

**In the name of God**





دانشکده برق و کامپیوتر

## Extracting User's Food Preferences by Sentiment Analysis

**Authors:**

**Elham Asani**

**Dr. Hamed Vahdat nejad**

**Saeed Hosseinabadi**

**Dr. Javad Sadri**



# Table of Content

- ❑ Introduction
- ❑ Research problem
- ❑ Related work
- ❑ Proposed method
- ❑ Evaluation
- ❑ Conclusion

# The rapid growth of online reviews on TripAdvisor websites

Introduction

Research problem

Related work

Proposed method

Evaluation

Conclusion



Introduction

Research  
problem

Related work

Proposed  
method

Evaluation

Conclusion

**The problem of this research is presentation of a method for extracting user's food preferences**

3

21

Introduction

Research  
problem

Related work

Proposed  
method

Evaluation

Conclusion

<b>The methods of extracting user preferences</b>	1-questionnaires	i 2 p: a restaurant recommendation system using preference queries over incomplete information
	2-star ratings	Personalized restaurant recommendation method combining group correlations and customer preferences
	3-individuals' text comments	Buon Appetito - Recommending Personalized Menus

# Overall schema of proposed method

Introduction

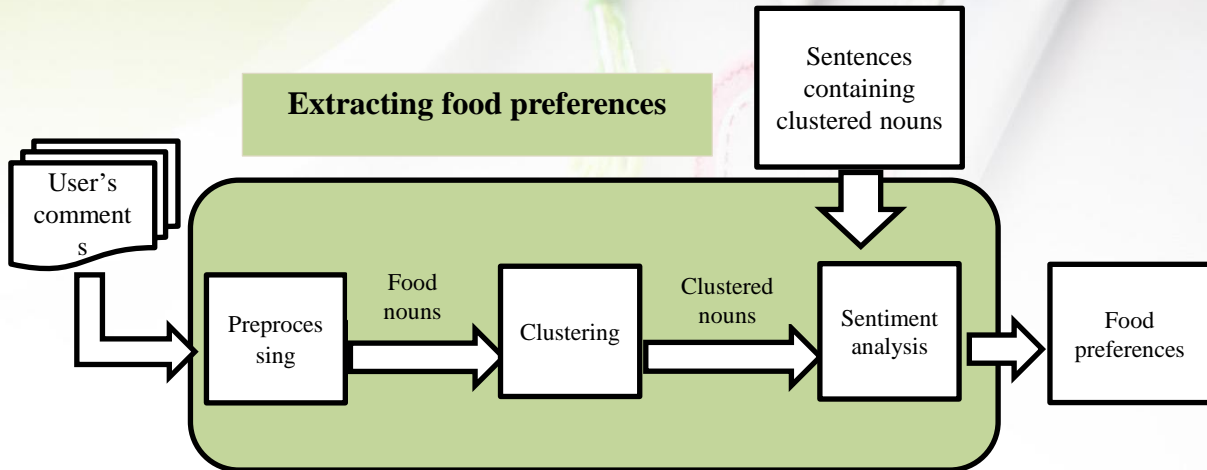
Research problem

Related work

Proposed method

Evaluation

Conclusion



# 1- Pre-processing

Introduction

Research problem

Related work

Proposed method

Evaluation

Conclusion





# 2-Clustering

Introduction

Research  
problem

Related work

Proposed  
method

Evaluation

Conclusion

food nouns

## Agglomerative hierarchical clustering

Calculating the  
distance  
between clusters  
using semantic  
similarity

Create a similarity  
matrix between  
nouns

Combining the  
clusters with  
most similarity  
(more than a  
threshold)

clustered nouns

# 3-Sentiment Analysis

Introduction

Research  
problem

Related work

Proposed  
method

Evaluation

Conclusion

Sentences containing  
clustered nouns

## Sentiment Analysis

Feature selection

Calculating the score of each feature

Calculating the score of each sentence

Calculating the score of each cluster

# 3-Sentiment Analysis

## 3-1: Feature selection

Introduction

Research  
problem

Related work

Proposed  
method

Evaluation

Conclusion

Part Of Speech  
tagging

Adverb

Adjective

Verb

Noun

Preposition

Pronoun

Conjunction

Adverb,  
Adjective

# 3-Sentiment Analysis

Introduction

Research  
problem

Related work

Proposed  
method

Evaluation

Conclusion

## 3-2: Calculating the score of each feature

$$\text{score}_{\text{pos}}(A) = \frac{1}{n} \sum_{i=1}^n \text{score}_{\text{pos}}(i) \quad (1)$$

$$\text{score}_{\text{neg}}(A) = \frac{1}{n} \sum_{i=1}^n \text{score}_{\text{neg}}(i) \quad (2)$$

$$\text{score}_{\text{obj}}(A) = \frac{1}{n} \sum_{i=1}^n \text{score}_{\text{obj}}(i) \quad (3)$$

**N= the number of synonyms for each feature**

# 3-Sentiment Analysis

Introduction

Research  
problem

Related work

Proposed  
method

Evaluation

Conclusion

## 3-3: Calculating the score of each sentence

$$\text{score}_{\text{pos}}(A) = \frac{1}{n} \sum_{i=1}^n \text{score}_{\text{pos}}(i) \quad (1)$$

$$\text{score}_{\text{neg}}(A) = \frac{1}{n} \sum_{i=1}^n \text{score}_{\text{neg}}(i) \quad (2)$$

$$\text{score}_{\text{obj}}(A) = \frac{1}{n} \sum_{i=1}^n \text{score}_{\text{obj}}(i) \quad (3)$$

**N= the number of feature in each sentence**

# 3-Sentiment Analysis

Introduction

Research  
problem

Related work

Proposed  
method

Evaluation

Conclusion

## 3-4: Calculating the score of each cluster

$$S_c = \frac{N_p - N_n}{N_p + N_n}$$

P= the number of positive sentences

N= the number of negative sentences

# 4- Evaluation

Introduction

Research  
problem

Related work

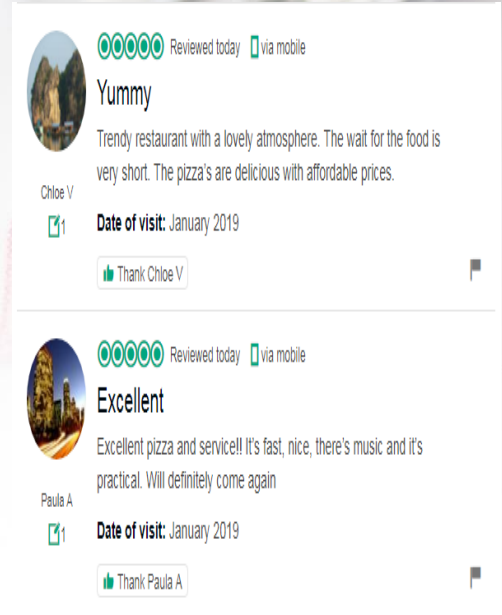
Proposed  
method

Evaluation

Conclusion

## Data gathering

The number of users	period	The number of comments	Resource
100	January to September of 2018	1826	TripAdvisor



The screenshot shows two TripAdvisor reviews. The first review is by 'Chloe V' with a 5-star rating, dated January 2019, describing the restaurant as 'Trendy' with a 'lovely atmosphere' and 'short wait'. The second review is by 'Paula A' with a 5-star rating, also dated January 2019, praising the 'pizza and service' and mentioning 'music'.

**Chloe V** Reviewed today via mobile  
**Yummy**  
Trendy restaurant with a lovely atmosphere. The wait for the food is very short. The pizzas are delicious with affordable prices.  
Date of visit: January 2019  
Thank Chloe V

**Paula A** Reviewed today via mobile  
**Excellent**  
Excellent pizza and service!! It's fast, nice, there's music and it's practical. Will definitely come again  
Date of visit: January 2019  
Thank Paula A

# Various criteria for measuring system accuracy

Introduction

Research  
problem

Related work

Proposed  
method

Evaluation

Conclusion

Precision= what percentage of the foods in the preferences are eaten in the test period

Recall= what percentage of the foods eaten in the test period are in the preferences

F-measure= $2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$



Introduction

Research  
problem

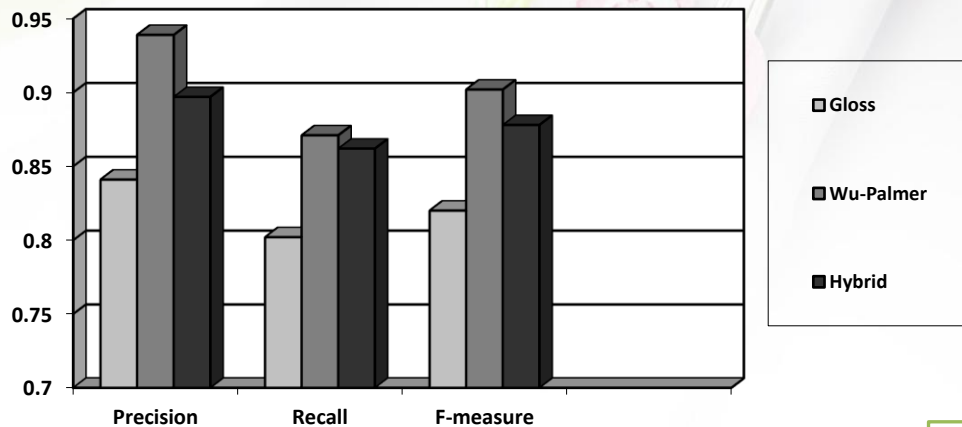
Related work

Proposed  
method

Evaluation

Conclusion

## Comparing the accuracy of extracting user preferences using three different semantic similarity criteria



Introduction

Research  
problem

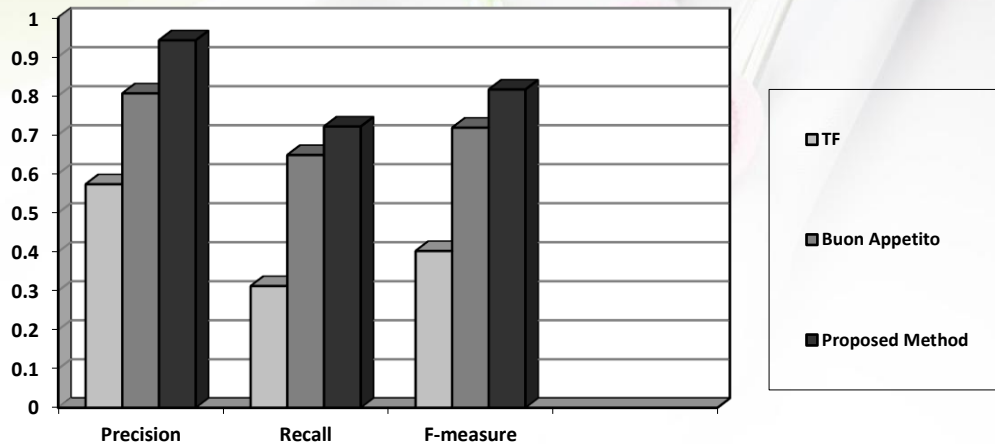
Related work

Proposed  
method

Evaluation

Conclusion

## Comparison results for top 3 using TF, Buon Appetito , Proposed method



Introduction

Research problem

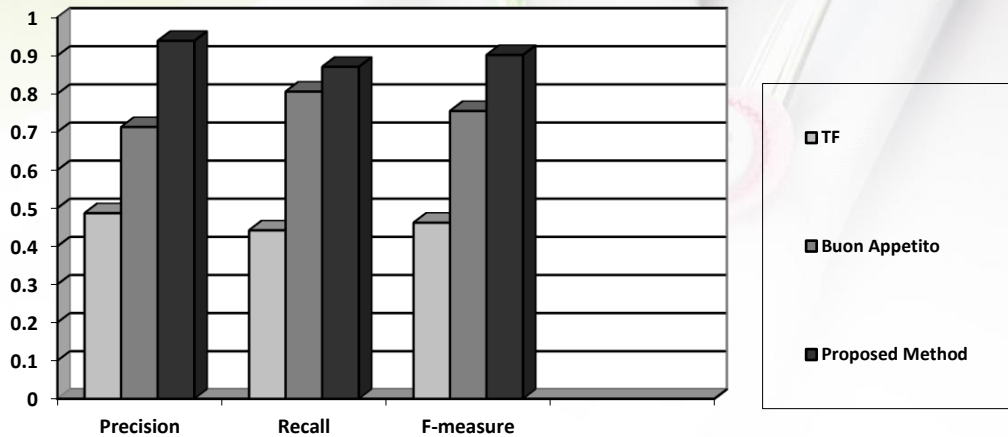
Related work

Proposed method

Evaluation

Conclusion

## Comparison results for top 5 using TF, Buon Appetito , Proposed method



Introduction

Research problem

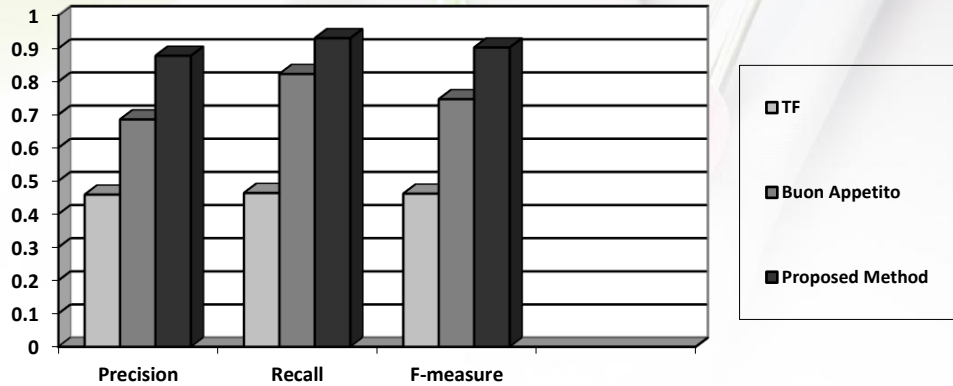
Related work

Proposed method

Evaluation

Conclusion

## Comparison results for top 7 using TF, Buon Appetito , Proposed method



Introduction

Research  
problem

Related work


Proposed  
method

Evaluation

Conclusion

- Increasing the accuracy of clustering using the Wu-Palmer criterion over other methods
- The accuracy of the proposed method is higher than the two methods compared in all modes top3, top5, top7
- The recall of the proposed method is higher than the two methods compared in all modes top3, top5, top7





**Thank you for your  
attention**