Chapter 72
Text Mining for Analysis of Interviews and Questionnaires

Yasuhiro Yamada
Shimane University, Japan

Kanji Kato
GK Sekkei Incorporated, Japan

Sachio Hirokawa
Kyushu University, Japan

ABSTRACT

Interviews and questionnaires are the basis for collecting information about the opinions, concerns and needs of people. Analysis of those texts is crucial to understand the kansei of people. Text mining is an approach to discover useful and interesting patterns, knowledge and information from texts. This chapter contains two sections on text mining for beginners of it. The first section gives a brief survey of basic text mining techniques, such as keyword extraction, word graphs, clustering of texts and association rule mining. The second section demonstrates an example of text mining applied to interview analysis. Two text mining systems - the concept graph system and the matrix search system - are applied to analyze 2,409 remarks about products and services from 19 people. The analysis shows that text mining systems with a search function achieve interactive analysis of texts and an examination of various problems that we targeted.

1. INTRODUCTION

There are various approaches to understand the kansei of people. Brain waves, body temperatures and blood pressure are typical physiological data being used to measure kansei. Features of kansei appear in such data. Kansei are also reflected by the language that people use. This chapter deals with the study on patterns of language usage to understand kansei.

In order to understand opinions, concerns and needs of people, interviews and questionnaires are used as a standard approach. These opinions, concerns and needs are deeply related to the kansei of people. Sometimes, they are held in common by many people. At other times, they are unique to

an individual. The kansei are implied by the form of the language, i.e. in sentences and in words. However, the kansei itself are behind these words and are not always expressed directly.

Interviews and questionnaires are stored in texts that can be analyzed by computers. Thanks to the progress of second storage capacity and the spread of the World Wide Web in the last decade, we can store and share a large amount of texts. Databases and search engines for texts help us to search easily and quickly. However, it is beyond the capability of one person to read, remember and manually analyze a large number of texts. At the very least, it requires a lot of time.

Even if an analyst was able to analyze a large amount of texts arising from interviews, the results may reflect the analyst's opinion and viewpoint. For example, they might depend on the analyst's personal knowledge, background and interest. Thus, if several researchers analyze a large amount of texts, the analyses are likely to be inconsistent from certain viewpoint. To eliminate bias, objective evidence is important. Traditionally, statistical methods have been used to support analysis. However, numeric analysis needs to be further interpreted. It should be explained as texts that can be readily understood.

Text mining is a technique to discover useful and interesting patterns, knowledge and information from texts. It is often able to discover information of which we are unaware. The results of mining are objective in the sense that they are obtained by the same method or algorithm even if input texts or analysts are different. Text mining techniques and systems are expected to help us analyze a large number of texts efficiently. Various such techniques and systems have been proposed.

This chapter contains two sections on text mining. The chapter assumes about readers which are going to use text mining for their research but do not know about it. The first section gives a brief survey of basic text mining techniques such as keyword extraction, word graphs, clustering of texts and association rule mining. Text mining

techniques are classified into two types - statistical and semantic methods. Statistical methods are based on the frequencies of words in the texts. Semantic methods consider the meanings of words and how the words feature in the sentence structure. The first section describes also the statistical methods. This section is useful to start studying text mining.

The second section demonstrates an example of text mining applied to interview analysis. Two text mining systems - the concept graph system and the matrix search system - are applied to analyze interview. Search keywords are used, as with search engines, to specify the target and the circumstances of the analysis. The concept graph system visualizes a hierarchy of characteristic words in the texts. We can understand important words and relations among them in the texts before we actually read the texts. The matrix search system displays a matrix of the distribution of clustering of the texts from two different aspects. We can interpret the relation of the two aspects in the texts by combining various aspects. The two systems achieve multiple and dialogical analysis of interview. Therefore, the systems enable us to examine various problems that we choose as targets.

2. TEXT MINING TECHNIQUE

Preliminaries

This section reviews some notation used to describe text mining techniques. Let $D=\{d_1, d_2, \ldots, d_n\}$ be a set of texts or documents. The number of texts in D is denoted by $|D|$. If w is a word that appears in some text d in D, we write $w \in D$. The term frequency of w in D, denoted by $tf(w, D)$, is the total number of occurrences of w in D. The document frequency of the word w in D, denoted by $df(w, D)$, is the number of texts in D that contain w. The number of texts containing both words u and v in D is denoted by $df(u*v, D)$.

Related Content

Anomaly Region Detection Based on DMST
Sulan Zhangand Jiaqiang Wan (2019). *International Journal of Data Warehousing and Mining (pp. 39-57).*
www.irma-international.org/article/anomaly-region-detection-based-on-dmst/223136

Scalable Biclustering Algorithm Considers the Presence or Absence of Properties
Abdelilah Balamane (2021). *International Journal of Data Warehousing and Mining (pp. 39-56).*
www.irma-international.org/article/scalable-biclustering-algorithm-considers-the-presence-or-absence-of-properties/272017

Rule-Based Data Mining Cache Replacement Strategy
Ramzi A. Haratyand Joe Zeitouny (2013). *International Journal of Data Warehousing and Mining (pp. 56-69).*
www.irma-international.org/article/rule-based-data-mining-cache/75615

Case Studies on the Use of Data Mining Techniques in Data Science
 (2023). *Principles and Theories of Data Mining With RapidMiner (pp. 193-204).*
www.irma-international.org/chapter/case-studies-on-the-use-of-data-mining-techniques-in-data-science/323375

A Decision Support System for Privacy Compliance
Siani Pearsonand Tomas Sander (2013). *Data Mining: Concepts, Methodologies, Tools, and Applications (pp. 1496-1518).*
www.irma-international.org/chapter/decision-support-system-privacy-compliance/73508