Assessing the Impact of Hotel Services to the Customer Rating Using Fuzzy String Matching and Belief Networks

Alexandros Bousdekis[1*] **Dimitris Kardaras**[1] **Stavroula Barbounaki**[2]

[1]*Business Informatics Lab, Department of Business Administration, School of Business, Athens University of Economics and Business (AUEB), Athens, Greece*

[2]*Merchant Marine Academy of Aspropyrgos, Aspropyrgos, Greece*

[*]*albous@mail.ntua.gr*

Presenter's Resume

Alexandros Bousdekis

- Current position
  - Post-doctoral Researcher (Athens University of Economics and Business)
    - Title of postdoctoral research: "Advanced data analytics and knowledge discovery for e-service customization"
- Education
  - PhD in Information Systems (National Technical University of Athens)
  - MSc in Manufacturing Systems Engineering (University of Warwick, UK)
  - BSc in Production and Management Engineering (Technical University of Crete)

Outline

- Introduction
- Research Methodology
- Results
- Conclusions & Future Work

Outline

Introduction

- Research Methodology
- Results
- Conclusions & Future Work

Motivation

- With the increased popularity of online bookings, **53% of travellers** state that they would be **unwilling to book a hotel that had no reviews**, while a **10% increase in travel review ratings would increase bookings by more than 5%**.

- These online reviews in the e-tourism era, in the format of both textual reviews (comments) and ratings, generate an **electronic Word Of Mouth (eWOM) effect**.

- In contrast to a pre-designed questionnaire survey, online textual reviews have an **open-structured form** and can:
  - show **customer consumption experiences**
  - highlight the product and service **attributes customers care about**
  - provide customers' **perceptions in a detailed way**.

Research Objective

- **Hotel quality evaluation from online reviews** is an **emerging research field**. However:
  - the exploitation of online textual reviews is still largely **under-explored**
  - there is a **lack of advanced data analytics** approaches for modelling complex dynamics of **online hotel review data**.

- The **increasing amount of online reviews** pose significant challenges for the development of advanced data analytics models providing a **higher level of intelligence** and thus, **increased business value**.

- In this paper, we propose an approach for **assessing the impact of hotel services to the customer rating** using **Fuzzy String Matching (FSM)** and **Bayesian Belief Networks (BBN)**.

- The **objective** is to provide a unified algorithm, which :
  - **mines customers' opinions from online hotel reviews** (review comments and rating)
  - **evaluates the hotel performance** by identifying **how the various attributes** (e.g., location, cleanliness, breakfast, etc.) **affect the overall review rating**.

Outline

- Introduction
- Research Methodology
- Results
- Conclusions & Future Work

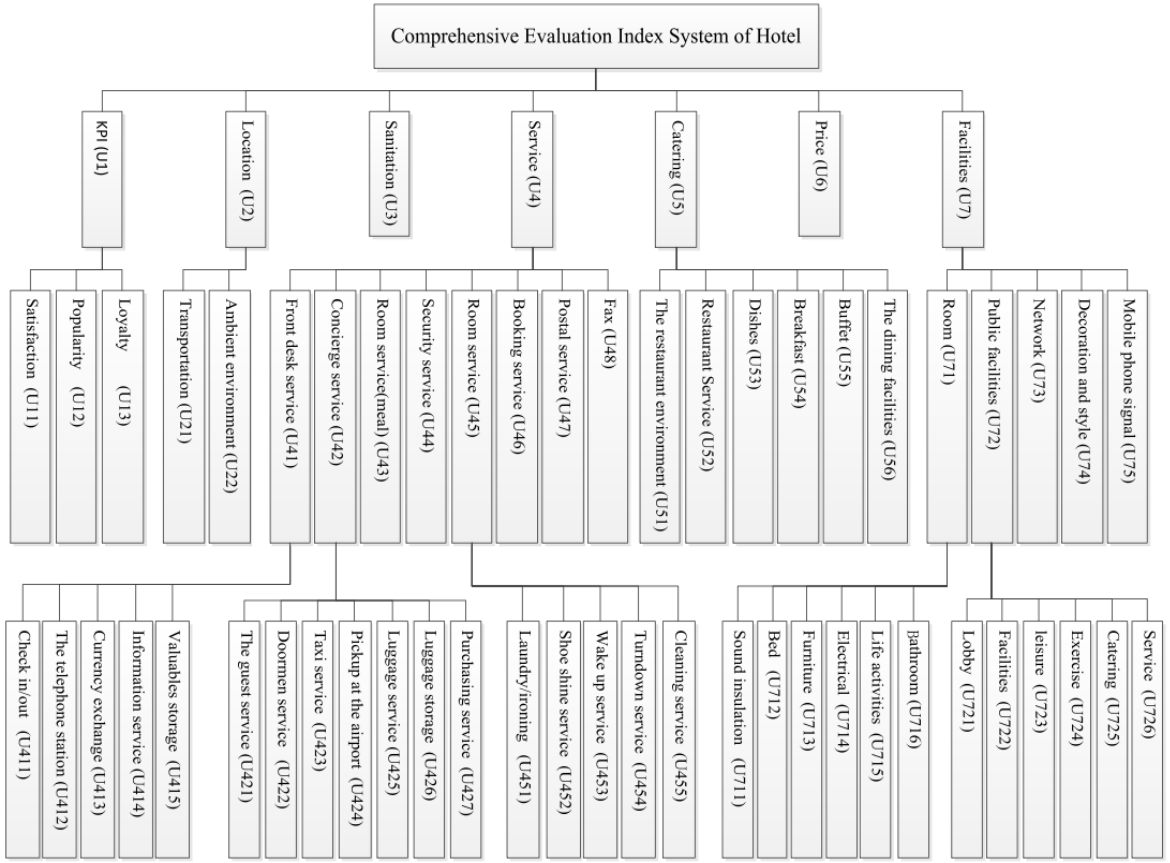
The Proposed Methodology

- The research methodology consists of 3 main steps:
  - Extracting the evaluation criteria from online comments
  - Mining customers' opinions using FSM
  - Assignment of sentiment scores to a discrete scale
  - Applying BBN for assessing the impact of hotel services to the customer rating

Extracting the Evaluation Criteria from Online Comments

- The proposed approach utilizes **3 fields from the online hotel reviews** in order to extract the evaluation criteria:
  - **review title**
  - **review comments**
  - **review rating**

- Based upon an evaluation index for hotel service quality, this step identifies the **criteria mentioned in the hotel reviews**. E.g.
  - Location
  - Price
  - Breakfast
  - room space
  - …



Mining Customers' Opinions Using Fuzzy String Matching (1/2)

- Since online comments are written in natural and informal language, there is the need to **mine customers' opinions**.

- FSM is able to take into account the **imprecision** and the **uncertainty** pervading values, which have to be compared in a matching process.

- In online review comments, **different customers may use different words or phrases to express their opinions**, while **the comments may be vague**.
  - For example, poor cleanliness can be expressed as: "The room was too dirty", "Very dirty", etc.
  - Regular expression is an efficient pattern match technology to identify the specific pattern strings from a long text.
  - However, the regular expression method causes a binary value result: match or not match.

Mining Customers' Opinions Using Fuzzy String Matching (2/2)

- In the proposed approach, we apply **FPMT** as an effective **fuzzy string matching method** to deal with the vagueness of the free text online comments.
  - Although this method results in some mismatched cases, this causes little impact on the final result, because there are **many redundant comments with similar semantics**.

- The **output** of customers' opinions mining is a **fuzzy evaluation of the extracted criteria**.
  - First, the extracted evaluation criteria of hotel quality are assigned to a **5-level Likert scale**.
  - Then, we consider the **median of the resulting responses** in order to represent the magnitude of causality among the evaluation criteria.

Assignment of Sentiment Scores to a Discrete Scale

- In this step, the sentiment scores are assigned to a discrete scale.
  - For example, if the review rating takes values between 1 and 5, the sentiment scores are classified to a respective discrete scale:
    - [-1, -0.6] is assigned to "DISASTER"
    - (-0.6, -0.2] is assigned to "MANY THINGS NEED TO BE IMPROVED"
    - (-0.2, +0.2] is assigned to "FAIR ENOUGH"
    - (+0.2, +0.6] is assigned to "PERFECT"
    - (+0.6, +1] is assigned to "ABSOLUTELY PERFECT"

Applying BBN for Assessing the Impact of Hotel Services to the Customer Rating

- The relationships between the sentiment discrete scale created in the previous step and the review rating of the customer are modelled in a probabilistic model with the use of BBN.

- The outcome indicates the **probability of having a specific value of the overall rating given the values of different services** (criteria).

- The user is able to perform **queries in order to assess the impact of each criterion to the review rating** but also combinations of criteria.

Outline

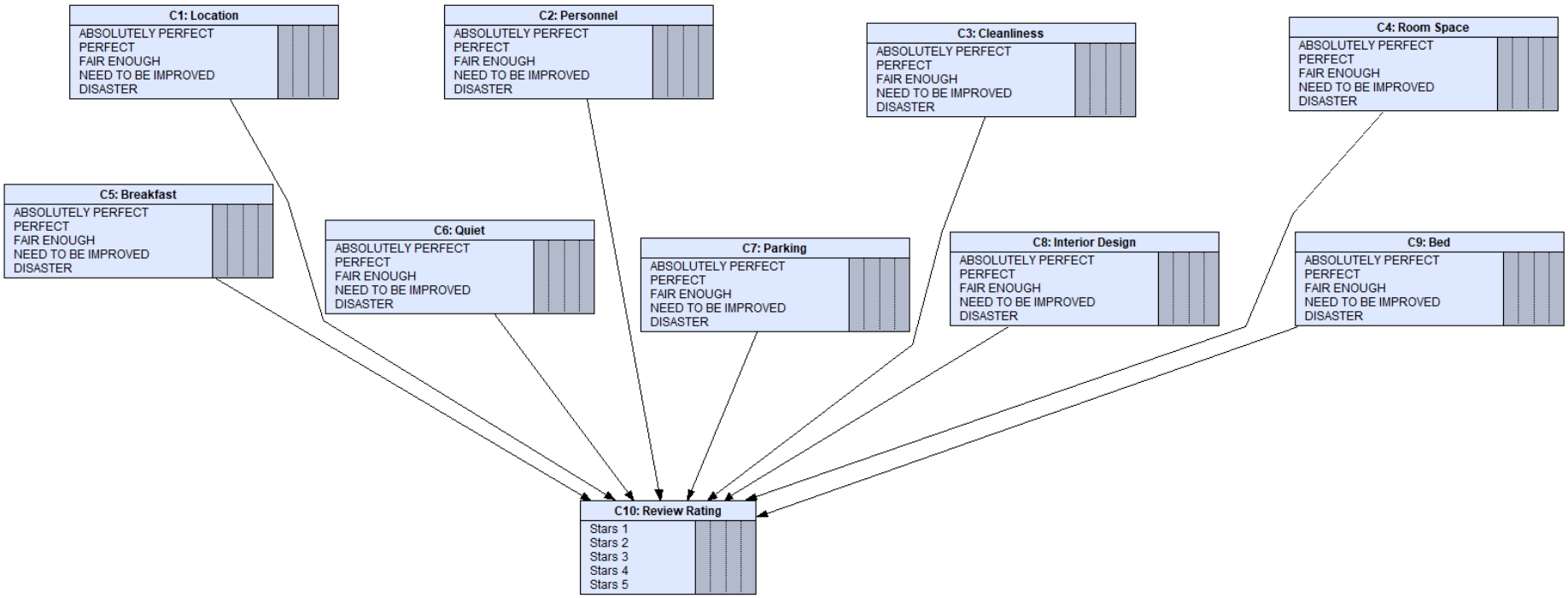
- Introduction
- Research Methodology
- Results
- Conclusions & Future Work

Data Collection and BBN Nodes

- The proposed methodology was applied to a **dataset from the TripAdvisor**.
- **The dataset consists of approx. 12,000 records** including, among others, the review title, the review comments, and the review rating.
- The **parent nodes** of the BBN are derived from the **extracted evaluation criteria** by FPMT (C1-C9) and the **child node** is the **review rating** (C10).

ID	Concepts	ID	Concepts
C1	Location	C6	Quiet
C2	Personnel	C7	Parking
C3	Cleanliness	C8	Interior Design
C4	Room Space	C9	Bed
C5	Breakfast	C10	Review Rating

The BBN Structure



Results from Indicative Queries

Values of Parent Nodes	Values of Child Node	$P(C10\|C_i)$
C1={FAIR ENOUGH}, C2={PERFECT}, C3={MANY THINGS NEED TO BE IMPROVED}, C4={PERFECT}, C5={PERFECT}, C6={FAIR ENOUGH}, C7={MANY THINGS NEED TO BE IMPROVED}, C8={FAIR ENOUGH}, C9={DISASTER}	3 stars	***0.332***
C1={MANY THINGS NEED TO BE IMPROVED}, C2={PERFECT}, C3={DISASTER}, C4={DISASTER}, C5={PERFECT}, C6={FAIR ENOUGH}, C7={FAIR ENOUGH}, C8={PERFECT}, C9={DISASTER}	2 stars	***0.241***
C1={PERFECT}, C2={ABSOLUTELY PERFECT}, C3={MANY THINGS NEED TO BE IMPROVED}, C4={PERFECT}, C5={PERFECT}, C6={ABSOLUTELY PERFECT}, C7={FAIR ENOUGH}, C8={FAIR ENOUGH}, C9={PERFECT}	4 stars	***0.214***
C1={ABSOLUTELY PERFECT}, C2={ABSOLUTELY PERFECT}, C3={FAIR ENOUGH}, C4={ABSOLUTELY PERFECT}, C5={FAIR ENOUGH}, C6={FAIR ENOUGH}, C7={FAIR ENOUGH}, C8={FAIR ENOUGH}, C9={PERFECT}	4 stars	***0.183***
C1={FAIR ENOUGH}, C2={PERFECT}, C3={MANY THINGS NEED TO BE IMPROVED}, C4={FAIR ENOUGH}, C5={PERFECT}, C6={PERFECT}, C7={MANY THINGS NEED TO BE IMPROVED}, C8={PERFECT}, C9={FAIR ENOUGH}	3 stars	***0.144***
C1={FAIR ENOUGH}, C2={ABSOLUTELY PERFECT}, C3={FAIR ENOUGH}, C4={PERFECT}, C5={ABSOLUTELY PERFECT}, C6={FAIR ENOUGH}, C7={MANY THINGS NEED TO BE IMPROVED}, C8={FAIR ENOUGH}, C9={FAIR ENOUGH}	3 stars	***0.139***
C1={MANY THINGS NEED TO BE IMPROVED}, C2={PERFECT}, C3={FAIR ENOUGH}, C4={PERFECT}, C5={ABSOLUTELY PERFECT}, C6={PERFECT}, C7={FAIR ENOUGH}, C8={PERFECT}, C9={FAIR ENOUGH}	4 stars	***0.091***
C1={PERFECT}, C2={PERFECT}, C3={FAIR ENOUGH}, C4={PERFECT}, C5={ABSOLUTELY PERFECT}, C6={PERFECT}, C7={FAIR ENOUGH}, C8={MANY THINGS NEED TO BE IMPROVED}, C9={FAIR ENOUGH}	5 stars	***0.073***

Classification Performance

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP) = 41	False Negative (FN) = 9
Actual Negative	False Positive (FP) = 3	True Negative (TN) = 32

$$Precision = \frac{TP}{TP + FP} = \frac{41}{41 + 3} = 93.1\%$$

$$Recall = \frac{TP}{TP + FN} = \frac{41}{41 + 9} = 82\%$$

Outline

- Introduction
- Research Methodology
- Results
- Conclusions & Future Work

Conclusions & Future Work

- This paper proposed an approach for **assessing the impact of hotel services to the customer rating** using **Fuzzy String Matching (FSM)** and **Bayesian Belief Networks (BBN)**.

- The proposed approach is able to **model the complex dynamics of online hotel review data**, which are derived from both the **textual nature** of the review comments and the **uncertain relationships** between these comments and the review rating.

- In our **future work**, we plan to:
  - apply our methodology to further datasets
  - to investigate the role of user profiling in hotel selection.

Thank you!