Proxy Setting For Developing A Quantitative Livability Forecasting Metric: Identifying Freight-traffic Role

Golnaz Sarram
University of Memphis
Intermodal Freight Transportation Institute
7th METRANS International Urban Freight Conference
October 19, 2017
Outline

1. Introduction
   Necessity of study | Principles

2. Background
   Past works | Data & study area

3. Problem Statement
   Methodology | Related Works

4. Detecting Relations in Livability Pattern
   Data Structure | Choosing classifiers

5. Implementation Approach
   Data | Ultimate Objectives
1. Introduction

- **Sustainable city development:**
  - Depends on the optimality of this relation
  - Need for understanding quality of life pattern

![Diagram showing relationships between freight movement, environment and safety, congestion, and their impacts on access and costs, talent attraction and retention, and quality of life.]
1. Introduction

- **Partnership for Sustainable Communities (2009):**
  - US Department of Transportation
  - Housing and Urban Development (HUD)
  - Environmental Protection Agency (EPA)

1. Provide more transportation choices
2. Promote equitable, affordable housing
3. Enhance economic competitiveness
4. Support existing communities
5. Coordinate and leverage federal policies and investment
6. Value communities and neighborhoods
2. Background

- **Defining Livability for Freight-Centric Communities (Ivey et. al -2013):**
  - Three survey instruments (residents, industries, and planners) reflecting
    - livability values and barriers
  - Residential survey was administered in Memphis
    - As a large Freight-centric urban community
    - Showing no differences in terms of **barriers** but in **conditions**

- **Analytical Hierarchy Process (AHP):**
  - Decision making technique (Pairwise comparison judgment)
  - Quantifies what we cannot measure
  - Prioritize values
  - Built the hierarchy based on the results of the residential survey and comments gathered during focus groups
2. Background

➢ Further Details:

1) Proceedings of the ASCE 2nd T&DI Green Streets, Highways and Development Conference, Austin, TX
   “Defining Livability for Freight-Centric Communities: Identifying Priorities of Residents of the Lamar Avenue Corridor in Memphis, TN.”

2) Mid-Continent Transportation Research Symposium. Madison, WI
   “Making Freight-Centric Communities More Livable: Examining Residential Perceptions and Priorities for Livable Communities”.

3) TRB 95th Annual Meeting
   “Evaluation of Factors Affecting Livability in a Freight-Centric Community in Memphis, Tennessee.”
3. Problem Statement

- **Problem Statement:**
  
  - The key obstacle:
    
    - Obtaining participants from a representative sample of community stakeholders
    - Statistical analysis for comparing the two distributions of scores were significantly different
    - The metric was not aligned with the stakeholders’ perceptions

- Interpreting a linkage between society stated preferences and quantitative measures of livability
  
  - Providing the opportunity to use additional methodologies with a larger dataset
  - Other cities freight-centric communities and states
  - Test the hypothesis that:
    
    - If a common definition and approach for measuring livability is possible
    - or if these are community dependent
3. Problem Statement

- Methodology:
  - Reviewing the scientific investigations of human response data related to opinion mining and sentiment analysis
  - Statistical approaches ➔ Univariate Methods
  - Machine learning technique applied to data from the previous livability survey obtained from urban freight-centric communities to identify key factors influencing perceptions of livability
    - Multivariate correlation and Group effect concepts
    - Explores algorithms that can learn from data by building a model from examples inputs and make data-driven predictions or decisions
3. Problem Statement

➢ Methodology:

• Understanding that the livability score decisions made by residential stakeholders can be appraised based on daily life choices and preferences

• Later allows consideration of a quantitative metric using a customer service satisfaction framework (or similar, such as predicting customer brand loyalty), to determine which pertinent indicators are important in this decision

• Anticipating no need to survey data (Can be used for unknown residents)

• Creating a community-based performance metric representing users’ perceptions of livability determined through data mining techniques
3. Problem Statement

- Related Works:

A) Approaches for Identifying Consumer Preferences for the Design of Technology Products: A Case Study of Residential Solar Panels

Heidi Q. Chen, Tomonori Honda and Maria C. Yang

[+] Author and Article Information

*J. Mech. Des* 135(6), 061007 (May 09, 2013) (12 pages)

Paper No: MD-12-1419; doi: 10.1115/1.4024232

History: Received August 19, 2012; Revised April 06, 2013

B) Towards Real-time Customer Experience Prediction for Telecommunication Operators

Ernesto Diaz-Aviles, Fabio Pinelli, Karol Lynch, Zubair Nabi, Yiannis Gkoufas, Eric Bouillet, and Francesco Calabrese

IBM Research – Ireland
## 4. Detecting Relations in Livability pattern

### Choosing Proper Learning Algorithm:

*The three components of learning algorithms*

<table>
<thead>
<tr>
<th>Representation</th>
<th>Evaluation</th>
<th>Optimization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instances</td>
<td>Accuracy/Error rate</td>
<td>Combinatorial optimization</td>
</tr>
<tr>
<td></td>
<td>Precision and recall</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Squared error</td>
<td>• Greedy search</td>
</tr>
<tr>
<td></td>
<td>Likelihood</td>
<td>• Beam search</td>
</tr>
<tr>
<td>Hyperplanes</td>
<td>Posterior probability</td>
<td>• Branch-and-bound</td>
</tr>
<tr>
<td></td>
<td>Information gain</td>
<td></td>
</tr>
<tr>
<td>Naive Bayes</td>
<td>K-L divergence</td>
<td></td>
</tr>
<tr>
<td>Logistic regression</td>
<td>Cost/Utility</td>
<td></td>
</tr>
<tr>
<td>Decision trees</td>
<td>Margin</td>
<td></td>
</tr>
<tr>
<td>Random forest</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sets of rules</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Propositional rules</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logic programs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neural networks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graphical models</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayesian networks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conditional random fields</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Choosing the set of classifiers in the hypothesis space of the learner that it can possibly learn
- Selecting most related features
- Learning many models, not just one

- Search among the classifiers in the language for the highest-scoring
4. Detecting Relations in Livability pattern

- **Data Structure:**

**Supervised learning**

- Training inputs and desired outputs
- Label data - a series of discrete responses to the ‘How do you rate your neighborhood for livability?’ (scale of 1-10) **Customer Satisfaction Score (CSAT)** – Multi-Class problem
- The goal is to learn a general rule that maps inputs to outputs

<table>
<thead>
<tr>
<th>Features (indicators)</th>
<th>Category 1</th>
<th>Category 2</th>
<th>Category 3</th>
<th>...</th>
<th>Target (Livability scores)</th>
</tr>
</thead>
<tbody>
<tr>
<td>transit service area</td>
<td>Bike Facilities</td>
<td>Affordability Index</td>
<td>Walking to</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Neighborhood-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neighborhood-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4. Detecting Relations in Livability pattern

- **Data Structure:**

*Supervised learning*

- Relative nature of the livability scores

- Respondents from diverse areas with varied backgrounds most likely have different expectations and opinions on how a livability score is defined

- It is concluded that there should be a categorization of the livability scores to provide the highest model performance

- Therefore, the machine learning classification was initiated by a manual process of grouping different subsets of the livability scores to determine an appropriate classification

- As a result, this repeated subset evaluation analysis identified two types of categorization schemes, respectively grouped scores include: {Class I (Medium: 4, 5, 6, 7) and (High: 8, 9 10)} and {Class II (Low: 4, 5), (Medium: 6, 7) and (High: 8, 9 10)}
5. Result

Table 1. Performance of classifiers before and after feature selection (n=472)

<table>
<thead>
<tr>
<th>Indicators set</th>
<th>Datasets</th>
<th>Metrics</th>
<th>Classifiers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Pre-SMOTE</td>
<td>SVM</td>
</tr>
<tr>
<td>Full features</td>
<td>Class I: Medium, High</td>
<td>Accuracy Rate</td>
<td>64.54%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weighted Avg. F-measure</td>
<td>0.633</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weighted Avg. ROC-Area</td>
<td>0.616</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G-mean</td>
<td>0.621</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-SMOTE</td>
<td>SVM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Accuracy Rate</td>
<td>73.93%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weighted Avg. F-measure</td>
<td>0.708</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weighted Avg. ROC-Area</td>
<td>0.708</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G-mean</td>
<td>0.7</td>
</tr>
<tr>
<td>Reduced features</td>
<td></td>
<td>Post-SMOTE</td>
<td>Wrapper</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Accuracy Rate</td>
<td>74.57%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weighted Avg. F-measure</td>
<td>0.737</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weighted Avg. ROC-Area</td>
<td>0.716</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G-mean</td>
<td>0.697</td>
</tr>
</tbody>
</table>
## 5. Result

### Table 2. List of extracted features in final training dataset

<table>
<thead>
<tr>
<th>Top Ranked Features In Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiencing <strong>negative environmental issues</strong> (smog, air pollution, noise, or otherwise)</td>
</tr>
<tr>
<td>Living in an economically thriving neighborhood</td>
</tr>
<tr>
<td>Knowing my neighbors</td>
</tr>
<tr>
<td>Minimal road congestion</td>
</tr>
<tr>
<td>Quality affordable housing</td>
</tr>
<tr>
<td>Having a park in my neighborhood</td>
</tr>
<tr>
<td>How often stuck in traffic due to <strong>trains</strong> (Response: Occasionally)</td>
</tr>
<tr>
<td>Feeling safe in my neighborhood</td>
</tr>
<tr>
<td>Experiencing presence of <strong>freight or heavy trucks</strong> traffic</td>
</tr>
<tr>
<td>Living close to school/work</td>
</tr>
<tr>
<td>Having alternative transportation options (walk, bike, public transit)</td>
</tr>
</tbody>
</table>
5. Result

Fig. 1. (a) Respondents distribution map  (b) Livability classification based on frequency of scores
5. Conclusion

- The results show that there is a rule relating neighborhood perceptions to participants’ livability scoring systems that can be revealed through machine learning techniques.
- Prioritizing livability indicators through data mining techniques and establishing a proxy metric also shows promise for developing decision-making tools.
- The procedure reshapes the consistency of proxy settings with reality, even if the community preferences change over time.
- Consistency with the previous Analytical Hierarchy Process (AHP) approach;
  - Uncovering more apparent impact of freight on livability perceptions than was revealed through the previous AHP.
- Although the pilot results are promising, it is believed that with additional research, larger datasets, and data from multiple settings, more efficient livability indicators can be identified, adopted, and employed for planning purposes.
6. Implementation approach

**Phase I – Identifying livability definition (Attitudinal Data)**

- Livability Index classification analysis in all divisions
- Prioritizing preferences of the users
- **Pilot Study:**
  - Hypothesis I: Customer Satisfaction (CS) analysis can classify the diversity of stakeholders’ quality of life perceptions accurately

**Phase II – Proxy Settings (Behavioral and Socio-Economic Data)**

- Learning the users’ mobility, commuting and mode choice patterns as a measure of Livability
- Correlating the behavioral metric to Livability Index

**Efficiency-Oriented Decision-Making Tool**

- The research outcomes (Multidimensional Livability Index (MPI)) provide a practical standard for urban metrics studies to assess their planning performance
- The updated urban modeling procedure can consider residential satisfaction in scenario planning through forecasting project impacts on urban configurations
6. Implementation approach

Data:

- Knowledge from the user experience data
- Users’ activity and mode choices data
- Assess planning performances
6. Implementation approach

- **Livability proxy setting ultimate objectives:**
  - Transportation development and decision-making depends on the associated comparisons for clarifying the differences among various sets of strategies within long-term financial constraints.
  - The evaluation process mostly ends up with hybrid scenarios from a combination of strategies, indicating the natural interconnected relationship of livability factors and complex outcomes of these decisions in the performance stage.
  - Optimizing this relationship:

![Diagram showing the relationship between Livability proxy setting ultimate objectives, Easy Access to opportunities, Efficiency-Oriented Decision-making, and Transportation/Land use Integration.](image)
Questions?

Golnaz Sarram

gsarram@Memphis.edu