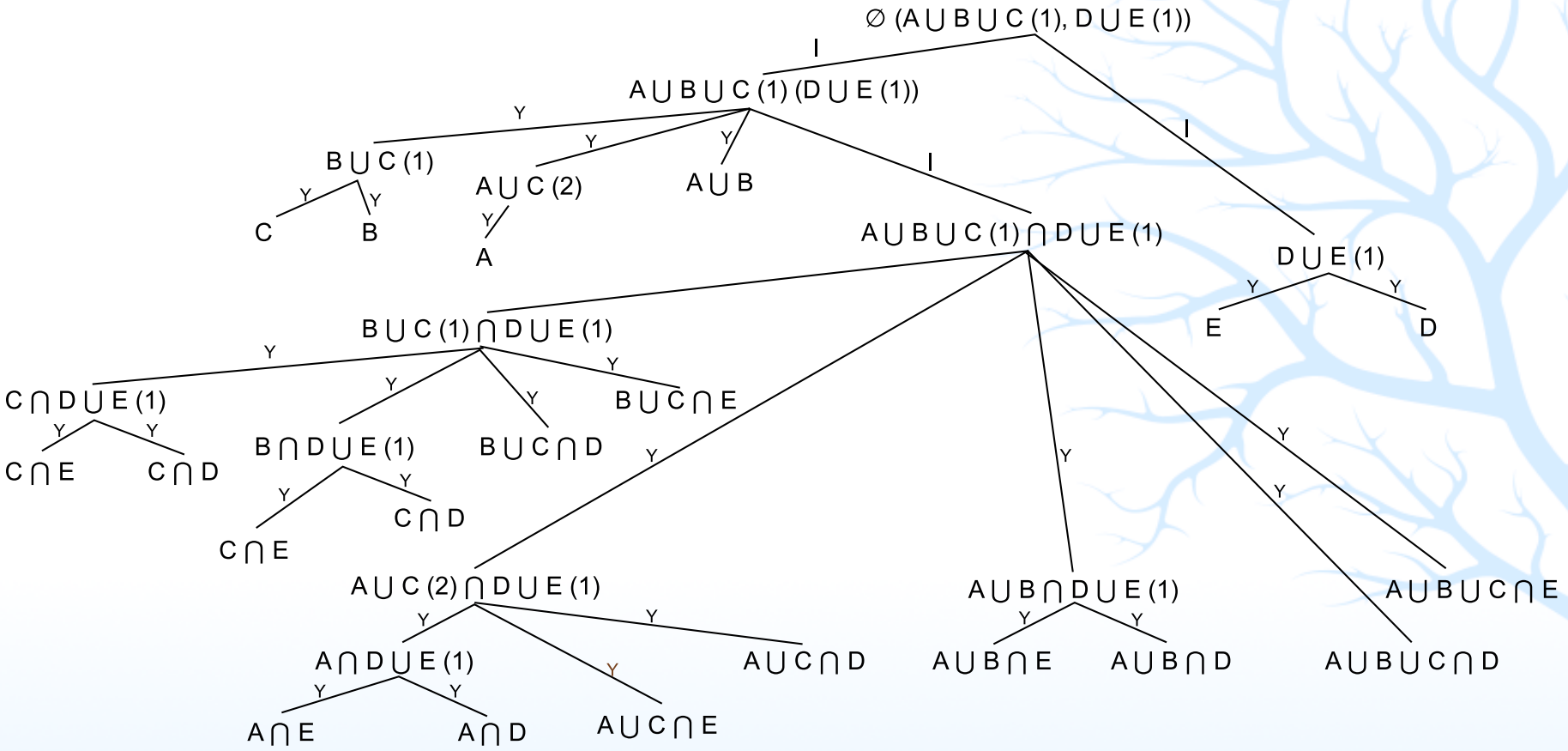
Combining context assertions into composite assertions using AND



Making enumerations less sharp.



Generation Pattern Space



Combination Space - moving from parent to child = □ Specificity, □ Sensitivity

Search

- Combination Space — oriented graph, edges *being more specific*
- Size is determined by available enumerations, their size and parameters and by available context operators
- Sensitivity / Specificity Threshold
- Evaluating Sensitivity for each node - branch cut off
- Than evaluating Specificity (more performance consuming)
- Maximum assertion length
- Maximum nodes

Resources

- Use maximum CPU with constant memory usage
- Fixed size FIFO queue
- Thread Pool
- Sensitivity and Specificity Task
- Breath First
  - If queue is full, process child nodes within current thread
- Depth First
  - If queue is full, poll queue and put children at the end of the queue, process polled tasks within current thread

Two Phase Search

Inconsistency in the way patients answer questionnaires - difficult reach Sensitivity / Specificity thresholds.

Dividing patients into groups based on their demographic and diagnostic data may establish groups with more consistent way of answering.

- Phase 1
  - Searching patient state space
  - Criteria: minimum positive examples
- Phase 2
  - Searching for interesting temporal developments within the patient groups

Cross Validation

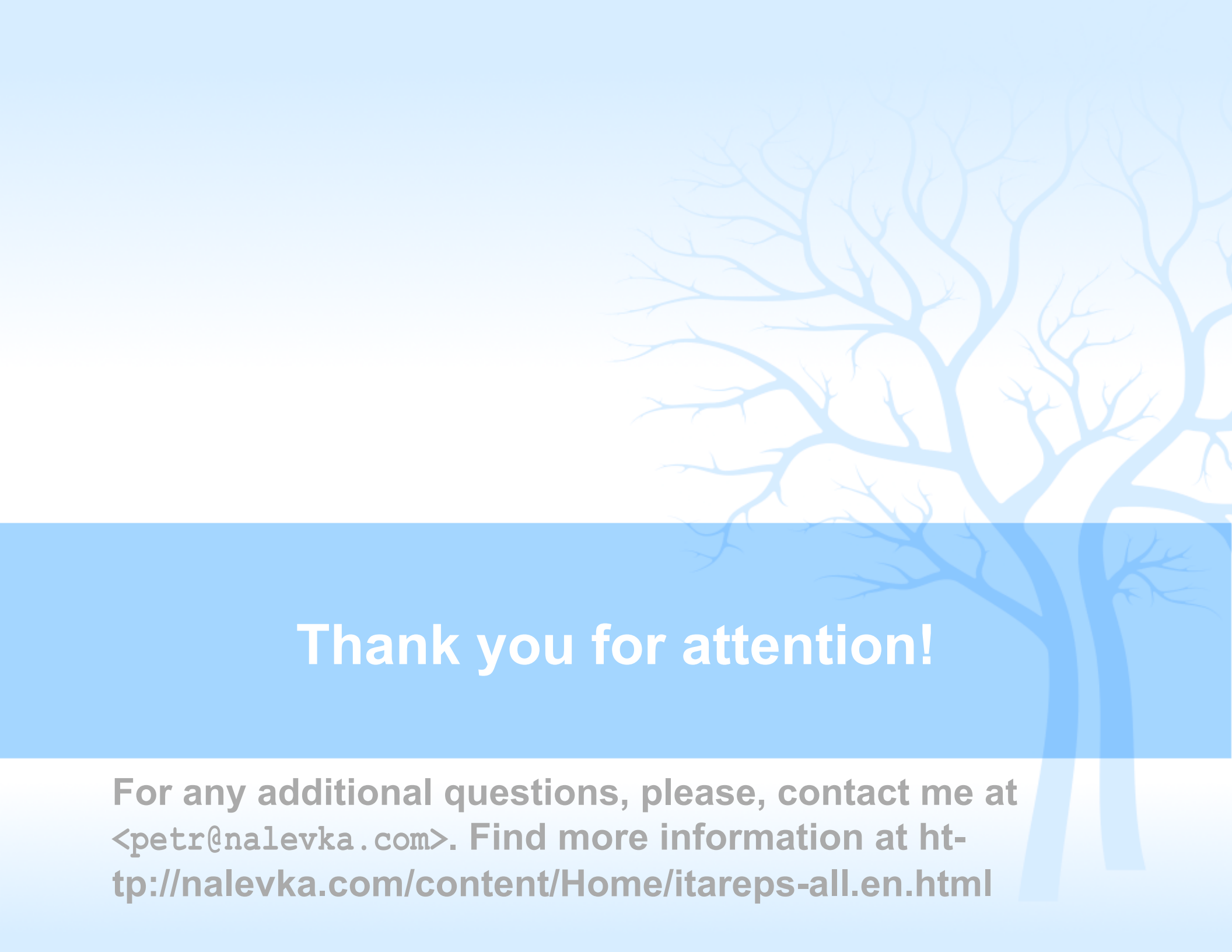
- Dividing each patient group into `X parts`.
- For each `Y` in `X`, `Y` is used as an *validation set* and the rest as the *training set*.
- Patterns are discovered in each *training set* and validated against the *validation set*.
- Suitability of patterns is again determined using Sensitivity and Specificity.

Results

Results

High Sensitivity and Specificity during search ≠ Suitability for prediction

- Even for heterogeneous patient groups - patterns with Sensitivity > 70, Specificity > 85.
- But cross validation is successful (Sensitivity + Specificity > 130) only for very specific groups
- Cross validation Sensitivity is very low < 30.



Thank you for attention!

For any additional questions, please, contact me at <petr@nalevka.com>. Find more information at http://nalevka.com/content/Home/itareps-all.en.html