The Ninth International Conference on Data Analytics
DATA ANALYTICS 2020
October 25, 2020 to October 29, 2020 - Nice, France

Hotel Quality Evaluation from Online Reviews Using Fuzzy Pattern Matching and Fuzzy Cognitive Maps

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 **hotel quality evaluation from online reviews** using **Fuzzy Pattern Matching (FPM)** and **Fuzzy Cognitive Maps (FCM)**.

- 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

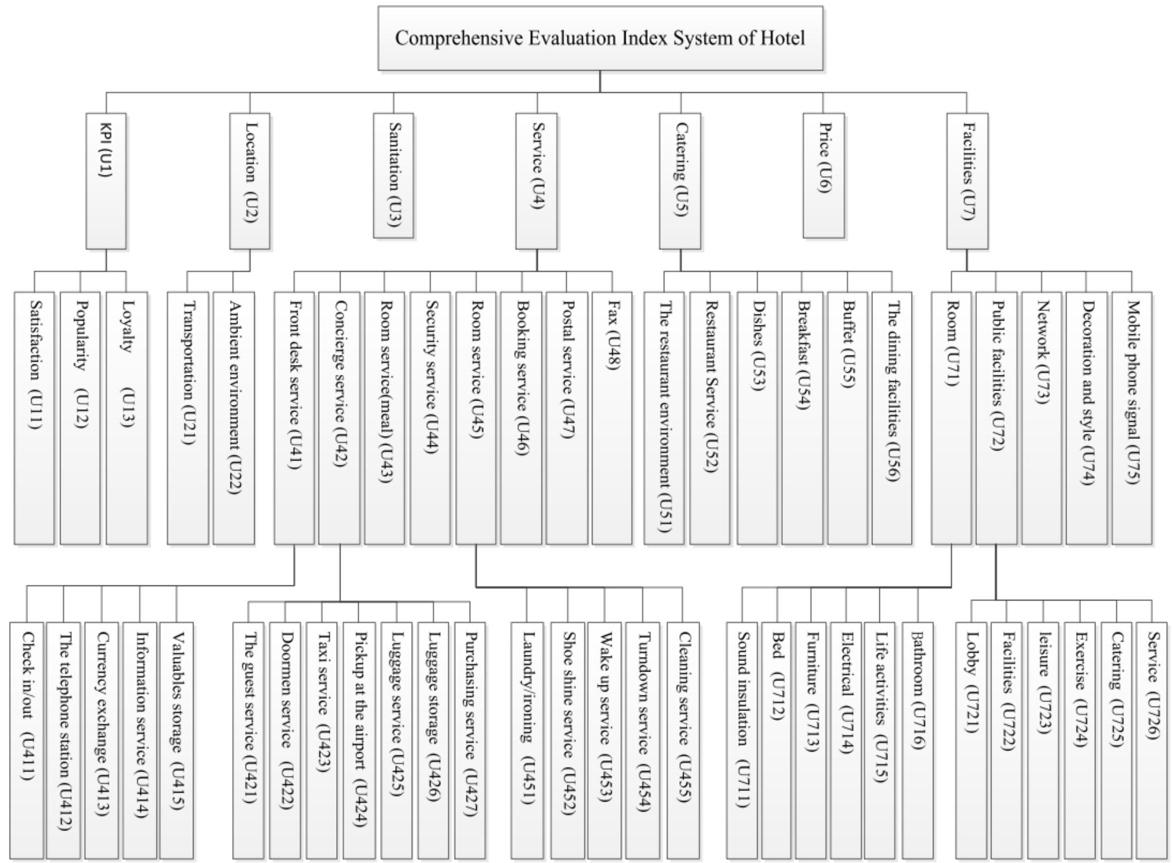
- Introduction
- Research Methodology
- Results
- Conclusions & Future Work

The Proposed Methodology

- The research methodology consists of 3 main steps:
  - Extracting the evaluation criteria from online comments
  - Mining customers' opinions using FPM
  - Applying FCM for attributes evaluation

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 Pattern Matching (1/2)

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

- FPM 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 Pattern Matching (2/2)

- In the proposed approach, we apply **FPMT** as an effective **fuzzy pattern 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 **to be used as FCM concepts**.

Applying FCM for attributes evaluation (1/2)

- This step applies **FCM** in order to:
  - **evaluate the quality of the hotels with respect to the extracted evaluation criteria (attributes)**
  - to identify the **effect of each criterion to the review rating**.

- The FCM suitability for hotel quality evaluation through online review is argued by considering that a **variety of what – if sensitivity simulations** can be performed effectively.
  - Through what – if simulations, hotels can **identify a set of relevant review factors**, pertaining to the **customer satisfaction** as well as **hotel services that need to be improved**.

- In the proposed approach, the **FCM concepts matrix consists of the extracted evaluation criteria plus an additional concept referring to the review rating**.

Applying FCM for attributes evaluation (2/2)

- The FCM is applied **separately for each hotel** in order to allow each hotel gaining meaningful insights for its performance.

- However, there is also the possibility for **aggregated results** of more than one hotel (e.g., in one region of interest, specific number of stars, same overall review rating, etc.) in the sense of an **"augmented topology"**.
  - Multiple weighted FCMs are combined into a single averaged FCM by adding their scaled and augmented adjacency weight matrix.

- If the FCMs involve different concepts, each causal matrix is augmented by adding a new column and row filled with zeros for each additional concept.

Outline

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

Data Collection and Evaluation Criteria

- The proposed methodology was applied to a **dataset including six 4-star hotels in Athens, Greece**.
- **Each hotel had 60 reviews** consisting, among others, of the review title, the review comments, and the review rating.
- The FCM concepts represent the **extracted evaluation criteria from FPMT (C1-C9) along with the review rating (C10)**.

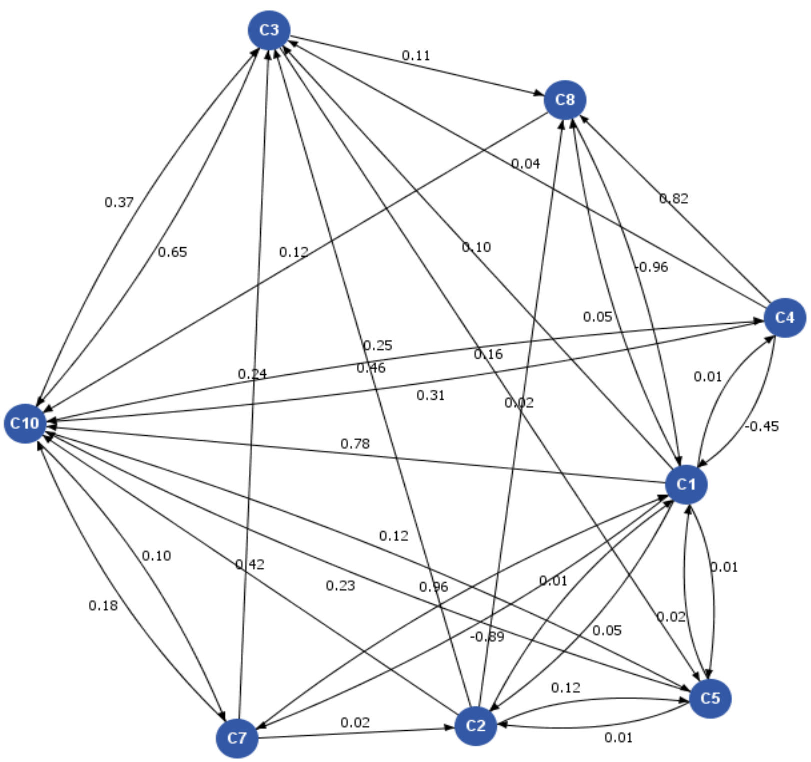
ID	Concepts		ID	Concepts
C1	Location		**C6**	Quiet
C2	Personnel		**C7**	Parking
C3	Cleanliness		**C8**	Interior Design
C4	Room Space		**C9**	Bed
C5	Breakfast		**C10**	Review Rating

The FCM Topology

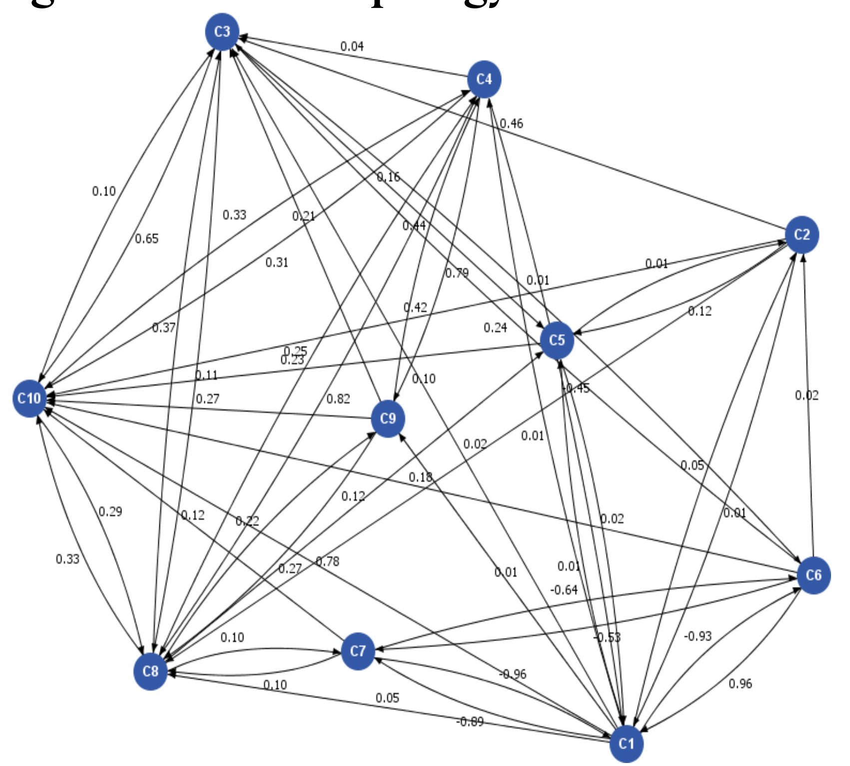
- After the fuzzy evaluation of the aforementioned concepts for each hotel, the **weight matrix** is created and is inserted to the FCM model.

- For all the 6 hotels, **the review rating (C10) is mainly affected by Location (C1), Cleanliness (C3), Room Space (C4) and Interior Design (C8)**.

FCM for 1 indicative hotel



Augmented FCM topology of all the 6 hotels



Degree Centrality of the FCM

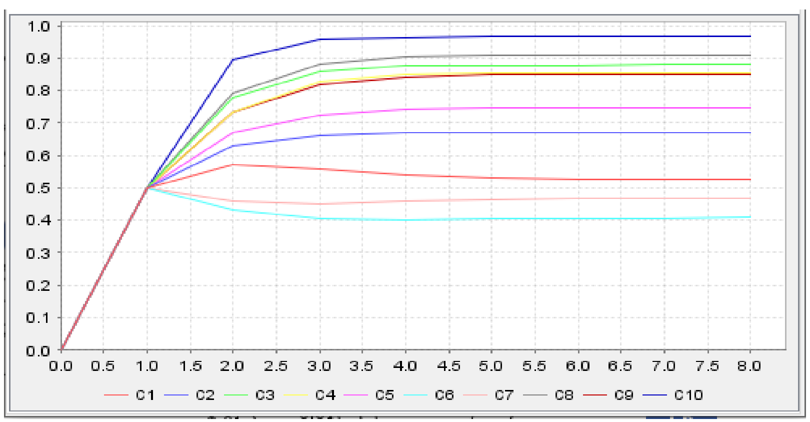
- **A local centrality measure determined by only its directed connections.**

Concepts	**Outdegree**	**Indegree**	**Centrality**
C1	2.06	2.40	4.46
C2	0.64	0.08	0.72
C3	0.49	1.52	2.01
C4	2.78	1.03	3.81
C5	0.03	0.41	0.44
C6	1.75	1.58	3.33
C7	1.70	1.52	3.22
C8	1.44	1.66	3.10
C9	0.92	1.02	1.94
C10	0.72	1.31	2.03

Inference until Convergence

- We performed **inference using various reasoning rules** in order to compute the **output vector including the weights of the concepts**:
  - Kosko's activation rule
  - Kosko's activation rule with self-memory
  - Rescaled activation rule with self-memory.

Indicative visualization of the iterations until convergence



Step	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
2	0.572	0.6318	0.779	0.734	0.6693	0.4305	0.4601	0.7908	0.733	0.8952
3	0.558	0.6628	0.8576	0.8257	0.7252	0.4042	0.4506	0.8832	0.8165	0.9563
4	0.5406	0.6695	0.8746	0.849	0.7414	0.4026	0.4571	0.9037	0.8414	0.964
5	0.5319	0.6708	0.8781	0.8543	0.7457	0.4052	0.4633	0.908	0.8477	0.9652
6	0.5283	0.6711	0.8789	0.8555	0.7467	0.4068	0.4665	0.9089	0.8491	0.9653
7	0.5269	0.6711	0.879	0.8557	0.7469	0.4075	0.468	0.9091	0.8494	0.9653
8	0.5263	0.6711	0.8791	0.8558	0.747	0.4078	0.4685	0.9092	0.8495	0.9653

Comparative Analysis for the Output Weight Vector

- The outcome of the non-linear Hebbian rule varies significantly compared to the outcomes of differential Hebbian learning and balanced differential Hebbian learning.
  - However, all the implementations result in the same order of significance.
    - For this example: C8 – C3 – C4 – C9 – C5 – C2 – C1 – C7 – C6.

Concepts	Non-linear Hebbian Learning	Differential Hebbian Learning	Balanced Differential Hebbian Learning
C1	0.5825	0.6466	0.6674
C2	0.6712	0.6624	0.6663
C3	0.8266	0.7079	0.6851
C4	0.8090	0.6942	0.6757
C5	0.7256	0.6740	0.6713
C6	0.4731	0.6131	0.6556
C7	0.5145	0.6211	0.6572
C8	0.8609	0.7133	0.6860
C9	0.8008	0.6920	0.6730
C10	0.9200	0.7542	0.6997

Outline

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

Conclusions & Future Work

- This paper proposed an approach for **hotel quality evaluation from online review comments and ratings** using FPM for mining customers' opinions and FCM for **evaluating the attributes that contribute to the review rating**.

- 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!