Data $\rightarrow$ [Machine learning] $\rightarrow$ Extracted Knowledge

Varun Ojha, Ph.D.
Postdoctoral Researcher.

ETH Zurich, Zurich 22.02.2017
Predictive Analysis – I / IV

Neural Tree: A simplified and advanced variant of Neural Networks
- Adaptive Predictive Model
- Low Complexity Models
- Dimensionally Reduction (Feature Selection and Feature Analysis).
- Application: Time-series, binary classification, regression

Fig. Typical Neural Tree
Predictive Analysis – II / IV

**Fuzzy Tree**: An advanced variant of Fuzzy Inference Systems
- Adaptive Inferential Predictive Model
- Low Complexity Models
- Dimensionally Reduction (Feature Selection and Feature Analysis).
- Application: Time-series, binary classification, regression

*Fig. Hierarchical Fuzzy Inference System*
Type-2 Fuzzy Inference System: For rule based predictive models

- IF X is A AND Y is B THEN Z is C
- C is a fuzzy set in Mamdani-FIS and a function in TSK-FIS
- Application: predictive analysis, multi attribute decision making

Fig A. Metaheuristic Optimization of Type-2 Fuzzy Inference System
Fig B. Type -1 Fuzzy Set
Fig C. Type -2 Fuzzy Set
Descriptive Analysis – I / V

Pattern Analysis and Clustering:
- Human Activity Recognition
- Hesitant Fuzzy Inference
- Associative Rule Mining
- Anomaly Detection
- Self Organizing Map (SOM)
- Evolutionary Feature Selection

Image Processing:
- Deep Learning: Convolutional Neural Network (CNN)
- Image Recognition (E.g., face, emotion, fingerprint detection)
- Image Pattern Matching
- Image Segmentation
Human Activity Recognition: For analyzing human activities based on the data collected from the smartphone, smartwatch, and smart-wear.

- Temporal Eco-State Recurrent Network was trained to classify activities.
Descriptive Analysis – III / V

Hesitant Fuzzy Inference System: For Multi Attribute/Criteria Decision Making

1. Determining multiple criteria for model evaluation (e.g., height, location, color, look, etc.)
2. Assign initial weights to criteria based on experts' knowledge
3. Assign linguistic terms (very poor, poor, average, good, very good) to each criteria.
4. Compute hesitant fuzzy set for each model and each criteria.
5. Optimize weights assigned to each criteria
6. Compute score to each models

- Example answer: B < D < A < C

Fig. Four models (symbolic) for experts evaluation and ratings.
**Associative Rule Mining**: To find frequent pattern (rule) in dataset.

- **IF** someone buy an item X **THEN** what is the possibility that the person will by the item Y
  - So compute $X \rightarrow Y$ (Support, Confidence)
  - Support: The probability that transaction contain both X and Y
  - Confidence: The conditional probability that the transaction containing X also contain Y.

- For minimum support **50%**
  - Frequent 1 item-sets: Beer: 3, Nuts: 3, Apple: 4, Eggs: 3,
  - Frequent 2 item-sets: {Beer, Apple}: 3

- For minimum confidence **50%**, the association rules are:
  - Beer $\rightarrow$ Apple (60%, 100%)
  - Apple $\rightarrow$ Beer (60%, 75%)
Descriptive Analysis – V / V

Anomaly Detection: For detecting abrupt behavior of a system. Also known as outlier detection.

1. Determine positive functioning of target sensor (variables)
2. Label input vectors as positive or negative example
3. Train a model with all positive examples
4. Cross validated to determine a minimum threshold (say E)
5. Computer probability of a test vector: \( P(\text{test vector}) \)
6. \( \text{IF } P(\text{test vector}) < E \text{ THEN } \) the test vector is anomalous.

Fig. Boeing-747 has about 6 million parts, sensors, and functional units.
ESUM: Analyzing trade-offs between the Energy and Social performance of Urban Morphologies
ESUM – Data and its analysis steps

Data Analysis Steps Performed/Planned:
1. Automatic clustering using Self Organizing Map:
2. Neural Tree or NN: To find if the participants behavior can be predicted. To perform Feature Selection and Analysis.
3. Fuzzy Tree or Fuzzy Inference Models: To find inferential (rule based) predictive models to understand if there are certain rule pattern emerge.
4. Evolutionary Feature Selection of Isovit Features
5. Associated Rule Mining: To find pattern in the relationship
6. To model survey questions as classification and fuzzy modeling.

Fig F. Data dependencies in the ESUM project (illustration)
Automatic clustering of participants data using SOM

Fig. SOM clustering map of participants (indicated by numbers)

Fig. Changing of participants behavior – biofeedback responses
ESUM – Predictive modeling (Non-inferential)

Assumption: Environment has direct influence on participants behavior.

Question: Can we predict the citizen's behavior based on certain environmental condition?

Case 1: Classification based predictive modeling results:

Hypothesis: If the environment has direct influence on the participants and the participants reacted differently the test classification accuracy will be high.

Case 2: Regression based predictive modeling results:

Hypothesis: If the environment has direct influence on the participants then the test correlation will be high.

Assumption: Environment has direct influence on participants behavior.

Question: Can we predict the citizen's behavior based on certain environmental condition?

Fig A. Classification results of 10-fold cross validation

Fig B. Correlation coefficients of 10-fold cross validation
ESUM – Feature Analysis

**Question:** What are the significant factors influencing a citizen’s behavior?

**Case 1: Classification based feature analysis results.**

**Hypothesis:** If the environment has direct influence on the participants, then the environment features should emerge prominent (high score) in feature analysis.

<table>
<thead>
<tr>
<th>Selected Feature</th>
<th>Discarded Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dust</td>
<td>Sound</td>
</tr>
<tr>
<td>Temp. Environment</td>
<td>Blood Volume Pressure (BVP)</td>
</tr>
<tr>
<td>Relative Humidity</td>
<td>Light</td>
</tr>
<tr>
<td>Heart Rate</td>
<td>Heart Rate</td>
</tr>
<tr>
<td>Electrodermal Activity</td>
<td>Electrodermal Activity</td>
</tr>
<tr>
<td>Temp Biofeedback</td>
<td>Temp Biofeedback</td>
</tr>
</tbody>
</table>

**Case 2: Regression based feature analysis results.**

**Hypothesis:** If the results of classification based feature analysis finds that the environment has direct influence on the participants, then to find which factors were significant towards measurements of HR, BVP, EDA, and TempBF is a significant step.

<table>
<thead>
<tr>
<th></th>
<th>Hear Rate</th>
<th>BVP</th>
<th>EDA</th>
<th>Temp BF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sound</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dust</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temp. Environment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative Humidity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Light</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ESUM – Predicting emotion of the citizens (Positive or Negative Emotion)

**Question:** Do the citizen’s exhibits positive or negative emotion towards certain environment?

**Case 1: Explicit response:**

<table>
<thead>
<tr>
<th>Response</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>65.02%</td>
</tr>
<tr>
<td>Neutral</td>
<td>05.99%</td>
</tr>
<tr>
<td>Negative</td>
<td>28.97%</td>
</tr>
</tbody>
</table>

**Emotion classification accuracy:**

- With only environment variables: 85.5%
- With only biofeedback variables: 92.25%

**Case 2: Implicit response:**

**Stress (feeling) as per EDA data**

- **Positive:** 81.37% (low stress)
- **Negative:** 18.63%

**Fact:** A certain change (a peak) in EDA response causes higher stress (negative emotion)
ESUM – Predictive modeling (Inferential)

Assumption: Environment has direct influence on participants behavior.

Question: Can we able to infer a detailed relationship between a citizens behavior and a city environment?

Fuzzy Inference model for emotional changes (specific study only)

Experiment: training a fuzzy rule learner (only for participants #1.)

Inputs attributes: Sound, Dust, Temp EN, RH.

Output attribute: Emotion (Positive Negative, Neutral)

# of Samples: 667

Rule 1: IF Temp EN ≥ 13.08 and Light ≥ 1549.15 THEN Emotion is Negative (Cf = 0.83)

Rule 2: IF RH ≥ 46.902667 THEN Emotion is Neutral (CF = 0.58)

Rule 3: IF Temp EN ≤ 12.849018 and RH ≤ 46.7825 THEN Emotion is Positive (Cf= 0.81)

Interpretation (Inference)

Comments on participants #1’s emotion: Participant #1 feels comfortable under a lower temperature and a lower humidity.

Cf indicates certainty factor
Another Idea: Deep learning (Convolutional Neural Network)

Multilevel stress labeling to the city images (or another form of architectural features) using EDA response/Survey response

224 x 224 image

Convolution 224 x 224 x 10

Pooling 112 x 112 x 10

3-output fully Connected NN layer

Positive
Neutral
Negative
Thank You

What an optimization algorithm can do for us?

Genetic Algorithm: http://rednuht.org/genetic_cars_2/