Food Recommender System Based on Online Reviews (Postprint)

Authors: Deng Hanxi, Zhihua Chen

Date: 2023-10-08T00:00:00+00:00

Abstract

Owing to the real-time demands of food recommendation, this study proposes a food recommendation system based on online reviews to furnish introductions and review summaries for individual restaurants. Specifically, the system primarily provides functionalities including a web content scraping robot and multi-document automatic summarization technology to automatically extract relevant reviews and blog articles and identify salient review sentences. Finally, the food recommendation system integrates cloud computing technology to implement parallel computation for the multi-document automatic summarization technology, thereby delivering real-time food review services.

Full Text

Preamble

A Food Recommendation System Based on Online Reviews

Hanxi Deng[1], Zhihua Chen[2]
[1]. Office of Domestic Cooperation and Exchange, Communication University of China, Beijing 100024
[2]. College of Computer Science and Big Data, Fuzhou University, Fuzhou, Fujian 350108

Abstract: In response to the real-time demands of food recommendations, this study proposes a food recommendation system based on online reviews to provide restaurant introductions and review summaries. The primary functions of the system include a web content extraction robot and multi-document summarization technology to automatically retrieve relevant reviews and blog articles while extracting important review sentences. Finally, the system integrates cloud computing technology to establish parallel processing for multi-document summarization, enabling real-time delivery of food review services.

Keywords: food recommendation; information retrieval; automatic text summarization; cloud computing
CLC Number: TP391
Document Code: A
Article ID: 1671-0134(2022)03-039-03
DOI: 10.19483/j.cnki.11-4653/n.2022.03.011
Citation Format: Deng H, Chen Z. A Food Recommendation System Based on Online Reviews [J]. China Media Technology, 2022(03): 39-41.

Introduction

In recent years, as people' s quality of life has gradually improved, individuals have become increasingly discerning about food. Not only must the cuisine satisfy customers' palates, but service and pricing must also meet their expectations [1]. Although the internet is now ubiquitous, allowing everyone to share dining experiences online, the sheer volume of reviews from numerous sources makes it difficult to quickly and accurately understand a restaurant.

To address the real-time needs of food recommendations, this paper proposes a food recommendation system based on online reviews called the "Food Situation Broadcasting System."This system provides restaurant introductions and review summaries, enabling people to quickly determine optimal dining locations and avoid disappointing experiences even in unfamiliar environments.

The "Food Situation Broadcasting System" integrates web content extraction robots, multi-document summarization technology (Multiple Document Summarization, MDS) [2], and cloud computing technologies. The web content extraction robot searches for relevant information using search engines like Baidu, identifies relevant food review content across various websites, and stores the crawled and parsed data as a Blog Corpus. Finally, multi-document summarization technology extracts food reviews from the relevant web corpus and generates summary formats to provide users with dining decision references. Users can query the system through a designed human-computer interface. The overall system processing flow is shown in [Figure 1: see original paper].

1.1 Web Content Extraction Robot

The web content extraction robot provides fuzzy search mechanisms, web crawlers, and HTML parsers, with each function described below.

1.1.1 Fuzzy Search Mechanism

The fuzzy search mechanism provides fuzzy operations and judgments, establishing a keyword database for active searching on Baidu.

1.1.2 Web Crawler

The web crawler crawls Baidu search results (such as content from returned web pages), tracks relevant links, and temporarily stores HTML content.

1.1.3 HTML Parser

The HTML parser decodes HTML tags from web pages obtained by the crawler, extracts main information, effectively removes special characters (such as single and double quotes), prevents database injection attacks, and establishes a Web Corpus to facilitate subsequent multi-document summarization inference.

1.2 Multi-Document Summarization Technology

The "Food Situation Broadcasting System" combines multi-document summarization technology to automatically summarize relevant food website reviews in real-time, effectively reducing information volume and extracting key review summaries. This allows users to quickly browse past consumers' opinions and experiences regarding restaurants or cuisine.

The multi-document summarization technology primarily references the MEAD toolkit [3] for system implementation. Food reviews from the web corpus are input into the automatic summarization module. Due to the large data volume requiring efficient and fast parallel computing, this module is implemented on the Hadoop platform using MapReduce. The process involves data preprocessing, feature selection, classifier, reranker, and summary generation steps for automatic extraction, with detailed functional designs described below.

1.2.1 Data Preprocessing

Data preprocessing extracts HTML content processed by the web content extraction robot, sequentially defines document IDs and sentence IDs, and performs weight calculation and summary generation for each sentence.

1.2.2 Feature Selection

The "Food Situation Broadcasting System" primarily employs two features: Thematic Words and Comment Terms.

1.2.2.1 Thematic Words Thematic word scoring calculates the frequency of thematic words in a sentence. Higher frequency indicates stronger relevance to the target theme [4]. For the i-th sentence s_i in a review document containing n_i words w, the thematic word score is calculated as shown in formula (1):

$$\text{ThematicWordScore}(s_i) = \sum_{w \in s_i} \text{score}(w)$$

where $\text{score}(w) = 1$ if w is a thematic word, and 0 otherwise.

1.2.2.2 Comment Terms Comment term scoring calculates the frequency of comment words in a sentence. Higher frequency indicates greater significance as a review statement [5]. For the i-th sentence s_i in a review document containing n_i words w, the comment term score is calculated as shown in formula (2):

$$\text{CommentTermScore}(s_i) = \sum_{w \in s_i} \text{score}(w)$$

where $\text{score}(w) = 1$ if w is a comment term, and 0 otherwise.

1.2.3 Classifier

Different features have varying degrees of importance. The classifier performs weighted summation to calculate the weight of each sentence, as shown in formula (3):

$$\text{SentenceWeight}(s_i) = \alpha \cdot \text{ThematicWordScore}(s_i) + \beta \cdot \text{CommentTermScore}(s_i)$$

where α and β are weighting parameters.

1.2.4 Reranker

The reranker recalculates similarity between sentences, sets a threshold for filtering, extracts important sentences with low mutual similarity, and finally performs extraction according to a set compression rate.

1.2.5 Summary Generation

The summary generation step maps the extracted sentences from the reranker according to document IDs, sentence IDs, and original review files to obtain multi-review automatic summary content. The final results are produced and provided to users for quick browsing and reference.

1.3 Cloud Computing Technology

The internet contains vast and complex web content. Analyzing this content generates substantial computational load due to the number of web pages and volume of text. For performance considerations, this study employs cloud computing for parallel processing, implemented on the Hadoop platform (Chen et al., 2012). Each sentence from review articles is processed separately to quickly calculate sentence scores and identify the most important statements for user decision reference.

2. System Implementation

The system designed in this research is available for general public use. Users can connect to the "Food Situation Broadcasting System" via mobile phones to obtain restaurant introductions and review summaries, enabling quick decisions about optimal dining locations and avoiding disappointing experiences even in unfamiliar environments.

In the "Food Situation Broadcasting System," users can access introductions to various restaurants and related food reviews. This study extracts relevant food review articles through the web content extraction robot and provides food review summaries through multi-document summarization technology to support user dining decisions. As shown in [Figure 2: see original paper], users can input a restaurant name to search for relevant food review summaries. Using "Xiamuni" as an example, when users input the restaurant name and click "Food Reviews," the system provides past consumer experiences, performs automatic text summarization, and extracts important review statements through algorithms for quick user decision reference, as shown in [Figure 3: see original paper]. Finally, a map navigation function ([Figure 4: see original paper]) guides consumers to the restaurant.

This research focuses on users' decision-making needs for restaurant selection, developing a food recommendation system based on online reviews called the "Food Situation Broadcasting System." Combining artificial intelligence and information retrieval technologies, the system facilitates the transformation from "media" to "intelligent media" [6-7], collecting and analyzing restaurant-related information and reviews. The system automatically summarizes online information (such as blogs, iPeen, and other food-related reviews) to help users quickly understand restaurants and evaluate their suitability as dining locations. Future work may apply this system model to review summarization and recommendation information across various industries, such as tourism.

References

[1] Zhou L, Li Q. Research on Huai' an Food Recommendation System Based on LBS Applications [J]. Food Safety Guide, 2021(21): 172-173.

[2] Wang Q, Song Zhang H, Li F. Automatic Summarization Generation Method Based on Multi-Dimensional Text Features [J]. Computer Engineering, 2020(9): 110-116.

[3] Hui-Fei Lin, Chi-Hua Chen, J.M. An Intelligent Embedded Marketing Service System Based on TV Apps: Design and Implementation through Product Placement in Idol Dramas [J]. Expert Systems with Applications, 2013(10).

[4] Luo F, Wang J, He D, Pu Q. Research on Text Automatic Summarization Method Fusing Theme Features [J]. Application Research of Computers, 2021(1): 129-133.

[5] Zhang C, Tong T, Zhou Q. Research on Automatic Book Review Summarization Based on Fine-Grained Review Mining [J]. Journal of the China Society for Scientific and Technical Information, 2021(2): 163-172.

[6] Xu M. From "Media" to "Intelligent Media" : The Impact of Artificial Intelligence Technology on Journalism [J]. China Media Technology, 2021(7): 56-58.

[7] Wu Z. The Mutual Construction and Domestication of Intelligent Media Technology and Commercial Communication [J]. China Media Technology, 2021(8): 103-105.

Author Profiles: Hanxi Deng (1986-), female, Beijing, lecturer, research interests: cultural industry, film and television media; Zhihua Chen (1984-), male, Taiwan, China, professor, research interests: artificial intelligence, Internet of Things.

(Responsible Editor: Yang Hu)

Note: Figure translations are in progress. See original paper for figures.