Incremental Set Recommendation Based on Class Differences

Yasuyuki Shirai\(^1\), Koji Tsuruma\(^1\), Yuko Sakurai\(^2\), Satoshi Oyama\(^3\), and Shin-ichi Minato\(^1,3\)

\(^1\) JST-ERATO MINATO Discrete Structure Manipulation System Project, Hokkaido University, Sapporo, Japan
\{shirai,tsuruma\}@erato.ist.hokudai.ac.jp

\(^2\) Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka, Japan
ysakurai@inf.kyushu-u.ac.jp

\(^3\) Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Japan
\{oyama,minato\}@ist.hokudai.ac.jp

Abstract. In this paper, we present a set recommendation framework that proposes sets of items, whereas conventional recommendation methods recommend each item independently. Our new approach to the set recommendation framework can propose sets of items on the basis on the user’s initially chosen set. In this approach, items are added to or deleted from the initial set so that the modified set matches the target classification. Since the data sets created by the latest applications can be quite large, we use ZDD (Zero-suppressed Binary Decision Diagram) to make the searching more efficient. This framework is applicable to a wide range of applications such as advertising on the Internet and healthy life advice based on personal lifelog data.

Keywords: recommendation, classification, collaborative filtering, zero-suppressed binary decision diagram.

1 Introduction

Several techniques on information filtering and information recommendation such as collaborative filtering and content-based filtering have been reported \([5][1][11]\). In conventional collaborative filtering, items are recommended on the basis of their relevance to the user’s preferences. Each item is recommended independently of the others; that is, the relationship of a recommended item to the other items is not considered.

In the real world, however, a user is often interested in a combination of items, such as the keywords in an advertisement and the places to be visited during a sightseeing tour. Recently proposed set recommendation techniques \([12][10]\) consider the unit of recommendation to be a set of items and the constraints and requirements among them.
In this paper, we extend this approach to incorporate the use of an algorithm to present recommendations for modifying the user’s initially chosen set. In our *incremental set recommendation* framework, it is assumed that each record (“item set”) in a database has been classified as a class such as positive/negative, and modifications are recommended that would change the item set so that it matched the target classification.

An example application of our framework is a recommendation system that uses a database in which the action history data for a group of people are stored. The data could be exercise history or dietary behavior, for example. Each person in the database is classified as either a success or failure w.r.t. to some target (e.g., weight loss). Those in the ”failure” group could use the system to obtain recommendations for specific behavior improvements that are based on the data for those in the ”success” group. The recommendations are made on the basis of the differences between the two groups and should change the user’s actions and lifestyle as little as possible. ”Behavior improvements” for the exercise history example means the addition and/or deletion of item sets representing the type and amount of exercises performed, while for the dietary behavior example, it means the addition and/or deletion of item sets representing the type and quantity of food eaten. Another example application is a system for describing the items for sale on an Internet shopping site. The descriptions of poorly selling items would be modified on the basis of the descriptions of items that sell well.

The rest of the paper is organized as follows: Section 2 gives the basic definitions. In Section 3 we describe the implementation of our framework using a Zero-suppressed Binary Decision Diagram (ZDD) data structure. We present and discuss the results of its evaluation in Section 4. We conclude in Section 5 with a brief summary, some additional comments, and a mention of future work.

2 Definition

We will provide some definitions and notations as follows:

**Definition 1 (Item).** An item is an atomic entity that represents a characteristic or feature and is denoted by a lower-case character, a, b, c, . . .. A set of all items to be considered is denoted by $\Sigma$.

In the exercise history example, each item could be the name of an exercise.

**Definition 2 (Data Record and Class).** A data record is a collection of items that represent the attributes or characteristics of the target object (we use $D$ to represent a data record). A class is a name for a set of data records, and is denoted by $\alpha$, $\beta$, $\gamma$, $\omega$ or $\phi$. Each data record belongs to only one class.

“Positive” or “Negative” is an example of a class.

**Definition 3 (Pattern Set/Class Membership).** A pattern set is a set of pairs, each of which consists of an item set and its weight (natural number). If the weight values are all the same, they can be omitted. A pattern set is denoted by $C_\omega$ where $\omega$ is the class identifier ($C_\omega = \{ p : w_p | p \in 2^\Sigma, w_p \in \mathbb{N} \}$). If $q : w_q \in C_\omega$ (simply we write $q \in C_\omega$), $q$ is called a pattern of class $\omega$. 