Research Topic

Jaesung Lee
Biography

Jaesung Lee

2001
CAUCSE
Bachelors’ degree course
(2001 – 2007)

Military Service

First Research Project (2007)
Feature Extraction Method for Imbalanced Bio Data, KOSAF

2007
2009 2010
CAUCSE
Master degree course
(2007 – 2009)

CAUCSE
Doctoral course
(2009 – 2013)

First SCI Paper (2010)

Top Class Journals
Pattern Recognition
Information Sciences
Expert Systems with Application

13 SCI(E) Papers (2016)

Year
2013 2014 2015 2016
CAUCSE
BK21 Research Professor
(2013 – Current)

Recent Research Project (2014)
Music Map: Investigating Relation of Meta Information, NRF
Research Interest

Machine Learning & Data Mining

Classification

Attribute Selection

Mutual Information

Greedy Search Algorithm

Genetic Search Algorithm

Sensor Signal Recognition

Classification

Music Recommendation

Classification

Mobile Robot Path Planning

Search Algorithm

Application

Theory

Improves accuracy and reduces time cost

Gives computers the ability to learn without explicit programming

Relates to expert system or decision support system
Research Topic: Classification

- Classification is identifying the membership of unseen example based on training examples.
- The membership can be car types, the meaning of sensor signals, and so on.
Research Topic: Classification

- After describing examples based on quantifiable properties, training data set was obtained.
- In this example, each car is described by four attributes: ID, Color, Height, and Length.

### Training Patterns

<table>
<thead>
<tr>
<th>ID</th>
<th>Color</th>
<th>Height</th>
<th>Length</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gray</td>
<td>1.5m</td>
<td>3.1m</td>
<td>Sedan</td>
</tr>
<tr>
<td>2</td>
<td>Blue</td>
<td>1.4m</td>
<td>3.2m</td>
<td>Sedan</td>
</tr>
<tr>
<td>3</td>
<td>Blue</td>
<td>1.6m</td>
<td>2.8m</td>
<td>Sedan</td>
</tr>
<tr>
<td>4</td>
<td>Black</td>
<td>2.1m</td>
<td>2.5m</td>
<td>SUV</td>
</tr>
<tr>
<td>5</td>
<td>White</td>
<td>2.3m</td>
<td>3.0m</td>
<td>SUV</td>
</tr>
</tbody>
</table>

### Unseen Pattern

<table>
<thead>
<tr>
<th>ID</th>
<th>Color</th>
<th>Height</th>
<th>Length</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Black</td>
<td>2.2m</td>
<td>3.1m</td>
<td>?</td>
</tr>
</tbody>
</table>
• Suppose that we consider Nearest Neighbor classifier to identify the type of unseen pattern.
• At first it calculates the similarity between each training-unseen pattern pairs.
• In this example, the first training pattern is compared to unseen pattern.
Research Topic: Classification

- Based on a distance measure, the dissimilarity between the pair’s is calculated.

Dissimilarity Calculation

ID | Color | Height | Length
---|-------|--------|--------
1  | Gray  | 1.5m   | 3.1m   
6  | Black | 2.2m   | 3.1m   

\[ 5.0 + 1.0 + 0.7 + 0.0 = 6.7 \]
Research Topic: Classification

- The calculation results indicate 4th pattern is the most similar pattern.
Research Topic: Classification

• **NN assigns the class information (car type) of training pattern to unseen pattern.**

• **Basically, this example also shows the core concept of machine learning-based classification;**
  
  *It utilizes the information obtained from training examples to achieve our task.*

---

Dissimilarity

Propagates class information of training pattern

**SUV** → **SUV**
Research Topic: Attribute Selection

- However, the information carried by ID or Color is unimportant to classify the type of car.
- Irrelevant attributes increase the time of the learning process.
- It also decreases the classification accuracy.

These cars can be classified as SUV.
Research Topic: Attribute Selection

- To avoid the above problems, important attributes can be selected.
- To select attributes, the importance of each attribute is calculated.
Research Topic: Attribute Selection

- Attributes with low importance are eliminated from data set.
- Only remaining attributes are used for learning.

### Training Patterns

<table>
<thead>
<tr>
<th>ID</th>
<th>Color</th>
<th>Height</th>
<th>Length</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gray</td>
<td>1.5m</td>
<td>3.1m</td>
<td>Sedan</td>
</tr>
<tr>
<td>2</td>
<td>Blue</td>
<td>1.4m</td>
<td>3.2m</td>
<td>Sedan</td>
</tr>
<tr>
<td>3</td>
<td>Blue</td>
<td>1.6m</td>
<td>2.8m</td>
<td>Sedan</td>
</tr>
<tr>
<td>4</td>
<td>Black</td>
<td>2.1m</td>
<td>2.5m</td>
<td>SUV</td>
</tr>
<tr>
<td>5</td>
<td>White</td>
<td>2.3m</td>
<td>3.0m</td>
<td>SUV</td>
</tr>
</tbody>
</table>

![Normalized Mutual Information Diagram]

- ID: 0.42 – Type
- Color: 0.51 – Type
- Height: 1.00 – Type
- Length: 0.02 – Type
The type of unseen pattern is still classified accurately. Moreover, learning process is much more simplified (Approximately 4x).
Research Achievement
Contribution to Research Field

- Electronics Letters (2012)
  Computationally-efficient score function considering correlation among attributes

- Pattern Recognition Letters (2013)
  First information theoretic score function that considers multiple concepts in the world

- Pattern Recognition (2015)
  Extremely-fast multi-label attribute selection method based on important concepts

- Information Sciences (2015)
  Memetic search method for identifying effective candidate attribute set was developed

  Upper and Lower bound of multi-label attribute selection was derived and reported

- Ph. D. Dissertation
### SCI Papers

<table>
<thead>
<tr>
<th>Seq.</th>
<th>Title</th>
<th>Journal</th>
<th>IF</th>
<th>Year</th>
<th>Role</th>
<th>IF Rank</th>
<th>ES Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Accelerating Multi-label FS based on Low-Rank Approximation</td>
<td>IEICE Transactions on Information and Systems</td>
<td>0.213</td>
<td>2015</td>
<td>Co-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>An Effective Initialization Method for GA-based RPP using a DAG</td>
<td>Information Sciences</td>
<td>4.038</td>
<td>2016</td>
<td>First</td>
<td>Top 4%</td>
<td>Top 1%</td>
</tr>
<tr>
<td>3</td>
<td>Fast Multi-label FS based on Information-theoretic Feature Ranking</td>
<td>Pattern Recognition</td>
<td>3.096</td>
<td>2015</td>
<td>First</td>
<td>Top 8%</td>
<td>Top 4%</td>
</tr>
<tr>
<td>4</td>
<td>MI-based Multi-label FS using Interaction Information</td>
<td>Expert Systems with Applications</td>
<td>2.240</td>
<td>2015</td>
<td>First</td>
<td>Top 15%</td>
<td>Top 1%</td>
</tr>
<tr>
<td>5</td>
<td>Memetic (GA) FS Algorithm for Multi-label Classification</td>
<td>Information Sciences</td>
<td>4.038</td>
<td>2015</td>
<td>First</td>
<td>Top 4%</td>
<td>Top 1%</td>
</tr>
<tr>
<td>6</td>
<td>Multi-label Learning (FS) using Mathematical Programming</td>
<td>IEICE Transactions on Information and Systems</td>
<td>0.213</td>
<td>2015</td>
<td>Co-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>ROBIL: RPP based on PBIL (GA) Algorithm</td>
<td>International Journal of Advanced Robotic Systems</td>
<td>0.526</td>
<td>2014</td>
<td>Co-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Efficient DTW for 3D Handwriting Classification</td>
<td>Expert Systems with Applications</td>
<td>2.240</td>
<td>2014</td>
<td>Co-</td>
<td>Top 15%</td>
<td>Top 1%</td>
</tr>
<tr>
<td>9</td>
<td>Fast Genetic Algorithm (GA) for Robot Path Planning (RPP)</td>
<td>Electronics Letters</td>
<td>0.930</td>
<td>2013</td>
<td>First</td>
<td>Top 10%</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>FS for Multi-label Classification using Multivariate MI</td>
<td>Pattern Recognition Letters</td>
<td>1.551</td>
<td>2013</td>
<td>First</td>
<td>Top 11%</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Approximating Mutual Information (MI) for Multi-label FS</td>
<td>Electronics Letters</td>
<td>0.930</td>
<td>2012</td>
<td>First</td>
<td>Top 10%</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Efficient Multivariate Feature Selection (FS) using Conditional MI</td>
<td>Electronics Letters</td>
<td>0.930</td>
<td>2012</td>
<td>First</td>
<td>Top 10%</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Classifying Data based on Adoptive Hamming Distance</td>
<td>IEICE Transactions on Information and Systems</td>
<td>0.213</td>
<td>2010</td>
<td>First</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- JCR Eigenfactor Score (ES): *A measure for rating the quality of journal*. IF and ES are two main evaluation measures of National Research Foundation.
A score function for multi-label attribute selection was derived theoretically.

\[
J = \sum_{l_i \in L} I(f^+, l_i) - \sum_{f_j \in S} \sum_{l_k \in L} I(f^+, f_j, l_k) - \sum_{l_u \in L} \sum_{l_w \in L} I(f^+, l_u, l_w)
\]

where \( I(\cdot) \) is multivariate mutual information, \( S \) is already-selected attribute subset, and \( L \) is given set of target labels we want to classify or predict.
Output

A bound for derived score function was investigated.

\[
\frac{3}{2} V_3(S') - \frac{n - 2}{2} V_2(S') \leq V_3(S') \leq \frac{n - 2}{2} V_2(S')
\]

where \( V_k(\cdot) \) is \( k \)-degree multivariate mutual information, \( S' \) is the power set of \( S \), and \( n \) is the number of variables in \( S \).
Output

The computational efficiency of calculating score function was improved.

\[
J = \tilde{V}_2(f^+ \times L'_1) + \sum_{k=3}^{b+1} \tilde{V}_k(f^+ \times Q'_{k-1})
\]

where \( \tilde{V}_k(\cdot) \) is an approximation of \( k \)-degree multivariate mutual information, \( Q \) is a set of important labels for preserving target structure.
A hybrid search method was developed to enhance the search capability.

Initialization
- Randomly generate subsets
- Evaluate each feature subset to obtain score value

Local Refinement
- Obtained subsets are refined by the greedy feature selection to accelerate the search

Create new solution
- Create new solutions randomly by referencing already obtained feature subsets

Maintain solutions
- Predefined number of subsets with higher scores are remained, and discard others

Iterates
Contribution to Research Field

Published Papers (Focus on applied problems)

Fast Genetic Algorithm for Robot Path Planning
Information Sciences 332(1), March 2016
Effective initialization method for GA-based robot path planning was proposed.

Fast Genetic Algorithm for Robot Path Planning
Electronics Letters 49(23), November 2013
Efficient search algorithm for robot path planning was developed.

Efficient Dynamic Time Warping for Smartphone Service
Expert Systems with Applications 49(23), November 2013
Efficient search algorithm for finding similar time series patterns was developed.
## Research Projects

<table>
<thead>
<tr>
<th>Title</th>
<th>Role</th>
<th>Period</th>
<th>Ack.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Music Map: Investigating Relation of Meta Information for Music Recommendation</td>
<td>PI (Director)</td>
<td>2014.11 - 2016.04</td>
<td>NRF</td>
</tr>
<tr>
<td>Techniques for multi-perceptual contents based on aesthetic synchronization</td>
<td>Research Assistant</td>
<td>2014.06 - 2017.03</td>
<td>KOCCA</td>
</tr>
<tr>
<td>Next Generation Big Data Research Team</td>
<td>Research Professor</td>
<td>2013.03 - 2020.02</td>
<td>NRF</td>
</tr>
<tr>
<td>Aesthetic Cognition-Emotion Inference for Personalized Music Service</td>
<td>Research Assistant</td>
<td>2013.06 - 2016.05</td>
<td>NRF</td>
</tr>
<tr>
<td>Korean Situation-Context-Emotion based Music Recommendation System</td>
<td>Project Leader</td>
<td>2010.05 - 2013.04</td>
<td>NRF</td>
</tr>
<tr>
<td>Stroke Prognosis Prediction Techniques using Temporal Clinical Data</td>
<td>Project Leader</td>
<td>2009.05 - 2011.04</td>
<td>NRF</td>
</tr>
<tr>
<td>Feature Extraction Method for Imbalanced Bio Data</td>
<td>PI (Director)</td>
<td>2007.09 – 2008.08</td>
<td>KOSAF</td>
</tr>
</tbody>
</table>

- Principal Investigator (PI)
## Patents

<table>
<thead>
<tr>
<th>Title</th>
<th>Domain</th>
<th>Year</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apparatus and Method for planning robot path using genetic algorithm</td>
<td>Robot Path Planning</td>
<td>2016</td>
<td>Registered</td>
</tr>
<tr>
<td>Method and apparatus for determining predictive tag value each of a plurality of music tag for generating mapping function of music</td>
<td>Music Recommendation</td>
<td>2016</td>
<td>Registered (Tech. Transf.)</td>
</tr>
<tr>
<td>Apparatus and Method for pronunciation training for beginners of foreign language learning</td>
<td>Smartphone Service</td>
<td>2015</td>
<td>Pending</td>
</tr>
<tr>
<td>Method and apparatus for providing music retrieval service based on tag combination using prototypical music content</td>
<td>Music Recommendation</td>
<td>2015</td>
<td>Registered</td>
</tr>
<tr>
<td>Method and apparatus for multiple meaning classification related music</td>
<td>Music Recommendation</td>
<td>2015</td>
<td>Registered</td>
</tr>
<tr>
<td>Method and apparatus for selecting feature used to classify multi-label</td>
<td>Attribute Selection</td>
<td>2015</td>
<td>Pending</td>
</tr>
<tr>
<td>English pronunciation training method and apparatus</td>
<td>Smartphone Service</td>
<td>2014</td>
<td>Pending</td>
</tr>
<tr>
<td>Pattern recognition method and apparatus within terminal</td>
<td>Smartphone Service</td>
<td>2014</td>
<td>Registered</td>
</tr>
<tr>
<td>Method and apparatus for path search</td>
<td>Robot Path Planning</td>
<td>2013</td>
<td>Pending</td>
</tr>
</tbody>
</table>
## Awards and Scholarship

<table>
<thead>
<tr>
<th>Title</th>
<th>Year</th>
<th>Activity and Results</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Excellence in Research Award 2012</strong></td>
<td>2013</td>
<td>School of Computer Science and Engineering, Chung-Ang University</td>
<td>4 SCI Papers</td>
</tr>
<tr>
<td>CAU Research Assistant (A)</td>
<td>2009</td>
<td>Chung-Ang University</td>
<td></td>
</tr>
<tr>
<td>Seoul Fellowship</td>
<td>2009</td>
<td>Seoul Metropolitan Government</td>
<td></td>
</tr>
<tr>
<td>CAU Scholarship</td>
<td>2007</td>
<td>Leading Research Team, Chung-Ang University</td>
<td></td>
</tr>
<tr>
<td>National Science and Engineering Scholarship</td>
<td>2007</td>
<td>Korea Student Aid Foundation</td>
<td></td>
</tr>
</tbody>
</table>

- Feature Selection for Multi-label Classification using Multivariate Mutual Information 34(2), 2013, First author
- Approximating Mutual Information for Multi-label Feature Selection, Electronics Letters 48(15), 2012, First author
- Classifying Categorical Data based on Adoptive Hamming Distance, IEICE Transactions on Information and Systems E93-D(1), 2010, First author
Thank you