Comput Syst Sci & Eng (2019) 4: 231–236

International Journal of
Computer Systems
Science & Engineering

Personalized Product Recommendation Model Based on User Interest

Jitao Zhang*

Jiangxi University of technology

The scale of e-commerce systems is increasing and more and more products are being offered online. However, users must find their own desired products among a large amount of unrelated information, which makes it increasingly difficult for them to make a purchase. In order to solve this problem of information overload, and effectively assist e-commerce users to shop easily and conveniently, an e-commerce personalized recommendation system technology has been proposed. This paper introduces the design and implementation of a personalized product recommendation model based on user interest. The "shopping basket analysis" functional model centered on the Apriori algorithm uses the sales data in the transaction database to mine various interesting links between the products purchased by the customers. Moreover, it helps merchants to formulate marketing strategies, reasonably arranges shelf-guided sales, and attracts more customers. This platform adopts a B/S structure and uses JSP+AJAX technology to achieve the dynamic loading of pages. In the background, the Struts2 framework is combined with the SQL Server database to establish the system function module, and then the function is gradually improved according to the requirement analysis, and the development of the platform is finally completed.

Keywords: Personalized Product Recommendation; User Interest; Association Rules

1. INTRODUCTION

At present, the number of users of e-commerce websites is rapidly increasing. As a new type of sales and shopping channel, the e-commerce model is a web-based sales platform. It provides convenient ways for merchants to sell and customers to purchase, but traditional e-commerce generally only lists products with aggregated information. They cannot provide clear points of personal interest for users browsing products, and cannot recommend products to customers. Therefore, how to recommend products of interest to customers through personalized recommendation services [1] becomes a problem that all major e-commerce websites need to solve. There are many personalized recommendation technologies, some of which are described below.

The collaborative filtering recommendation is quite a successful technique in the personalized recommendation system [2], and is used more in e-commerce. Users' similar purchasing behaviors are used to recommend products to other users. The recommendation idea is based on collaborative filtering by calculating similarity: when two similar users purchase a product, they also recommend the same product to another user. With the collaborative recommendation technology [3], it is necessary to make recommendations based only on the user's purchase behavior without paying attention to the type of product. However, with the increase of website structure, content complexity and the number of users, collaborative filtering algorithms encounter real-time, data sparsity, scalability and cold start issues.

Content-based recommendation [4] addresses the problem of quickly and efficiently obtaining user interest information from a massive amount of data. For this purpose, a content-based personalized recommendation algorithm has been proposed. The content-based recommendation algorithm is based on obtaining a common interest user group, firstly adopting a content similarity algorithm to mine a content recommendation item, and then obtaining a new content recommendation item

*Email: zjtbgs@qq.com

Table 1 Market basket trading table

Transaction ID	Trading items
T1	{beef, cabbage}
T2	{salvage, diaper, beer, egg}
T3	{beef, diaper, beer, apple}
T4	{salvage, milk, diaper, beer}
T5	{salvage, beef, diaper, apple}

according to the user's interest. The basic idea is to establish a relevant configuration file for the user and the product, and update the configuration file according to the user's purchase behavior and the type of product. The interest document is constructed based on the user's history information, and the degree of similarity between the document and the product is calculated, thereby generating a recommendation for the user.

There are many mining techniques based on user interest. The most classic mining algorithm for association rules is Apriori [5][6]. This algorithm first discovers all frequent itemsets, and then generates association rules from frequent itemsets. The application of association rules in data mining is highly suitable for finding meaningful relationships between data in large data sets. One of the reasons is that it is not limited by the choice of only one dependent variable. This paper will use the Apriori algorithm of association rules to analyze the user's browsing and purchasing interest patterns and establish a user behavior model. This study builds on the design and implementation processes of the personalized recommendation system for e-commerce. Valuable products can be promoted to consumers by improving the efficiency of data mining [7]. According to the user's characteristics and purchase behavior, it offers the user any information of interest and provides a commodity recommendation. By finding the association rules among the items in the sales transaction database, the Apriori algorithm of association rules is used to analyze and establish the user behavior model of the user's browsing and purchasing interest patterns. By using the market basket analysis in association rules [8] and the prediction algorithm [9], the empirical evaluation shows that the algorithm is better than the known algorithm of a large database. The experiment shows that the algorithm has a linear relationship with the number of transactions, effectively enhancing the relevance of the commodity information, and improving the ability of the e-commerce website to attract customers and sales.

2. ASSOCIATION RULE ALGORITHM

Association rule definition: Assuming that $I = \{i_1, i_2, i_3, \ldots i_n\}$ is a collection of items, given a transaction database $D = \{\mathrm{t}_1, \mathrm{t}_2, \mathrm{t}_3, \ldots, \mathrm{t}_\mathrm{n}\}$, where each transaction t is a non-empty subset of I, that is, $\mathrm{t} \in I$, each transaction has a unique identifier TID (Transaction ID) correspondence. An association rule is an implication of the form $X \Rightarrow Y$, where $X, Y \in I$ and $X \cap Y = \phi$, X and Y are called the antecedent or left-hand-side (LHS) and consequent or right-hand-side (RHS) of the association rule, respectively [6]. The support degree of the association rule $X \Rightarrow Y$ in D is the percentage of the transaction containing $X \cup Y$ in D, that is, the probability $P(X \cup Y)$. Confidence is the percentage of Y that contains both X, that is, the conditional probability $P(Y|X)$. If the minimum support threshold and the minimum confidence threshold are satisfied, the association rule is called interesting [10]. Simply, given a transaction data set T, find out all the association rules that SUPPORT >= MIN_SUPPORT, CONFIDENCE >= MIN_CONFIDENCE [11].

3. SYSTEM STRUCTURE DESIGN

In this paper, a personalized product recommendation model based on user interest is designed. Based on the weighted moving average method and market basket analysis method, the association rule mining algorithm and prediction algorithm are designed. The model uses a three-tier architecture [12]: the presentation layer, the business layer, and the persistence layer. The connection is set up between the controller and the middle tier, and the middle tier is in contact with the database. The model uses a market basket analysis algorithm and a prediction algorithm to generate a recommendation model.

The recommendation model of this paper mainly adopts the market basket analysis method and the prediction algorithm in the association rule mining algorithm to recommend. Ordinary users can use the market basket analysis algorithm to obtain recommendations. System administrators can view the data analysis and statistics generated by the prediction algorithm.

3.1 Association Rules Mining Algorithm Design

The so-called market basket analysis is based on the analysis of the information in the market basket to discover the customer's buying habits, as shown in Table 1.

The table above is the database D of the market basket and contains 5 transactions. Each row represents the item of the market basket (without considering the number of times). Item set I = {beef, cabbage, diaper, beer, egg, apple}.

For the association rules: diaper ⇒ beer, transaction T2, T3, T4 contain both beer and diaper, support = 3 ÷ 5 = 60 %, confidence = 3 ÷ 4 = 75 %. If min_support = 0.6 and min_confidence = 0.6 are given, it is considered that the associative rule diaper ⇒ beer is interesting; that is, there is a correlation between the diaper and the beer.

$$\text{support(diaper} \Rightarrow \text{beer)} = \text{P(diaper} \cup \text{beer)} \tag{1}$$

$$\text{confidence(diaper} \Rightarrow \text{beer)} = P(\text{beer|diaper}) = \frac{\text{support(diaper} \Rightarrow \text{beer)}}{\text{support(diaper)}} \tag{2}$$

$$\text{lift(diaper} \Rightarrow \text{beer)} = \frac{\text{conf (diaper} \Rightarrow \text{beer)}}{\text{supp(beer)}} = \frac{\text{supp (diaper} \cup \text{beer)}}{\text{supp(diaper) supp(beer)}} \tag{3}$$

The support and confidence of the rules are two ways of measuring the rules. Support is used to determine the frequency of rules in the data set. If the support of an item set is greater than the threshold, this set of items can be called a frequent itemset. Confidence is used to determine how often Y appears in a transaction that contains A.

(1) Find out the item set whose support is greater than the threshold [13];

(2) On the basis of this item set, find the rules whose confidence is greater than the threshold. These rules are called strong association rules [14].

3.2 Prediction Algorithm

A very important theoretical basis for the prediction process is that certain forms of demand models should play basically the same role in the past, present, and future. However, when the actual situation changes abruptly, predictive monitoring determines whether or not the model is still valid at this time.

The so-called tracking signal is the ratio between the RSFE (running sum forecast error) and the average absolute error. The formula is:

$$\text{TS} = \frac{RSFE}{bdMAD} = \frac{\sum_{i=0}^{N}(A_i - F_i)}{MAD} \tag{4}$$

First, the average of the number is taken for the first two days. If the numerical average is higher than the alert value, no matter what kind of trend the data represents, the user's interest in this category of goods is increasing; then, the recommended number in this category is increased. If the numerical average is below the alert value, the moving weighted average is calculated. Compare the arithmetic average with the moving weighted average. If the moving weighted average is less than the numerical average, the data is considered not to increase, and the recommended amount is not increased; otherwise, it assumes that the data shows an increasing trend, and then compares the amount of the previous day with the alert value. If the alert value is exceeded, the recommended amount is increased to the limit value; otherwise, the amount is not increased.

IF(average < alert value)
    Then IF average > moving weighted average
        Then IF the previous day < alert value
Then the recommended amount can be increased to the limit //It is growing, and the previous day has reached the alert value.
    Else do not increase the recommended amount //It is growing, but did not reach alert value the previous day.
    Else no increase in recommendable amount //No growth trend
    Else increase the recommended amount to the limit //The overall level is high

Among them, the commodity recommendation module is an important function of this system, and it plays an important role in perfecting this system. Product recommendation is a personalized recommendation based on users' interests. Following the statistical analysis of user purchase records and traffic records, the previous association rule algorithm is used to find out the rules implied therein, and a prediction algorithm is used to predict the product categories that are of interest to the user. This enables users to see recommended related products when purchasing a product.

3.3 Algorithm Design

(1) First, we need to generate data that can be mined by association rules. We use themoving weighted average method to predict user interest. First, take a day as a benchmark, and then take the access data for the first two days of all product categories, find theirweighted averages, and then compare thesewith actual values to determine trends. This results in data that can be directly mined for association rules. Sid category array

    Result weighted average

    Real actual value

    For I From 0 To sid.length

    Result $\longleftarrow$ weighted average of sid.[I]

    Real $\longleftarrow$ average of sid.[I]

    IF Result>Real

    On this day, the data showed a downward trend. In Excel, the trend value of the two-dimensional table is -1.

    Moreover, on this day, the data showed an upward trend. In Excel, the trend value of the two-dimensional table is 1.

    End IF

    End For

(2) After obtaining statistics for the data generated above, use Excel to calculate the support and confidence, and then screen out the rules of support greater than the threshold, and confidence greater than the threshold. These rules are effective for producing the desired outcome.

This function allows users to find more quickly what they want to buy, which will encourage them to revisit the website. At the same time, product recommendations can also promote the sale of website items. In addition, the system is more intelligent as it automatically recommends goods.

3.4 The Algorithm Implementation of the Recommended Model

The implementation of the association rule mining algorithm relies mainly on Java daemon to evaluate the data and record it in the Excel two-dimensional table. Then Excel is used to calculate the support and confidence, and finally filter out the rules with support and confidence greater than the threshold.

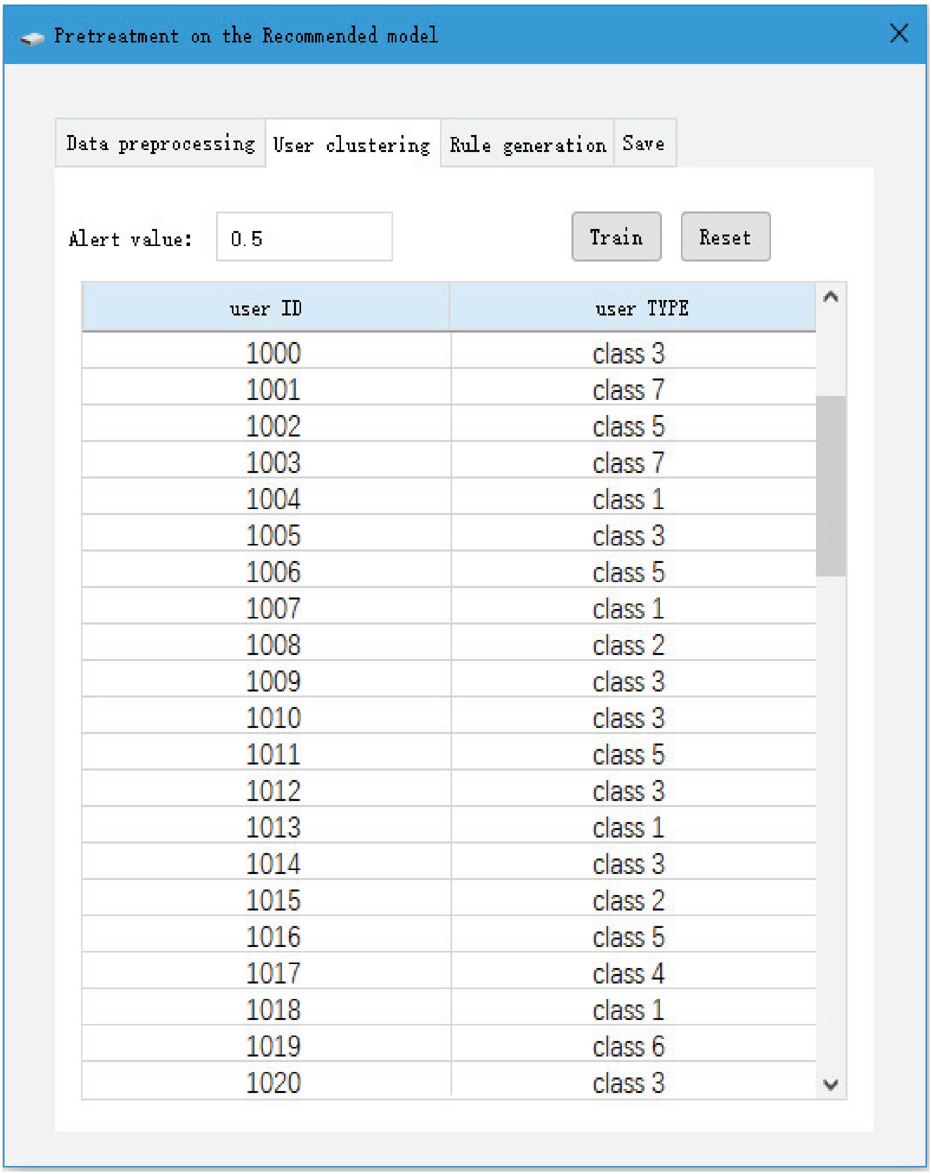

Figure 1 User clustering

These are very strong, and therefore very effective, association rules.

Some formulas used in Excel are:

the number of rising occurrences = COUNTIF (B2:B12, "1")

the number of drop occurrences = COUNTIF (B2:B12, "−1")

support = COUNTIFS(B2:B12,"= 1", C2:C12,"= 1")/ total

confidence = COUNTIFS(B2:B12,"= 1", C2:C12,"= 1")/ the number of rising occurrences

In this paper, 60% is used as a support and confidence threshold.

The specific implementation is shown in Table 2.

4. SIMULATION VERIFICATION

4.1 Data Preparation

The data used in the experimental program included 50 users, 58 products, and 400 randomly-generated transaction records. SQL Server 2000 was used as the database storage, and the related database included four original data forms: a commodity transaction table, transaction detail table, commodity information table, and user information table. The main information contained in each table is as follows: the commodity transaction table contains a large number of records, each record has its lower attributes: transaction number, shopping time and customer number (non-empty for membership) and so on. The data in the transaction detail table consists of the commodity number, commodity quantity, commodity price and transaction number and so on. The commodity information table mainly includes commodity number, commodity name, commodity category, commodity price and so on. The user information table includes user numbers, gender,

Table 2 Mining Results of Association Rules

	women	makeups	Baby Pro	Digital	Home tex	living	men
2017/6/1	1	−1	−1	1	−1	1	1
2017/6/2	−1	1	1	−1	1	−1	−1
2017/6/3	1	1	−1	−1	1	1	1
2017/6/4	1	1	1	1	1	1	1
2017/6/5	1	−1	−1	−1	−1	−1	−1
2017/6/6	1	1	1	1	−1	1	1
2017/6/7	−1	1	−1	−1	1	1	1
2017/6/8	1	1	1	1	−1	1	1
2017/6/9	−1	−1	−1	−1	1	1	−1

Table 3 Relationship

Pre rules	Rules after	supports	confidences
Philips SHB1300	PanasoniMD10000GK	24	0.76
Pinnacle PCTV	PHILIPS	26	0.76
SONY Memory Stick Adapter	Lezhibang E301	27	0.76
CASIO EW -V2000	PHILIPS	29	0.75
Haitiandi camera	UT820 TV stick	28	0.73
…			

age, marital status, monthly income and login times and so on.

4.2 Data Preprocessing

Although the original data contains the main content needed by the project research institute, they still have some deficiencies, such as the absence of data in the table or some errors and so on. Therefore, the data must be preprocessed before the relevant actions are performed.

1) Delete irrelevant data

Select the user attributes that can affect the recommendation, such as the user's gender, age, marital status, monthly income and other attributes, and delete some useless attributes such as the user's name, contact phone number, address and so on, as none of this attribute information will affect the results.

2) Data conversion

The data table used by the original data is not Boolean and needs to be converted accordingly, including the following two aspects:

(1) Discretization of quantized attributes: Some attributes are numerical attributes that need to be divided into several intervals. For example, the "monthly income" field in the user information table is a numerical attribute. According to the distribution law of the value, we divide it into four groups: G1 (less than 1000 yuan), G2 (1000 yuan to 3500 yuan), and G3 (3500 yuan to 7000 yuan), G4 (7000 yuan to 10,000 yuan) and G5 (more than 10,000 yuan). Other numerical attributes can be divided into several intervals according to this method, and they can be represented by corresponding symbols, thereby converting numerical attributes into Boolean types.

(2) Conversion of category attributes: Some attribute values are selected from alternatives, such as the "gender" field in the customer information table, and are further converted into Boolean types as follows: G1 (male) and G2 (female). Other category attributes and so on are further converted to Boolean types.

(3) Using market basket analysis to user clustering: As shown in Figure 1, we set the alert value to 0.5 so that the users will be divided into six categories. The higher the alert value, the more user categories will be generated. The more user categories, the fewer will be the association rules being generated.

(4) Generate association rules: The clustering results and related transaction data are used as mining data, and the relevant support is set to 0.05 (that is, support for at least 20 records in 400 records); the confidence is 0.5; and the mining association rules apply to both the relationship between products, and the relationship between user categories and product items as shown in Table 3.

(5) Online recommendation: When the user logs into the system and purchases or queries related products, he or she will obtain some useful recommendations, such as the following rules in the association rules repository:

Soopen (QQ) QQ camera ⇒ Weifeng WT330A tripod

Soopen (QQ) QQ camera ⇒ Sumdex NTC-103BK black digital camera bag

User category 3 ⇒ SSK All in One card reader

Soopen (QQ)QQ camera ⇒ Tianma 2.0X teleconverter 52mm

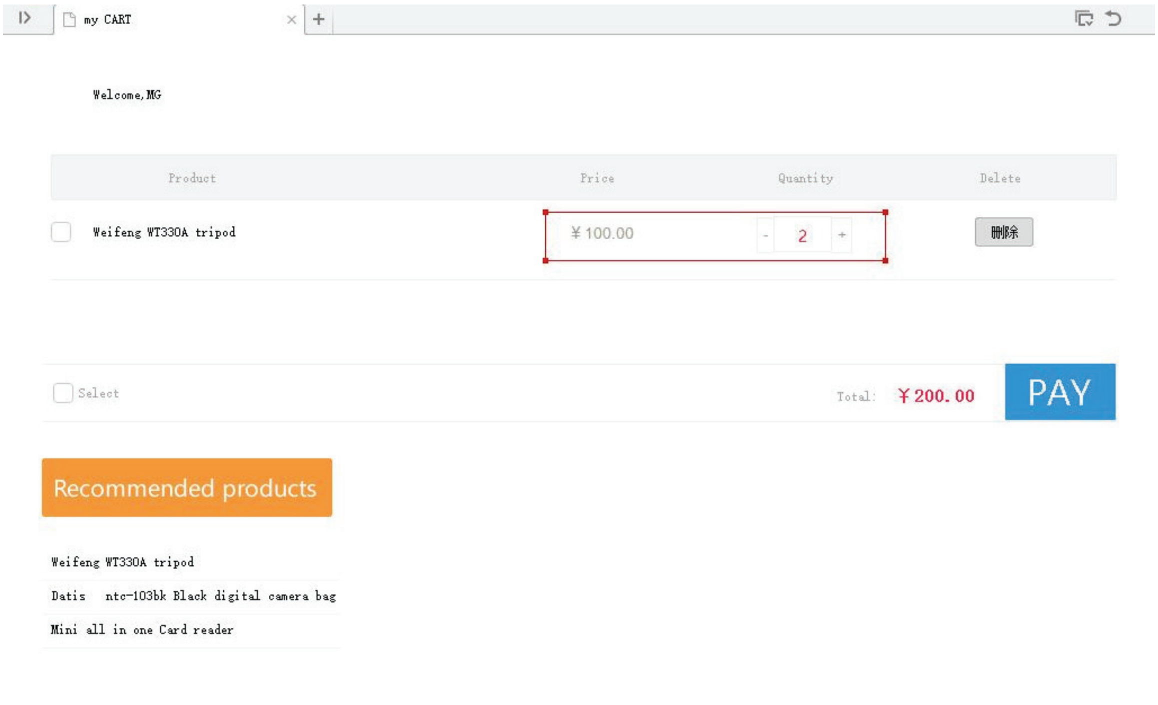

Figure 2 Interface diagram of recommended results

When a user logs in and buys "Soopen (QQ) QQ camera", the recommended product status of the user is as shown in Figure 2.

5. CONCLUSION

This paper discusses a personalized recommendation model based on user's interest; analyzes the recommendation model based on the market basket analysis method and prediction algorithm association rules; and presents the implementation and experimental results obtained under the laboratory conditions. This recommendation model relies on the data sources and commercial effects accumulated by the extensive use of e-commerce web sites. The evaluation of theory and experience is only one aspect of system evaluation. The system must eventually be optimized through application in practice.

REFERENCES

1. Sun L, Xu W. E-Commerce Personalized Recommendation System Based on Web Mining Technology Design and Implementation[C]// International Conference on Intelligent Transportation, Big Data and Smart City. IEEE, 2016:347–350.
2. Wang Y, Qiu F, Zhao J A, et al. Research on Personalized Recommendation of Learning Resources Based on Collaborative Filtering Recommendation Technology[J]. Journal of Distance Education, 2011.
3. Zhang Z Q, Qian S J. The Research of E-commerce Recommendation System Based on Collaborative Filtering Technology[C]// icise. IEEE Computer Society, 2012:3136–3138.
4. Wang J, Tang X C. Personalized recommendation algorithm research based on content in social network[J]. Application Research of Computers, 2011, 28(4):1248–1250.
5. Toivonen H. Apriori Algorithm[J]. Encyclopedia of Machine Learning, 2011:39–40.
6. Guo Y, Wang M, Li X. Application of an improved Apriori algorithm in a mobile e-commerce recommendation system[J]. Industrial Management & Data Systems, 2017, 117(2):287–303.
7. Agrawal R, Srikant R. Fast Algorithms for Mining Association Rules in Large Databases[C]// Proc. International Conference on Very Large Data Bases, Santiago, Chile, Sept. 1994:487–499.
8. Chen Y L, Tang K, Shen R J, et al. Market basket analysis in a multiple store environment[J]. Decision Support Systems, 2005, 40(2):339–354.
9. Meng B, Au O C, Wong C W, et al. Efficient intra-prediction algorithm in H.264[C]// International Conference on Image Processing, 2003. ICIP 2003. Proceedings. IEEE, 2003:III-837-40 vol.2.
10. Liu Y A, Yang B. Research of an improved Apriori algorithm in mining association rules[J]. Journal of Computer Applications, 2007, 27(2):418-420.
11. Klemettinen M. Finding interesting rules from large sets of discovered association rules[C]// Cikm 94 Third International Conference on Information & Knowledge Management. 1994:401–407.
12. Imielienskin T, Swami A, Agrawal R. Mining association rules between set of items in large databases[J]. Acm Sigmod Record, 1993, 22(2).
13. Ramirez A O. Three-Tier Architecture[J]. Linux Journal, 2000, 2000(75es):7.
14. Baralis E, Chiusano S, Garza P. On support thresholds in associative classification[C]// ACM Symposium on Applied Computing. ACM, 2004:553-558.
15. Dasseni E, Verykios V S, Elmagarmid A K, et al. Hiding Association Rules by Using Confidence and Support[C]// International Workshop on Information Hiding. Springer-Verlag, 2001:369–383.