

Advanced in Control Engineering and Information Science

## A Cloud-Based Recommender System - A Case Study of Delicacy Recommendation

Chi-Hua Chen<sup>a,\*</sup>, Ya-Ching Yang<sup>a</sup>, Pei-Liang Shih<sup>b</sup>, Fang-Yi Lee<sup>a</sup>, Chi-Chun Lo<sup>a</sup>

<sup>a</sup>Institute of Information Management, National Chiao-Tung University, Taiwan, R.O.C.

<sup>b</sup>Degree Program of Computer Science, College of Computer Science, National Chiao-Tung University, Taiwan, R.O.C.

---

### Abstract

Delicacy recommendation services are the trend of the future. In this paper, we propose an effective *decision support systems* (DSS), the *Cloud-Based Recommender System* (CBRS), which provides the introduction and commentaries of delicacies and restaurants with relevant recommendation. CBRS provides the *web content retrieval agent* (WCRA) and *multiple document summarization* (MDS) technology to generate summary of commentaries. Finally, CBRS combines with the cloud computing for MDS to provide delicacy recommendation services.

© 2011 Published by Elsevier Ltd. Selection and/or peer-review under responsibility of [CEIS 2011]

*Keywords:* Delicacy Recommendation; Web Content Corpus; Automatic Text Summarization; Cloud-Based Recommender System.

---

### 1. Introduction

The rise of the quality of life index together with the improvement of economic growth leads to increase delicacy requirements. Delicacy recommendation services are the trend of the future. In this paper, we propose an effective *decision support systems* (DSS), the *Cloud-Based Recommender System* (CBRS), which is a three-tier system composed of the users, *multimedia application server* (MAS), and *database server* (DS) to provide the introduction and commentaries of delicacies and restaurants with relevant recommendation. CBRS provides the *web content retrieval agent* (WCRA) and *multiple document summarization* (MDS) technology to generate summary of commentaries [2]. Finally, CBRS combines with the cloud computing for MDS to provide delicacy recommendation services.

---

\* Corresponding author. Tel.: +886-975292259

E-mail address: [chihua0826@gmail.com](mailto:chihua0826@gmail.com)

## 2. Related Work

Web contents (e.g., blogs) have increased greatly in recent years because of the rapid development of computer technology and the spread of Internet. Users may get a huge amount of information through the intelligent agency from the blogs. However, it's not easy to filter the useless and repeated information for the users. To solve this problem, CBRS use the MDS to get rid of the repeated information and to simplify search results so the users can save the searching time and get the important information.

We refer MEAD which is a public domain portable multi-document summarization system based on Linux [5] to design MDS modules. The processes of MEAD are implemented in Perl programming language. The main procedures of MEAD are as follows.

(1) *Preprocess*: The intelligent agent to retrieve the web contents using HTML format to segment the sentences in original document in order to facilitate follow-up to the weight computing [2].

(2) *Feature selection*: In this paper, MDS is designed to consider several features to compute the weight of each sentence by words and phrases. The main three features are centrality, sentence length, and position [2, 4-7].

(3) *Classifier*: The scores of every sentence are computed through the weight with each feature [3].

(4) *Reranker*: Because the Classifier is carried out only in accordance with score of sentence similarity calculation and sorting, there is often high similarity between sentences, especially in multi-document summarization. MEAD designs a Reranker mechanism to recalculate the sentence with the syntactic similarity and set the threshold to filter out high similarity sentences to reduce the redundancy ratio. Finally, the summary is done by extracting the sentences from original document by the compression ratio [4].

(5) *Summarization*: Summarization can retrieve and recombine words and phrases in the original document according to the order of the sentences by Reranker sorting.

(6) *Evaluation*: CBRS is used to measure the performance of text summarization system including the effect of output results as well as users' satisfaction.

## 3. System Design Principles

In this paper, the design of the CBRS provides functions which are WCRA, MDS, and cloud computing algorithm. WCRA searches the web contents using Google blog search engine and ippeen.com search engine, and it finds comments about the delicacies in web pages and stores the crawl and parse into web content corpus. CBRS uses MDS technology to provide the introduction and commentaries of delicacies and restaurants with relevant recommendation [2]. Finally, cloud computing algorithm can support MDS to summarize delicacy comments rapidly. Users then use the system interface to query relevant information. Overall system processes are shown as Figure 1.

### 3.1. Web Content Retrieval Agent

The WCRA provides functions which are fuzzy search, HTML crawler, and HTML parser. The functions are shown as follows.

(1) *Fuzzy search*: Fuzzy search provides fuzzy computing and judge. It establishes the keywords corpus and uses the terms in corpus to search the articles via Google blog search engine and ippeen.com search engine.

(2) *HTML crawler*: HTML crawler is used to create a copy of all the visited web pages for later processing by a fuzzy search. In this paper, CBRS uses the results of Google blog search and ippeen.com search in various web contents and track related page links which HTML contents.

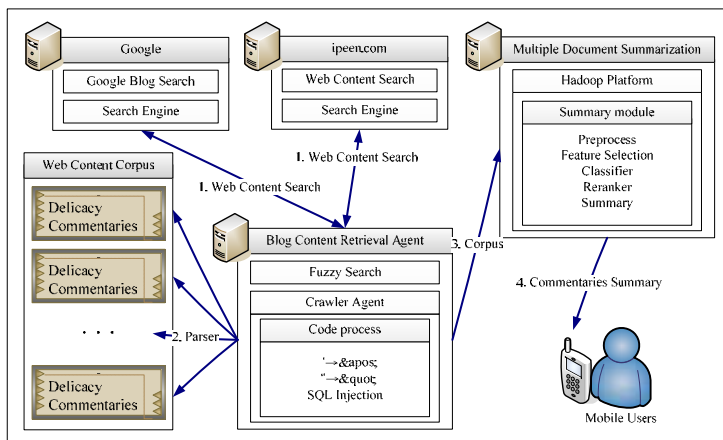


Fig. 1. System architecture of the CBRS

(3) *HTML parser*: HTML parser analyzes the HTML tags generated from HTML crawler to get the key information. After that, it removes the special characters (such as single quotes and double quotes), and avoid hacking attacks. Finally, we would establish web content corpus to get the summarization from multiple documents to provide relevant recommendation.

### 3.2. Multiple Document Summarization

CBRS combines MDS technology to summarize automatically the various delicacy comments in real-time and reduce the amount of information effectively so that users can quickly browse the tourist of consumers' point of view and the past experience.

We refer the procedures of MEAD to design MDS modules and combine the cloud computing algorithm in CBRS. The relevant good comments in the web content corpus are inputted into the MDS modules which are (1) Preprocess, (2) Feature Selection, (3) Classifier, (4) Reranker, and (5) Summary to get text summarization generated automatically.

#### (1) Preprocess

In first step, preprocess would transfer the format of original HTML documents from a web content entry. And then, we set the documents ID and Sentence ID sequentially in order to calculate the weight of sentences in each document for the summary.

#### (2) Feature Selected

After that, CBRS uses features including (i) thematic terms and (ii) comment terms to calculate the weight of each sentence.

1). *Thematic terms*: There are  $n$  terms in the sentence  $s$ . If the  $i$ -th word  $w_i$  is a thematic term, the score  $a_i$  set to 1. Otherwise the score  $a_i$  set to 0.

$$Score_{Thematic}(s) = 1 + \sum_{i=1}^n a_i, \text{ where } a_i = \begin{cases} 1, & \text{word } w_i \text{ in } s \text{ and } w_i \in \text{thematic terms} \\ 0, & \text{otherwise} \end{cases}$$

(1)

2). *Comment terms*: There are  $n$  terms in the sentence  $s$ . If the  $i$ -th word  $w_i$  is a comment term, the score  $b_i$  set to 1. Otherwise the score  $b_i$  set to 0.

$$Score_{Comment}(s) = 1 + \sum_{i=1}^n b_i, \text{ where } b_i = \begin{cases} 1, & \text{word } w_i \text{ in } s \text{ and } w_i \in \text{comment terms} \\ 0, & \text{otherwise} \end{cases}$$

(2)

(3) Classifier

In the third step, we select and consider some important features to set the different weights according to these features. We summarize these features and describe how we use their weights to calculate the overall score of each sentence. The expression shows in the following.

$$Score(s) = Score_{Thematic}(s) \times Score_{Comment}(s)$$

(3)

(4) Reranker

In this step, we would judge the correlation in sentences which is decreased the redundancy. Next, we would set the threshold for filtering, and set the compression to extract.

(5) Summery

After that, we get the extract from Reranker and map the extract to summarize from document ID and sentence ID in this step of preprocess. Finally, MDS technology is used to provide summary of commentaries for users.

### 3.3. Cloud Computing Algorithm

There are a large number of web contents which are distributed in different web sites. We analyze these web contents to infer and summarize the delicacy comments. For performance improvement, we use Hadoop platform to build cloud computing and parallel processing environments [1]. Then we can implement MapReduce program to analyze and compute the scores of sentences in each comment document for user’s delicacy decision support [1].

Each Mapper uses Eq. (1), (2), and (3) to compute the  $Score(s_j)$  of sentence  $s_j = \{w_1^j, w_2^j, \dots, w_n^j\}$  in document  $d_i$  shown as Figure 2. There are total  $y$  sentences in  $x$  documents. After computing the score of each sentence, we sort these sentences by their score and select top  $n$  sentences to generate the delicacy comment summarization.

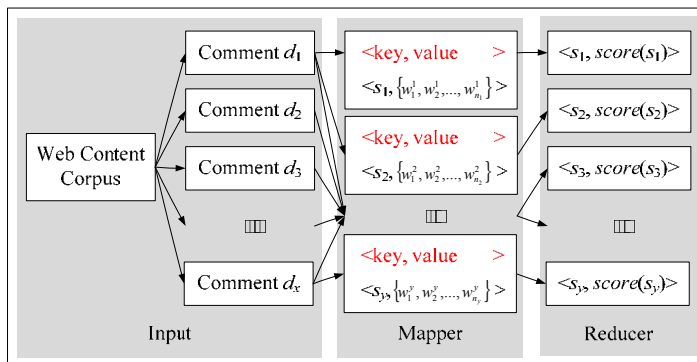


Fig. 2. Cloud Computing Algorithm

## 4. System Usage

Mobile users can utilize various terminal devices that include *personal computer* (PC), notebook, tablet PC, *personal digital assistant* (PDA), and smart phone to access CBRS to get the delicacy comment summarization and delicacy recommendation services. The user can acquire (i) the delicacies and restaurants introduction and (ii) the relevant comments of each delicacy. For example, mobile user inputs and submits his requirement “千葉火鍋” into CBRS shown as Figure 3. CBRS then retrieves the relevant

comments from web contents via WCRA and provides the delicacy comment summaries via MDS based on cloud computing to help mobile users make their delicacy decision shown as Figure 4.



Fig. 3. Query Message



Fig. 4. Delicacy Summary

## 5. Conclusions

This paper focuses on providing restaurant information and comments for delicacy recommendation services and describes the CBRS, an integrated service platform, which provides relevant recommendations for finding delicacy recommendation services. CBRS provides the WCRA and MDS technology to generate summary of commentaries. Finally, CBRS combines with the cloud computing for MDS to provide delicacy recommendation services.

## References

- [1] B.Y. Lin, C.H. Chen, H.C. Chang, C.C. Lo, "A Network Behavior Analysis System for Cloud Computing Service", *Information-An International Interdisciplinary Journal*, Vol. 14, No. 3, pp. 931-938, 2011.
- [2] C.C. Lo, C.H. Chen, D.Y. Cheng, H.Y. Kung, "Ubiquitous Healthcare Service System with Context-awareness Capability: Design and Implementation", *Expert Systems with Applications*, Vol. 38, No. 4, pp. 4416-4436, 2011.
- [3] K. Kaikhah, "Text Summarization Using Neural Networks", *WSEAS Transactions on Systems*, Vol. 3, No. 2, pp. 960-963, 2004.
- [4] D.R. Radev, T. Allison, S. Blair-goldensohn, J. Blitzer, A. Çelebi, S. Dimitrov, E. Drabek, A. Hakim, W. Lam, D. Liu, J. Otterbacher, H. Qi, H. Saggion, S. Teufel, A. Winkel, Z. Zhang, "MEAD - a platform for multidocument multilingual text summarization", In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, Lisbon, Portugal, 2004.
- [5] D.R. Radev, H. Jing, M. Budzikowska, "Centroid-based summarization of multiple documents: sentence extraction, utility-based evaluation, and user studies," the *Proceedings of the ANLP/NAACL 2000 Workshop on Automatic Summarization*, 2000.
- [6] D.R. Radev, A. Winkel, and M. Topper, "Multi Document Centroid-based Text Summarization", In *ACL 2002*, Philadelphia, PA, 2002.
- [7] J.Y. Yeh, H.R. Ke, W.P. Yang, and I.H. Meng, "Text Summarization Using a Trainable Summarizer and Latent Semantic Analysis", *Information Processing and Management*, Vol. 41, No. 1, pp. 75-95, 2005.