(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 July 2001 (26.07.2001)

PCT

(10) International Publication Number
WO 01/53973 A2

(51) International Patent Classification[7]: G06F 17/00

(21) International Application Number: PCT/US01/01643

(22) International Filing Date: 19 January 2001 (19.01.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

| | | |
|---|---|---|
| 60/177,213 | 21 January 2000 (21.01.2000) | US |
| 09/520,837 | 8 March 2000 (08.03.2000) | US |

(71) Applicant: **NET PERCEPTIONS, INC.** [US/US]; 7901 Flying Cloud Drive, Eden Prairie, MN 55344 (US).

(72) Inventors: **FRANKOWSKI, Daniel**; 3216 Colfax Avenue South, Minneapolis, MN 55408-3554 (US). **BIEGANSKI, Paul**; 6461 Regency Lane, Minneapolis, MN 55344 (US). **DRISKILL, Robert**; 5890-66th Lane North, Greenfield, MN 55357 (US). **GURALNIK, Valerie**; 7235 Stewart Drive, Eden Prairie, MN 55346 (US). **MULIER, Filip**; 3774 Thomas Court, Vadnais Heights, MN 55127 (US).

(74) Agents: **GARRETT, Arthur, S.** et al.; Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P., 1300 I Street, N.W., Washington, DC 20005-3315 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published:**
— *without international search report and to be republished upon receipt of that report*

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

WO 01/53973 A2

**(54) Title:** RECOMMENDATION METHOD AND SYSTEM BASED ON RATING SPACE PARTITIONED DATA

**(57) Abstract:** Methods and systems consistent with the present invention provide a recommendation system that uses positive data and/or negative data separately to locate neighbors and provide recommendation to users. Such methods and systems use the data to locate potential neighbors based on users' ratings. Methods and systems consistent with the present invention calculate affinity values between the user and potential neighbors located to determine whether the potential neighbor's ratings are closely related to that of the user's ratings. If a user and a potential neighbor have an affinity greater than a predetermined threshold, that neighbor is considered close enough to the user to provide a recommendation for various items. Affinity values are calculated from a series of affinity equations available to the recommendation system.



META-1033
IPR2025-01248

# RECOMMENDATION METHOD AND SYSTEM BASED ON RATING SPACE PARTITIONED DATA

## BACKGROUND OF THE INVENTION

### A. Field of the Invention

This invention relates generally to data processing systems, and more particularly, to recommendation systems.

### B. Description of the Related Art

Recommendation systems are becoming widely used in e-commerce business activities. Recommendation systems allow e-commerce operators to take advantage of customer databases to provide valuable personalized service to customers. For example, systems that make personalized recommendations are used as a marketing tool to turn "window shoppers" into buyers, increase cross-sells and up-sells, and deepen customer loyalty.

Existing recommendation systems make recommendations to customer with unary data or data from a well-known Likert scale. Unary data is a set of user-item pairs that indicate an event of interest to the user has occurred. An example of unary data is purchase record data where a user-item pair indicates that the user has purchased a particular item. In general, unary data does not characterize the event, but rather it indicates that the event has occurred. Likert scale data indicates a user's preferences about an item. Typically, Likert scales give the user options, such as: like very much "5"; like a little "4"; don't mind either way "3"; dislike a little "2"; and strongly dislike "1." Likert data based calculations use a form of correlation calculations that assumes negative information and positive information within the same data. Thus, calculations using

Likert data must accept, and cannot differentiate or separate, positive and negative data. For example, these calculations handle low positive data (e.g., a user rating an item a "1") as a negative connotation such as "strongly dislike." In performing the correlation calculations, negative data is treated no different than positive data. Therefore, Likert based calculations associate customers regardless of the type of data in common, so long as the customers have some data in common. Therefore, Likert calculations may produce erroneous results.

Although existing recommendation systems provide recommendations based on unary data and Likert scale data, these systems are not capable of treating positive data and negative data differently. There exists a need to improve existing recommendation systems to provide recommendations while using negative data in a different manner from that of positive data.

## SUMMARY OF THE INVENTION

Methods and systems consistent with the present invention provide a recommendation system that uses positive data and negative data separately to locate neighbors and provide recommendation to users. Such methods and systems use the data to locate potential neighbors based on users' ratings. Methods and systems consistent with the present invention calculate affinity values between the user and potential neighbors located to determine whether the potential neighbor's ratings are closely related to that of the user's ratings. If a user and a potential neighbor have an affinity greater than a predetermined threshold, that neighbor is considered close enough to the user to provide

a recommendation for various items. Affinity values are calculated from a series of affinity equations available to the recommendation system.

Consistent with the present invention a computer-implemented method provides recommendations based on stored data corresponding to each one of a set of users with respect to a first item. The method analyzes partitioned preference data that reflects positive and negative preferences expressed by each one of a set of users with respect to the first item. The method also provides a recommendation based on the determined affinity.

Consistent with the present invention a computer-implemented method provides a recommendation using resource allocation data that indicates strength of a user's interest in a particular item. The method obtains click-stream data corresponding to the user, locates a plurality of neighbors with click-stream data similar to the user's click-stream data, and determines an affinity between the user and one of the plurality of neighbors based on the resource allocation data. Once determined, the method includes the one of the located neighbors meeting predetermined criteria on a neighbor list, and provides a recommendation to the user based on the neighbor list.

Consistent with the present invention a computer-implemented method provides a recommendation using resource allocation data that indicates a user's strength of an interest in a particular item. The method locates, in a database that contains resource allocation data for a plurality of users, other users with a similar strength of an interest in the particular item as the user, and determines an affinity between the user and one of the other users based on the similar strength of an interest. The method then provides a

recommendation to the user based on a list that contains a set of other users meeting predetermined criteria.

Consistent with the present invention a computer-implemented method provides a recommendation for an item based on likes and dislikes of a user. The method locates a plurality of neighbors with positive data similar to the user's positive data using a search strategy, and determines an affinity between the user and each one of the plurality of neighbors based on weighted agreements between the user and each one of the plurality of neighbors. The method also includes the one of the located neighbors meeting predetermined criteria on a neighbor list, and provides a recommendation to the user based on the neighbor list.

Consistent with the present invention a computer-implemented method provides a recommendation that indicates a user's likes and dislikes for a particular item. The method locates, in a database that contains positive and negative data for a plurality of users, other users with a similar positive likes for the particular item as the user, and determines an affinity between the user and one of the other users. The affinity is composed of weighted agreements. The method also provides a recommendation to the user based on the determined affinity.

Consistent with the present invention, a method generates a recommendation for an item. The method locates at least one potential neighbor for a user from a pool of candidates using a search strategy, and determines an affinity between the user and a potential neighbor. The affinity composed of a weighted agreement between the user and a potential neighbor. If the affinity is above a threshold, the method decides that the potential neighbor is a neighbor of the user.

Consistent with the present invention a computer-implemented recommendation method is disclosed. The method permits a user to submit a request for a recommendation, provides user ratings corresponding to each one of a set of users with respect to a first item to a rating database, and determines an affinity between a first user and another user by analyzing partitioned preference data that reflects positive and/or negative preferences expressed by each one of a set of users with respect to the first item. The method also provides a recommendation based on the determined affinity.

## BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings, which are incorporated in and constitute a part of this specification, illustrate an implementation of the invention and, together with the description, serve to explain the advantages and principles of the invention. In the drawings,

Figure 1 depicts a data processing system suitable for practicing methods and systems consistent with the present invention;

Figure 2 depicts a more detailed diagram of the client computer depicted in Fig. 1;

Figure 3 depicts a more detailed diagram of the recommendation server depicted in Fig. 1;

Figure 4 depicts a flow chart of the steps performed when collecting data in a manner consistent with the principles of the present invention;

Figure 5 depicts a flow chart of the steps performed when providing a recommendation consistent with the principles of the present invention;

Figure 6A depicts a rating table for use with methods and systems in a manner consistent with the present invention;

Figure 6B depicts an agreement table for use with methods and systems in a manner consistent with the principles of the present invention;

Figure 6C depicts an interest rating table for use with methods and systems consistent with the present invention;

Figure 6D depicts a normalized interest rating table for the interest rating table of Fig. 6C

Figure 6E depicts a second interest rating table for use with methods and systems consistent with the present invention;

Figure 6F depicts a second normalized interest rating table for the second interest rating table of Fig. 6E;

Figure 7A depicts an embodiment of an electronic commerce server for use with the present invention when using interest data; and

Figure 7B depicts another embodiment of an electronic commerce server for use with the invention when using RSP data.

## DETAILED DESCRIPTION

The following detailed description of the invention refers to the accompanying drawings. Although the description includes exemplary implementations, other implementations are possible, and changes may be made to the implementations described without departing from the spirit and scope of the invention. The following detailed description does not limit the invention. Instead, the scope of the invention is defined by the appended claims. Wherever possible, the same reference numbers will be

used throughout the drawings and the following description to refer to the same or like parts.

Overview

Methods and systems consistent with the present invention provide a recommendation server capable of using rating space partitioned (RSP) data to provide a recommendation to a user. RSP data is a type of data that represents positive and negative preferences expressed by a user for an item. An item may be any item available to recommend to a user. For example, an item may be a book, CD, or a movie. Positive data reflects the fact a desired event has occurred, or is likely to occur. For example, positive data may be a purchase event, adding an item to a shopping cart, or explicitly stating that a user likes an item. Positive data may also indicate positive preferences for an item, such as liking the item, or an interest in the item. Negative data reflects the fact that a desired event will not occur, or is likely not to occur. Negative data may also indicate negative preferences for the item, such as a dislike.

Since RSP data is partitioned (positive and negative), it is easy to determine whether a user likes or dislikes a particular item. RSP data may be based on implicit user interaction. That is, if a user spends less than three seconds on a particular web page, the RSP data for that particular user/page may be negative data.

Unlike Likert data, RSP data contains partitioned positive data and negative data that is separated and treated differently. This allows the recommendation system to create search strategies that accurately locate potential neighbors from a large pool of candidates. For example, a search strategy may use positive data to initially narrow the

number of potential neighbors, whereas the negative data may be used for minor corrections in the search strategy.

A special type of RSP data, called interest data, is a type of data that represents a measure of the level of interest someone has expressed in an item. Since interest data is always a positive measure, interest data recommendations may only use positive data (e.g., it is assumed that a user cannot show an interest in an item they dislike). Interest data may also be resource allocation data. Resource allocation data is a type of data where the user indicates, not only an item of interest, but also how much interest the user has in the items. For example, if a user has $1000 to spend on mutual funds, he may allocate his resource (money) to have $250 in mutual fund A, $750 in mutual fund B, and 0 in mutual fund C. The user has a higher interest in mutual fund B than in mutual fund A, and no interest in mutual fund C.

Interest data may also be based on user purchase data. That is, the interest data would include a list of items recently purchased by the user. A user that purchases more of item A than item B would have a higher interest in A than B. For example, if a user recently purchases item A and item B, and afterwards purchases ten more of item A, the user has a higher interest in item A than B. Resource allocation data may be considered to have multiple levels of interest (e.g., preferences), whereas purchase data is considered to have one level of interest (e.g., existence of a rating)

Recommendation systems that incorporate RSP data and interest data treat negative data differently than positive data to produce superior results from conventional recommendation systems. Also, RSP data enables quick drill down search strategies to limit the pool of candidates to provide accurate recommendations. Different search

strategies may be used based on the type of data to search, or recommendation. Interest recommendation systems provide recommendations when the data is only positive. Thus, a person's dislike in a particular item will not skew the results.

To provide recommendations, the recommendation system consistent with the present invention uses a data collector to track and record any user interaction. Recommendations may be used in a variety of situations. For example, a recommendation may be used as part of marketing campaigns that recommend items to users who are interested in similar items; as part of knowledge-management systems in large corporations that recommends reports and documents to employees based on the employees business or research interests; and as part of call centers that provide recommendations for merchandise to consumers placing orders.

System Components

Fig. 1 depicts a data processing system 100 suitable for practicing methods and systems consistent with the present invention. Data processing system 100 comprises a client computer 112 connected to recommendation server 120 via a network 130, such as the Internet. A user uses client computer 112 to provide various information to recommendation server 120.

Although only one client computer 112 is depicted, one skilled in the art will appreciate that data processing system 100 may contain many more client computers and additional client sites. One skilled in the art will also appreciate that recommendation server 120 may be located at various places on network 130. For example, the functions of recommendation server 120 may be included in a merchant server by client computer to, for example, purchase an item. Alternatively, the same functions may be incorporated

in client computer 112. A subset of the Internet is commonly referred to as the world wide web (www). Certain computers and/or servers connected to the web (referred to as web sites) offer information in the form of web pages. A web page may include digital content, such as text and/or images, audio streams, or instructions to obtain recommendation requests from a user using hypertext markup language (HTML), Java or other techniques.

Figure 2 depicts a more detailed diagram of client computer 112, which contains a memory 220, a secondary storage device 230, a central processing unit (CPU) 240, an input device 250, and a video display 260. Memory 220 includes browser 222 that allows users to interact with recommendation server 120 by transmitting and receiving files, such as web pages. An example of a browser suitable for use with methods and systems consistent with the present invention is the Netscape Navigator browser, from Netscape Corp.

As shown in Figure 3, recommendation server 120 includes a memory 310, a secondary storage device 320, a CPU 330, an input device 340, and a video display 350. Memory 310 includes recommendation engine 312 and data collector 314. Recommendation engine 312 determines if an item should be recommended to the user. It may use many different techniques to generate recommendations based on RSP data. One technique that may be used to generate recommendations is automated collaborative filtering as described in Resnick, Iacovo, Susha, Bergstrom, and Riedl, "GroupLens: An Open Architecture For Collaborative Filtering Of Netnews," Proceedings of the 1994 Computer Supported Collaborative Work Conference (1994). Other recommendation techniques are described in U.S. application serial no. 08/729,787, filed October 8, 1996,

U.S. application serial no. 08/733,806, filed October 18, 1996, attorney docket no. 7744-6000, filed September 23, 1999, U.S. application serial no. no.09/404,597, filed September 24, 1999, and U.S. application serial no. 09/438,846, filed November 12, 1999, all incorporated by reference. Recommendation systems may also be based on well-known Collaborate filtering (CF) systems, logical rules derived from data, or on statistical or machine learning technology. For example, a recommendation system may use well-known rule-induction learning, such as Cohen's Ripper, to learn a set of rules from a collection of data as described in Good, N., Schafer, J.B., Konstan, J., Borchers, A., Sarwar, B., Herlocker, J., and Riedl, J., "Combining Collaborative Filtering with Personal Agents for Better Recommendations," Proceedings of the 1999 Conference of the American Association of Artifical Intelligence (AAAI-99).

Recommendation systems may also be based on well-known data mining techniques that include a variety of supervised and unsupervised learning strategies and produce "surprising" results expressed as associations or rules embedded in a data set. Recommendation systems may also contain rating functions (models) programmed by a system administrator. The rating functions are either a formula or a table of ratings that determines business goals (e.g., the formula may specify a low rating for low-stock and out-of-stock items). These systems also require user data as input to produce personalized recommendations for users.

Data collector 314 monitors user interaction with various applications, such as an online store or any other e-commerce application (not shown) that may be operating on server 120 or elsewhere in network 100. For example, user interaction may include

explicit feedback, such as survey data, or responses to recommendations, (e.g., a user informs the application not to recommend a particular item by selecting a "not-for-me" button); implicit feedback, such as click stream data that reflects the user's navigation through web pages, time spent on web pages, links traverses; shopping and purchase data, such as number of items viewed, contents of shopping baskets, amount of purchases, or returns; or explicit preference ratings, such as numerical (like or dislike) or qualitative ratings of items. Data collector 314 may also monitor user interaction with various applications by obtaining web page logs from online stores or e-commerce applications. A web page log comprises a set of records of all activity on a particular web site. For example, a web page log may contain a user's time on a web page and information reflecting the web pages viewed during time passed. Regardless of the method used and/or statistics received by data collector 314, data collector 314 identifies a user and item, and stores the information as ratings in database 322.

Data collector 314 may include a web page, Application Program Interface (API), or other input interface to receive data from a user or an application. An API is a set of routines, protocols, or tools for communicating with software applications. APIs provide efficient access to data collector 314 without the need for additional software.

Secondary storage device 320 includes a rating database 322 that stores various user ratings, such as RSP data. Rating database 322 obtains data by receiving parsed user interaction from data collector 314. One skilled in the art will appreciate that database 322 may contain other types of data, such as unary data, and Likert.

Data Collection Process

Figure 4 depicts a flow chart of the steps performed by data collector 314 when collecting data. The first step is for data collector 314 to receive user interaction data (step 402). To receive user interaction data, data collector 314 parses various web page logs from an e-commerce application. If the received data requires processing (step 404), then data collector 314 processes the data (step 406). That is, data collector 314 extracts user and item information from the collected data and converts the data into positive and negative preference data. The RSP algorithms and affinity equations (described below) use either positive-only (e.g., interest data) or both positive and negative data to provide recommendations. Finally, data collector 314 stores the data in rating database 322 (step 408).

Recommendation Process with RSP Data

Figure 5 depicts a flow chart of the steps performed when generating a recommendation with RSP data. The first step is to receive a request for a recommendation from a user (step 502). The request may come in many forms. For example, a recommendation request may come from an e-commerce application that will display a list of items to a user before "check-out." The request may also come from a particular web page viewed by a user, or by monitoring "click-stream" data. Click-stream data is data obtained by monitoring users actions on particular web pages. Either way, the request is submitted to recommendation engine 312 using an API. For example, the e-commerce application may query recommendation engine with a "predict" API at the time the user displays finalizes his shopping cart.

A request for a recommendation may also come from an item, or a group of items (e.g., items that are within the same category). For example, recommendation engine 312 may recommend to the item a list of users that may be interested in that item.

Once recommendation engine 312 receives the request, recommendation engine 312 may, depending upon the type of equation, extract either positive ratings and/or negative ratings of the user or item from rating database 322 (step 504). If no data is available for the user, recommendation engine 312 may provide a default list. A default list would contain a preprogrammed list of items to recommend to the user. For example, if the user has never used recommendation engine 312 before, it may provide a top ten list of best selling items to the user.

Recommendation engine 312 uses the extracted data to locate potential neighbors in a candidate neighbor list using a specified search strategy (step 506). The term "neighbor" means a user identified in rating database 322 with similar interests as the first user. For example, if another user in rating database 322 has rated similar items as the first user, the other user may be considered a potential neighbor. A candidate neighbor list is defined as a pool of candidates from which to choose potential neighbors. For example, a candidate neighbor list may be geographically constrained, or consist of an entire database. A neighborhood is the list of neighbors found. Since rating database 322 may contain many potential neighbors, it is desirable to first reduce the set of candidates using the search strategy. Recommendation engine 312 may employ a search strategy using positive data to locate potential neighbors. Positive data is generally preferred in the search for potential neighbors since positive data contains more valuable information than negative data. That is, since rating database 322 contains a large

number of user ratings that are generally negative, and users only rate a small portion of the entire item space, using positive data leads to worthwhile rating data (and potential neighbors) more quickly.

Negative data may be later incorporated to provide recommendations using the affinity and agreement equations (described below) to make minor adjustments to the recommendation. Thus, negative data is used when providing recommendations using potential neighbors selected from a large pool of candidates.

Figure 6A depicts an exemplary portion of rating database 322 containing positive, negative and unrated items. User 2 and User 3 may be potential neighbors of User 1 when using a positive data search strategy since User 1 and both User 2 User 3 rated positively similar items (items 1 and 2). However, both Users are considered *potential neighbors* of User 1 since an affinity between the users still needs to be determined, as further described below. For example, an ideal neighbor for a user would be a neighbor that has rated all items that the user has also rated. Moreover, User 4 will never be a potential neighbor of User 1, since User 4 has not rated positively rated items similar to that of User 1.

If no potential neighbors are found (step 508), recommendation engine 312 attempts to locate any neighbor to provide a recommendation (step 510). If neighbors are not located, then recommendation engine 312 uses a default list instead of providing a recommendation, as described above (step 520). If, however, recommendation engine 312 locates neighbors (step 510), recommendation engine 312 uses the located neighbors to provide a recommendation (step 522).

If, however, at least one potential neighbor is found (e.g., using a positive data search strategy) (step 508), recommendation engine 312 computes an affinity between the user and the potential neighbor using an affinity equation (step 512). Affinity equations consist of any combination of weighted agreement components, such as positive agreement, negative agreement, or disagreement, described below.

Positive Agreement

A positive agreement measures a common level of positive preference between the user and the potential neighbor. The agreement computes a function using positive co-ratings of the user and the potential neighbor. The positive agreement may be computed using the following equation where "R" is the first user, "r" is the potential neighbor, and "positive_coratings" is the number of items both have rated positively:

$$Positive_agreement = \frac{\sum_{positive_coratings} 1}{\sum_{positive_ratings_R \cup positive_ratings_r} 1}$$

One skilled in the art will appreciate that other agreement equations, such as Mutual Normalized Interests, and Fuzzy Evidence Set Similarity, may be used to compute the positive agreement. These equations are further described below.

Mutual Normalized Interests

The mutual normalized interest equation is a positive agreement equation that uses normalized interest information, such as normalized ratings, to return a common interest level between the user and the potential neighbor. To do so, the equation computes the sum of the minimum normalized coratings. A corating is a pair of ratings

for the users. For example, Figure 6C depicts an interest rating table 620 containing common ratings between a user and a potential neighbor. Figure 6D depicts a normalized interest table 630 containing normalized data from interest rating table 620.

The mutual normalized interest is computed using the following equation where "R" is the first user, "r" is the potential neighbor, and "coratings_i" is the number of items both users have rated:

$$affinity = \sum_{coratings_i} \min(R'_i, r'_i)$$

Using the normalized data in normalized interest table 610, methods and systems consistent with the present invention may provide an affinity value between the user and potential neighbor by computing the sum of the minimum of the coratings: .1 + .1 + .4 = .6. The value ".6" is an affinity value between the user and the potential neighbor.

Fuzzy Evidence Set Similarity

The fuzzy evidence set similarity equation is another positive agreement equation that uses normalized interest information, such as normalized ratings, to return the amount of interest overlap between the user and the potential neighbor. More information regarding fuzzy evidence may be found in Zimmerman, H.J., "Fuzzy Set Theory - And Its Applications," Second Revised Edition, 1991, hereby incorporated by reference. For example, Fig. 6E depicts an interest rating table 640 containing some common ratings between a user and a potential neighbor. Fig. 6F depicts a normalized interest table 650 containing normalized data from interest rating table 640.

The fuzzy evidence set similarity is computed using the following equation, where "noncorating_j" is the set of items "r" has not rated that "R" has rated, "noncorating_k"

is the set of items "R" has not rated that "r" has, and "coratings_i" is the number of items both users have rated:

$$affinity = \frac{\sum_{coratings_i} \min(R'_i, r'_i)}{\sum_{noncorating_j} R'_j + \sum_{noncorating_k} r'_k + \sum_{coratings_i} \max(R'_i, r'_i)}$$

The sum of the minimum of coratings for Figure 6D is: (.05) + (.25)

The sum of the coratings not available to user "R" is: 0

The sum of the coratings not available to the potential neighbor "r" is (.2) + (.5)

The sum of the maximum of coratings is: (.5) + (.5)

Thus, the affinity measure between the user "R" and the potential neighbor "r" is: ".17".

Although two interest equations are explained above, one skilled in the art will appreciate that other interest equations may be used, such as a cosine similarity equation. The cosine equation is as follows:

$$affinity = \frac{\sum_{coratings_i} R'_i * r'_i}{\sqrt{\sum_j R'^2_j * \sum_j r'^2_j}}$$

More information on cosine similarity equations may be found in Salton, G., "The SMART Retrieval System: Experiments in Automatic Document Processing," Prentice Hall, Englewood Cliffs, NJ, 1971, hereby incorporated by reference.

Negative Agreement

A negative agreement measures a common level of negative preference between the user and the potential neighbor. The agreement computes a function using negative

co-ratings of the user and the potential neighbor. The negative agreement may be computed using the following equation where "R" is the first user, "r" is the potential neighbor, and "negative_coratings" is the number of items both have rated negatively:

$$Negative_agreement = \frac{\sum_{negative_coratings} 1}{\sum_{negative_ratings_R \cup negative_ratings_r} 1}$$

One skilled in the art will appreciate that any positive agreement equation may be used to computer the negative agreement on negative data.

Disagreement

A disagreement measures a level of disagreement in preferences between the user and the potential neighbor. The disagreement computes a function using opposite co-ratings of the user and the potential neighbor. The opposite agreement be computed using the following equation where "R" is the first user, "r" is the potential neighbor, and "opposite_coratings" is the number of items both have rated opposite:

$$Disagreement = \frac{\sum_{opposite_coratings} 1}{\sum_{ratings_R \cup ratings_ratings_r} 1}$$

Figure 6B depicts a completed agreement table 610 with values for various types agreements correlating to table 600 in Figure 6A. For example, the Positive_Agreement between User 1 and User 2 is "2/4" since they share two of four positive ratings; both have rated items 1 and 2 positively, User 1 has also rated item 5 positively, and User 2

has rated item 7 positively.

RSP Affinity Equations

Once the agreements are determined, an affinity may be determined for the user and a potential neighbor using various RSP affinity equations. Listed below are four exemplary RSP affinity equations, where "Wp" is a positive weight, "Wn" is a negative weight, "Wd" is a disagreement weight:

(1) positive affinity = Positive_Agreement

(2) positive & negative affinity = Wp*Positive_Agreement + Wn*Negative_Agreement

(3) positive affinity without disagreement = Wp*Positive Agreement - Wd*Disagreement

(4) general affinity = Wp*Positive_Agreement + Wn*Negative_Agreement - Wd*Disagreement

Equations 1-3 are special cases of equation 4, with Wp, Wn, and Wd set to different values.

Using equation 2 and Wp=Wn=1, methods and systems consistent with the present invention may provide an affinity value between User 1 and User 2 as follows: 1* "2/4" + 1*"1/3" = "10/12," and User 1 and User 3 as follows: 1*"2/4" + 1*"2/3" = "14/12." Thus, User 3 has a higher affinity with User 1 that User 2 has with User 1.

Any of the above listed affinity equations may be used in any combination to compute an affinity value based on the level of agreements measured. Based on the landscape of the candidate neighbors, one affinity equation may be more useful than

another. In one example, when the ratings database consists of mostly positive data, and all users have similar ratings (e.g., a rock music store), positive data (e.g., interest data) may be useful as a search strategy, however disagreement data may be useful to refine and/or select neighbors. Thus, equation 3 would be used. In another example, when the ratings database consists of unreliable or sparse positive data (e.g., positive agreement on few items), negative data may be useful to increase the trustworthiness of the overall affinity values.

After each affinity value is computed for a user and a potential neighbor using an equation as described above, recommendation engine 312 determines if the affinity value is above a predetermined threshold value (step 514). One skilled in the art will appreciate that the threshold value may be a maximum value, minimum value, or a range of values. If the affinity value is above the threshold value, the potential neighbor is added to a neighbor list (step 516). Each neighbor on the neighbor list provides rating information to recommendation engine 312 that is used to compute a recommendation for the user. Otherwise, if the affinity value is below the threshold value, the potential neighbor is dropped and the next potential neighbor is located in rating database 322 (step 506).

Recommendation engine 312 locates neighbors until enough neighbors have been located (step 518). For example, to provide a quick recommendation, recommendation engine 312 may require ten neighbors. However, to provide a more accurate recommendation, recommendation engine 312 may require fifty neighbors. Once the requisite number of neighbors has been located, recommendation engine 312 may provide a recommendation to the user using well-known recommendation techniques (step 524).

Positive Interest Data Example

As an example of an application suitable for methods and systems consistent with the present invention are suitable for use with an augmented electronic mutual fund server on the Internet. Fig. 7A illustrates a recommendation system integrated into a web-based electronic mutual fund site (e-commerce site). The user at computer 702 connects using a network 704 to a web server 706. A commerce server 708, connected to web server 706, processes all financial transactions for the user and contains a database of various mutual funds for sale. Web server 706 presents this set of products for sale to the user.

A recommendation server 710 coupled to the web server 706 and commerce server 708 receives purchase information from commerce server 708. The recommendation server 710 uses web server 706 and commerce server 708 to provide the user with specifically targeted content, such as recommendations to purchase specific items, recommendations to view specific items, or targeted advertisement. Recommendation server 710 does so by maintaining records of previous purchases and quantity of purchases by the user and other users.

As a specific example of the recommendation system implemented as described above, a user may purchase $1000 of mutual fund A, $500 of mutual fund B, and $2000 of mutual fund C. Each time user 602 buys or sells a mutual fund, the commerce server records the purchase and provides the recommendation server with the data. The recommendation server may then compare user 702 portfolio to other user's portfolios maintained in the recommendation server using an interest affinity equation. The users that have high affinities with user 602 are considered neighbors and are included on a

neighbor list that is used to provide recommendations to user 602. For example, if another user has $1000 mutual fund A, $1000 mutual fund B, and $1000 in mutual fund D, recommendation server 610 may recommend that user 602 consider mutual fund D as a potential investment.

RSP Data Example

As another example of an application suitable for methods and systems consistent with the present invention are suitable for use with CD stores on the Internet. Fig. 7B illustrates a recommendation system integrated into various CD stores (e-commerce sites) using RSP data to provide recommendations. The user at computer 712 connects using a network 714 to an e-commerce server 716. E-commerce servers 716 process all purchase transactions for user 712 and contain databases of various CDs for sale.

A recommendation server 718 coupled to network 714 receives purchase and return information from e-commerce servers 716. The recommendation server 718 uses this information to provide user 712 with specifically targeted content, such as recommendations to purchase specific CDs, recommendations to view specific CDs, or targeted advertisement at each e-commerce server 716. Recommendation server 718 does so by maintaining records of previous purchases and quantity of purchases by user 712 and other users.

As a specific example of the recommendation system implemented as described above, a user may purchase CDs A, B, and C, and return CD A, from different e-commerce servers 716. Each time user 712 purchases a CD, an e-commerce server 716 records the user interaction and provides the purchase information as a positive rating for

that CD to recommendation server 718. Each time user 712 returns a CD, an e-commerce server 716 records the user interaction and provides the return information as a negative rating for that CD to recommendation server 718. Recommendation server 718 may use the positive data to locate potential neighbors and compare user 712 purchase/return history to located potential neighbors' purchase/return history using predetermined RSP affinity and agreement equations. Other users that have high affinities with user 712 are considered neighbors to user 712 and will be included on a neighbor list that is used to provide recommendation to user 712. For example, if another user has purchased CD A, B, and D, recommendation server 718 may recommend that user 712 consider to purchase CD D.

Conclusion

Methods and systems consistent with the present invention provide a recommendation server capable of using RSP data to provide a recommendation to a user. The recommendation server contains software to provide RSP data recommendations to the user. Alternatively, the software may provide recommendations of users to an item, or groups of items. To provide the recommendations, the recommendation server applies an affinity equation to the set of RSP data. The foregoing description of an implementation of the invention has been presented for purposes of illustration and description. It is not exhaustive and does not limit the invention to the precise form disclosed. Modifications and variations are potential in light of the above teachings or may be acquired from practicing of the invention. For example, positive data definitions and negative data definitions may be reversed. Also, although methods and systems consistent with the present invention describe providing a recommendation for an item

to a user, conversely, recommendations may also be provided for a user to an item. That is, a recommendation may be generated including a list of suitable users (e.g., by using item affinities). Moreover, the described implementation includes software but the present invention may be implemented as a combination of hardware and software or in hardware alone.

WHAT IS CLAIMED IS:

1. A computer-implemented method for providing recommendations based on stored data corresponding to each one of a set of users with respect to a first item, comprising:

determining an affinity between a first user and another user by analyzing partitioned preference data associated with the stored data that reflects positive and negative preferences expressed by each one of a set of users with respect to the first item; and

providing a recommendation based on the determined affinity.

2. The method of claim 1, wherein the data is resource allocation data.

3. The method of claim 1, wherein determining an affinity between a first user and another user, further includes:

locating other users with individual interest considered similar to the first user based on resource allocation data; and

including one of the located users meeting predetermined criteria on a list.

4. The method of claim 1, wherein determining an affinity between a first user and another user, further includes:

using a search strategy to locate other users; and

including located users meeting predetermined criteria on a list.

5. The method of claim 4, wherein positive co-rating data is used in the search strategy to locate the users.

6. The method of claim 4, wherein determining an affinity further includes using a weighted sum one of at least positive agreement, negative agreement, or opposite agreement between the first user and the another user.

7. The method of claim 6, wherein determining an affinity further using a weighted positive agreement further includes using one of at least a mutual normalized interest equation, a fuzzy evidence equation, and a cosine similarity equation.

8. The method of claim 1, wherein determining an affinity further includes appending the another user to a list when an affinity value exceeds a predetermined value.

9. The method of claim 8, further comprising the step of truncating the list when a predetermined number of users on the list has been met.

10. The method of claim 8, further comprising the step of selecting a predetermined number of users from the list that meet predetermined criteria.

11. The method of claim 1, wherein collecting data further includes extracting the data from a database.

12. The method of claim 1, wherein collecting data further includes measuring implicit ratings from the set of users.

13. The method of claim 1, wherein collecting data further includes determining a time period associated with a user's viewing a web page.

14. The method of claim 1, wherein collecting data further includes providing purchase data to indicate how much money each one of a set of users spent.

15. The method of claim 1, wherein providing a recommendation further includes providing the recommendation based solely on user purchase data.

16. The method of claim 1, wherein collecting data further includes providing data to indicate item purchases and/or item returns.

17. The method of claim 1, wherein collecting data further includes determining whether an item has been rejected by a user.

18. The method of claim 1, wherein the data is RSP data.

19. A computer-implemented method for providing a recommendation using resource allocation data that indicates strength of a user's interest in a particular item, comprising the steps of:

obtaining click-stream data corresponding to the user;

locating a plurality of neighbors with click-stream data similar to the user's click-stream data;

determining an affinity between the user and one of the plurality of neighbors based on the resource allocation data;

including the one of the located neighbors meeting predetermined criteria on a neighbor list; and

providing a recommendation to the user based on the neighbor list.

20. A computer-implemented method for providing a recommendation using resource allocation data that indicates a user's strength of an interest in a particular item, comprising the steps of:

locating, in a database that contains resource allocation data for a plurality of users, other users with a similar strength of an interest in the particular item as the user;

determining an affinity between the user and one of the other users based on the similar strength of an interest; and

providing a recommendation to the user based on a list that contains a set of other users meeting predetermined criteria.

21. The method of claim 20, wherein determining an affinity further includes:

using at least a mutual normalized interest equation, a fuzzy evidence equation, and a cosine similarity equation to compute the affinity.

22. The method of claim 20, wherein determining an affinity further includes:

including one of the other users on the list when an affinity value exceeds a predetermined number.

23. The method of claim 22, further comprising the step of truncating the list when a predetermined number of the other users on the list has been met.

24. The method of claim 20, wherein the resource allocation data further includes implicit ratings from the user.

25. A computer-implemented method for providing a recommendation for an item based on likes and dislikes of a user, comprising the steps of:

locating a plurality of neighbors with positive data similar to the user's positive data using a search strategy;

determining an affinity between the user and each one of the plurality of neighbors based on weighted agreements between the user and each one of the plurality of neighbors;

including the one of the located neighbors meeting predetermined criteria on a neighbor list; and

providing a recommendation to the user based on the neighbor list.

26. A computer-implemented method for providing a recommendation that indicates

a user's likes and dislikes for a particular item, comprising the steps of:

locating, in a database that contains positive and negative data for a plurality of users, other users with a similar positive likes for the particular item as the user;

determining an affinity between the user and one of the other users, wherein the affinity is composed of weighted agreements; and

providing a recommendation to the user based on the determined affinity.

27. The method of claim 26, wherein determining an affinity further includes:

using a weighted sum one of at least positive agreement, negative agreement, or disagreement to determine the affinity.

28. The method of claim 26, wherein determining an affinity further includes including one of the other users on a list when the affinity exceeds a predetermined number.

29. The method of claim 28, further comprising the step of truncating the list when a predetermined number of the other users on the list has been met.

30. A method for generating a recommendation for an item, comprising the steps, executed in a data processing system, of:

locating at least one potential neighbor for a user from a pool of candidates using a search strategy;

determining an affinity between the user and a potential neighbor, wherein the affinity comprises a weighted agreement between the user and a potential neighbor;

determining if the affinity is above a threshold and if so, deciding that the potential neighbor is a neighbor of the user.

31. The method of claim 30, wherein the search strategy uses positive data to locate

potential neighbors.

32. A system for providing a user a recommendation for an item, comprising:

a commerce server that processes purchase and return transactions for a plurality of users; and

a recommendation server that provides a recommendation for a first user based on a determination of an affinity composed of weighted agreements between the first user and another user,

wherein the commerce server transmits information associated with the processed transactions to the recommendation server.

33. The system of claim 32, wherein the recommendation server determines the affinity by using a search strategy to locate other users, and wherein positive co-rating data is used in the search strategy to locate the other users.

34. The system of claim 33, wherein the affinity is a weighted sum one of at least positive agreement, negative agreement, or disagreement between the first user and a located user.

35. The system of claim 32, wherein the recommendation server determines the affinity by locating other users with interest considered similar to the first user based on resource allocation data, and includes one of the located users meeting predetermined criteria on a list.

36. The system of claim 35, wherein the affinity is determined by using one of at least a mutual normalized interest equation, a fuzzy evidence equation, and a cosine similarity equation.

37. The system of claim 32, wherein the recommendation server further appends the another user to the list when the affinity exceeds a predetermined number.

38. A system for providing recommendations, comprising:

data collecting means for collecting data corresponding to each one of a set of users with respect to a first item;

determining means for determining an affinity between a first user and another user by analyzing partitioned preference data associated with the collected data that reflects positive and negative preferences expressed by each one of a set of users with respect to the first item; and

providing means for providing a recommendation based on the determined affinity.

39. The system of claim 38, wherein the data is resource allocation data.

40. The system of claim 38, wherein the determining means further includes using one of at least a mutual normalized interest equation, a fuzzy evidence equation, and a cosine similarity equation.

41. The system of claim 38, wherein the determining means further includes locating means for locating other users with individual interest considered similar to the first user based on resource allocation data and including one of the located users meeting predetermined criteria on a list.

42. The system of claim 38, wherein the determining means uses a search strategy to locate other users, and wherein the determining means further includes means for including located users meeting predetermined criteria on a list.

43. The system of claim 42, wherein positive co-rating data is used in the search

strategy to locate the users.

44. The system of claim 42, wherein the determining means determines an affinity using a weighted sum one of at least positive agreement, negative agreement, or disagreement between the first user and the another user.

45. The system of claim 38, wherein the determining means further includes appending means for appending the another user to a list when an affinity value exceeds a predetermined value.

46. The system of claim 45, further comprising truncating means for truncating the list when a predetermined number of the located users on the list has been met.

47. The system of claim 38, wherein the data collection means further includes extracting means for extracting the data from a database.

48. The system of claim 38, wherein the data collecting means further includes measuring means for measuring implicit ratings from the set of users.

49. The system of claim 38, wherein the data collecting means further includes determining means for determining a time period associated with a user's viewing a web page.

50. A computer-implemented recommendation method, comprising:

permitting a user to submit a request;

providing user ratings corresponding to each one of a set of users with respect to a first item to a rating database;

determining an affinity between a first user and another user by analyzing partitioned preference data that reflects positive and/or negative preferences expressed by each one of a set of users with respect to the first item; and

providing a recommendation to the user based on the determined affinity.

51. The system of claim 50, wherein the user ratings are interest ratings.

52. A computer readable medium for controlling a data processing system to perform a method for providing recommendations based on stored data corresponding to each one of a set of users with respect to a first item executed in a data processing system, the computer readable medium comprising:

5 a determination module for determining an affinity between a first user and another user by analyzing partitioned preference data associated with the stored data that reflects positive and negative preferences expressed by each one of a set of users with respect to the first item; and

a recommendation module for providing a recommendation based on the determined affinity.

53. The computer readable medium of claim 52, wherein the data is resource allocation data.

54. The computer readable medium of claim 52, wherein the determination module further includes:

a locating module for locating other users with individual interest considered similar to the first user based on resource allocation data and including one of the located users meeting predetermined criteria on a list.

55. The computer readable medium of claim 52, wherein the determination module further uses a search strategy to locate other users, and includes located users meeting predetermined criteria on a list.

56. The computer readable medium of claim 55, wherein positive co-rating data is

used in the search strategy to locate the users.

57. The computer readable medium of claim 55, wherein the determination module determines an affinity by using a weighted sum one of at least positive agreement, negative agreement, or opposite agreement between the first user and the another user.

58. The computer readable medium of claim 57, wherein the determination module determines an affinity by using the weighted positive agreement, wherein the weighted positive agreement includes one of at least a mutual normalized interest equation, a fuzzy evidence equation, and a cosine similarity equation.

59. The computer readable medium of claim 52, wherein the determination module further includes an appending module for appending the another user to a list when an affinity value exceeds a predetermined value.

60. The computer readable medium of claim 59, further including a truncating module for truncating the list when a predetermined number of users on the list has been met.

61. The computer readable medium of claim 60, further including a selecting module for selecting a predetermined number of users from the list that meet predetermined criteria.

62. The computer readable medium of claim 52, further including an extracting module for extracting the data from a database.

63. The computer readable medium of claim 52, further including a measuring module for measuring implicit ratings from the set of users.

64. The computer readable medium of claim 52, further including a determining

module for determining a time period associated with a user's viewing a web page.

65. The computer readable medium of claim 52, further including a providing module for providing purchase data to indicate how much money each one of a set of users spent.

66. The computer readable medium of claim 52, wherein the providing module provides the recommendation based solely on user purchase data.

67. The computer readable medium of claim 52, further including a providing module for providing data to indicate item purchases and/or item returns.

68. The computer readable medium of claim 52, further including a determining module for determining whether an item has been rejected by a user.

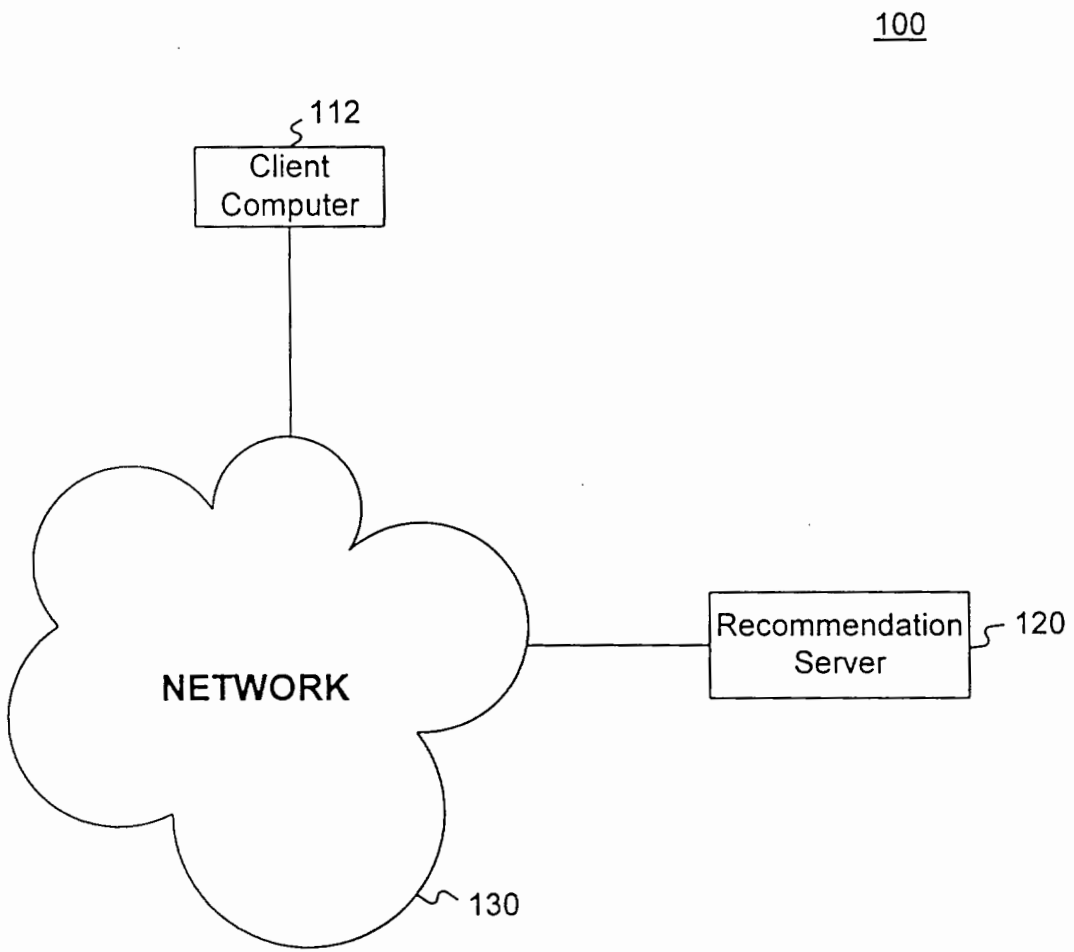
100
112
Client Computer
NETWORK
Recommendation Server
120
130

FIG. 1

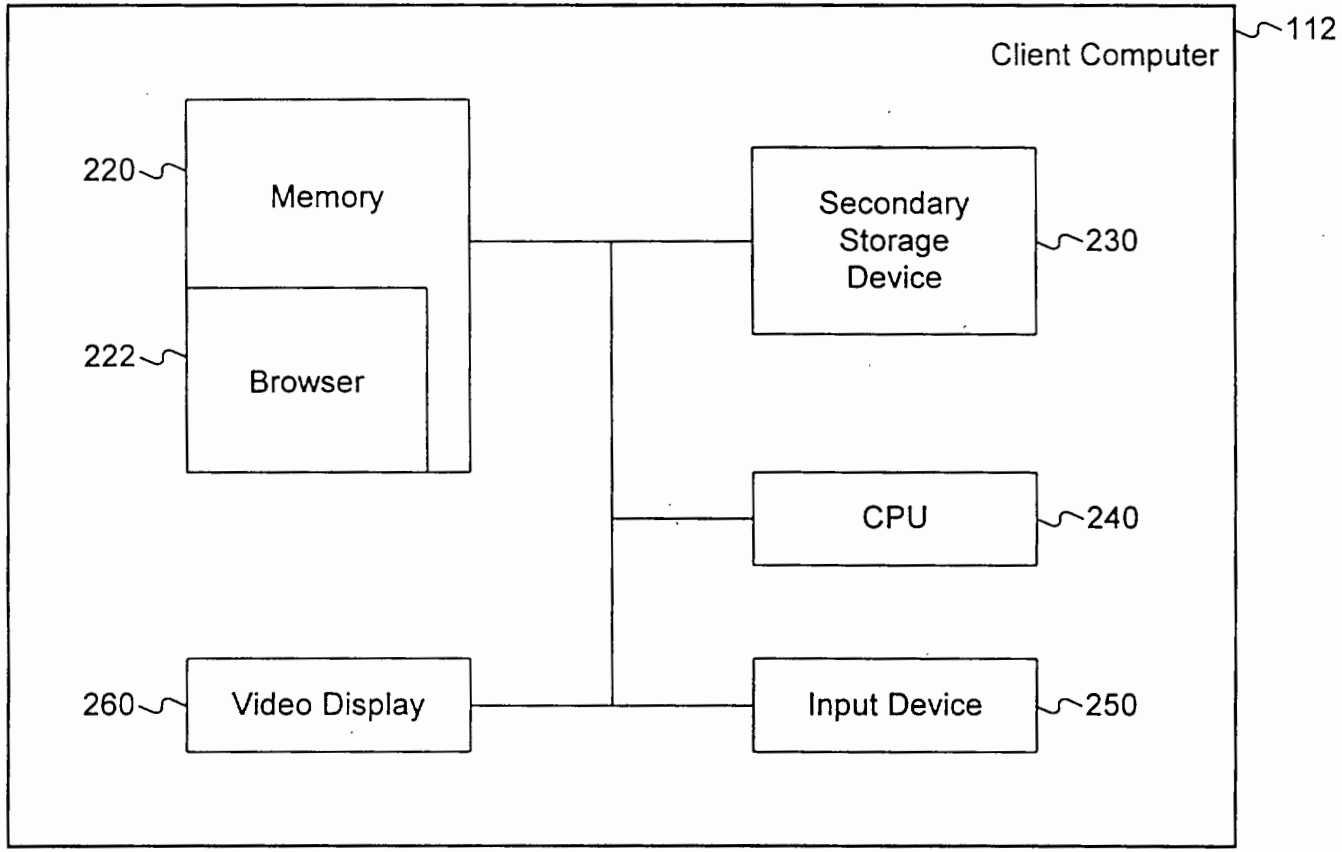
Client Computer
112
220
Memory
222
Browser
Secondary Storage Device
230
CPU
240
260
Video Display
Input Device
250

FIG. 2

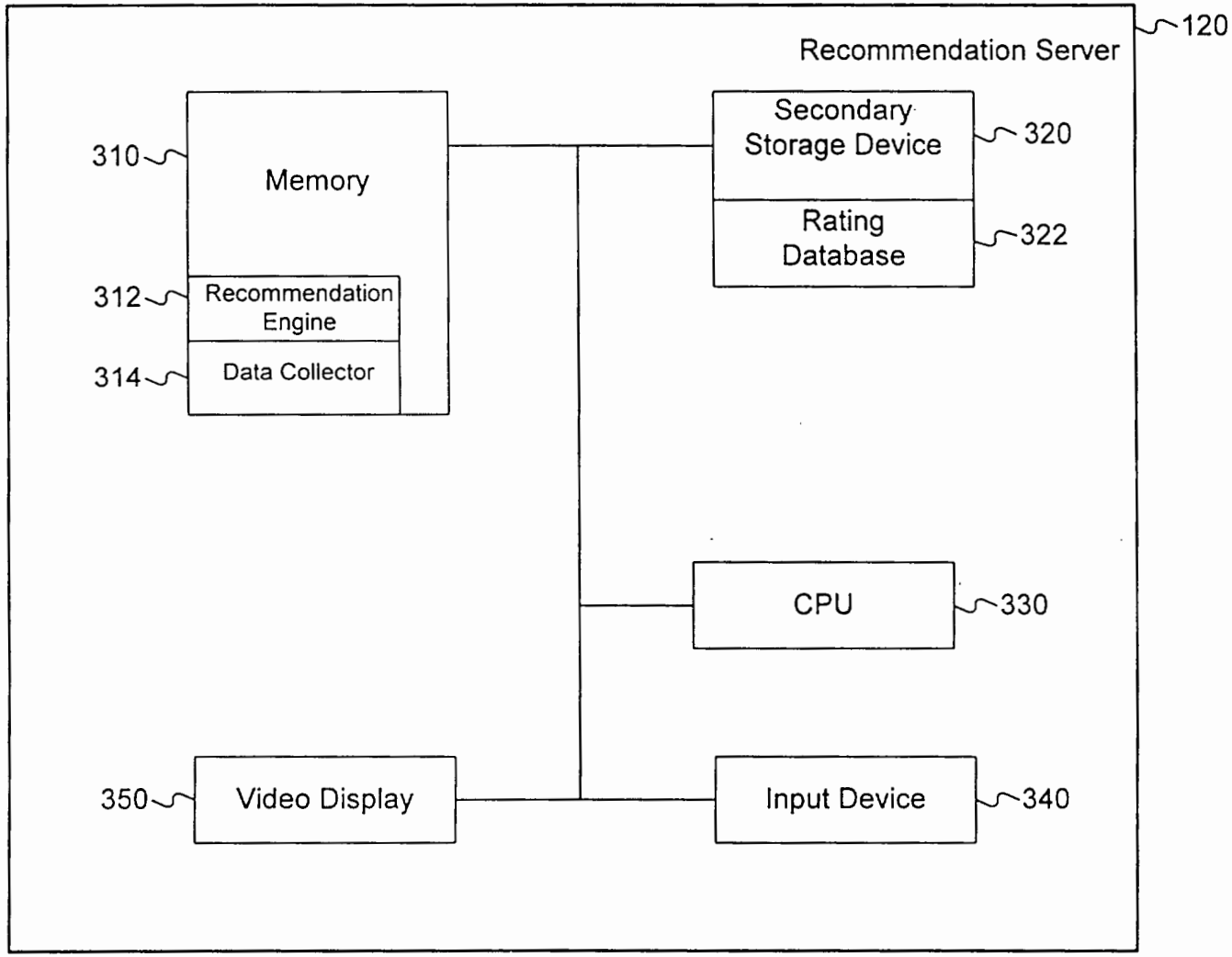
120
Recommendation Server
310
Memory
312
Recommendation Engine
314
Data Collector
Secondary Storage Device
320
Rating Database
322
CPU
330
350
Video Display
Input Device
340

FIG. 3

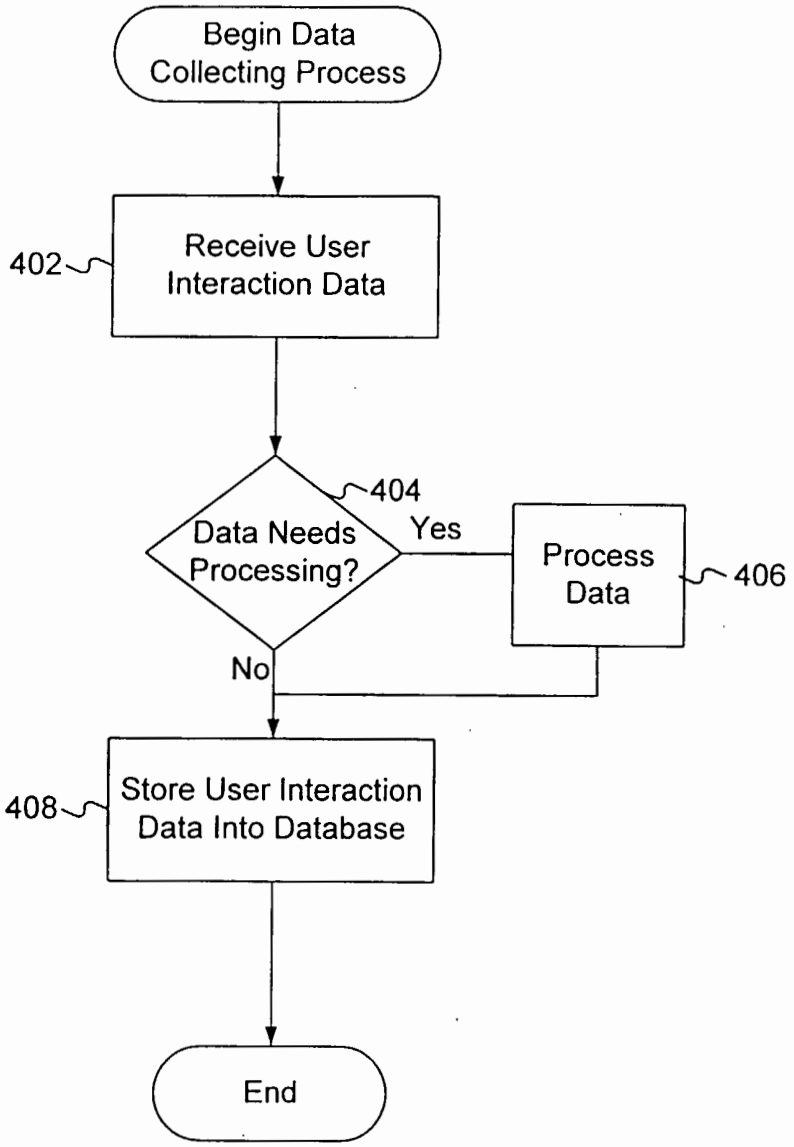
Begin Data Collecting Process
402
Receive User Interaction Data
404
Data Needs Processing?
Yes
Process Data
406
No
408
Store User Interaction Data Into Database
End

FIG. 4

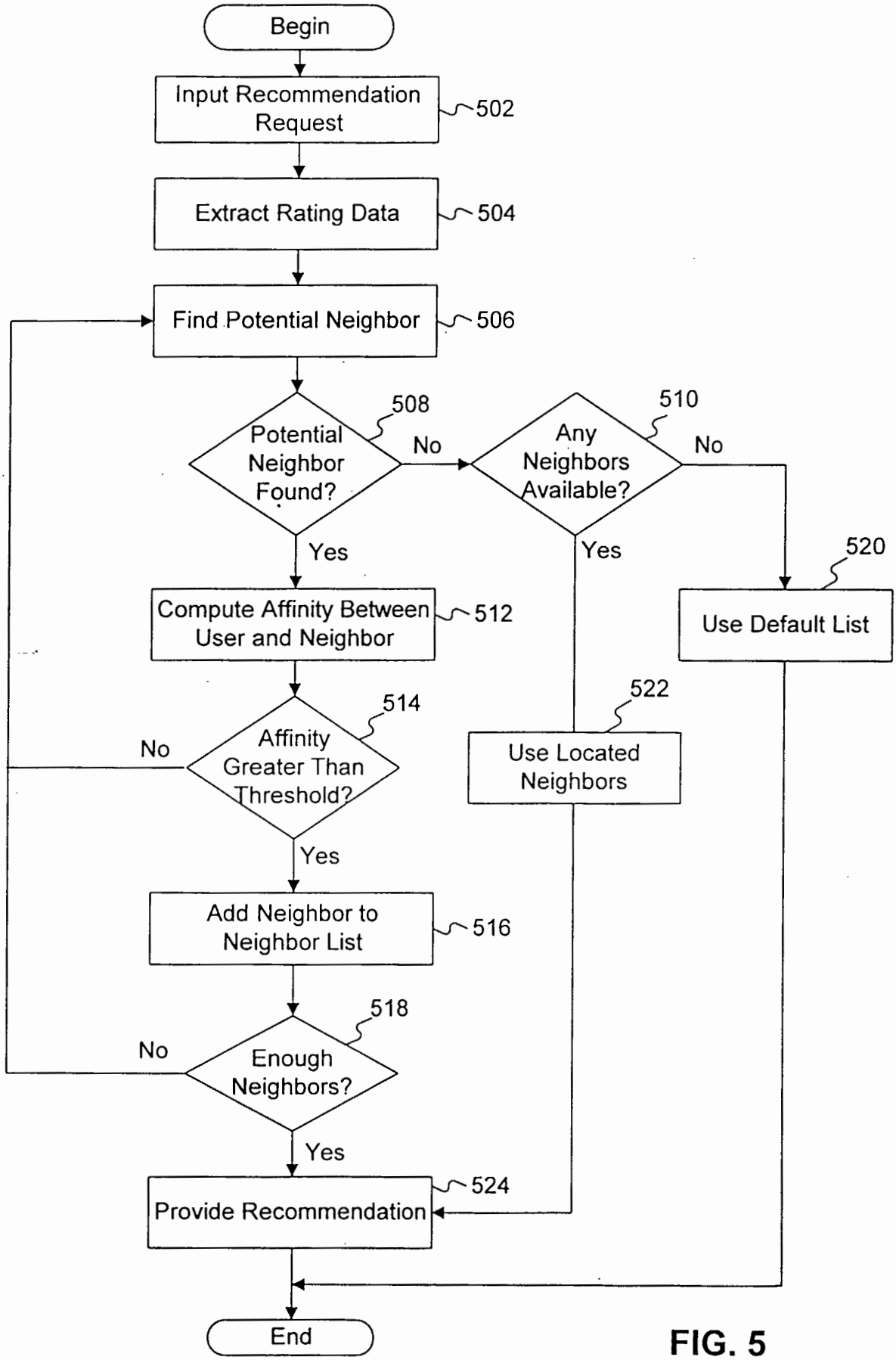
Begin
Input Recommendation Request
502
Extract Rating Data
504
Find Potential Neighbor
506
Potential Neighbor Found?
508
No
Any Neighbors Available?
510
No
Yes
Yes
520
Compute Affinity Between User and Neighbor
512
Use Default List
522
Affinity Greater Than Threshold?
514
No
Use Located Neighbors
Yes
Add Neighbor to Neighbor List
516
518
No
Enough Neighbors?
Yes
524
Provide Recommendation
End

FIG. 5

600

BINARY RSP RATINGS

| | Item 1 | Item 2 | Item 3 | Item 4 | Item 5 | Item 6 | Item 7 |
|---|---|---|---|---|---|---|---|
| User 1 | 1 | 1 | -1 | -1 | 1 | — | — |
| User 2 | 1 | 1 | — | -1 | — | -1 | 1 |
| User 3 | 1 | 1 | -1 | -1 | -1 | 1 | — |
| User 4 | -1 | -1 | -1 | -1 | — | 1 | 1 |

FIG. 6A

610

| Subject | Potential Neighbor | Positive Agreement | Negative Agreement | Disagreement | Positive+ Negative | Positive- Disagreement |
|---|---|---|---|---|---|---|
| User 1 | User 2 | 2/4 | 1/3 | 0 | 10/12 | 2/4 |
| User 1 | User 3 | 2/4 | 2/3 | 1/6 | 14/12 | 4/12 |
| User 1 | User 4 | 0 | 2/4 | 2/6 | 2/4 | -2/6 |

FIG. 6B

620

INTEREST RATINGS

| | Item 1 | Item 2 | Item 3 |
|---|---|---|---|
| User 1 | 4 | 1 | 5 |
| User 2 | 1 | 5 | 4 |

FIG. 6C

640

INTEREST RATINGS

| | Item 1 | Item 2 | Item 3 | Item 4 |
|---|---|---|---|---|
| User | 4 | 1 | 5 | 10 |
| User 2 | — | 5 | 5 | — |

FIG. 6E

630

Normalized Interest Ratings

| | Item 1 | Item 2 | Item 3 |
|---|---|---|---|
| User | .4 | .1 | .5 |
| User 2 | .1 | .5 | .4 |

FIG. 6D

650

Normalized Interest Ratings

| | Item 1 | Item 2 | Item 3 | Item 4 |
|---|---|---|---|---|
| User | .2 | .05 | .25 | .05 |
| User 2 | — | .5 | .5 | — |

FIG. 6F

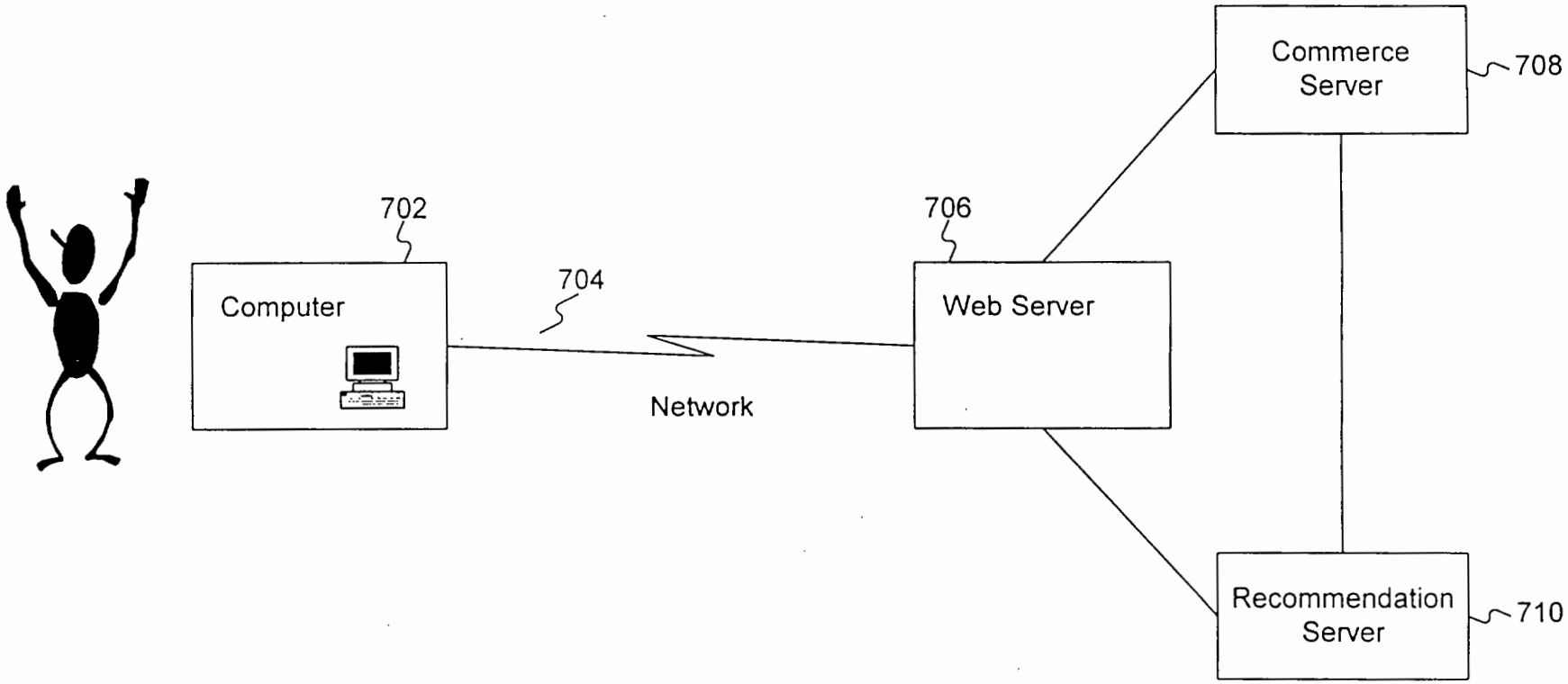
Commerce Server
708
702
706
704
Computer
Web Server
Network
Recommendation Server
710

FIG. 7A

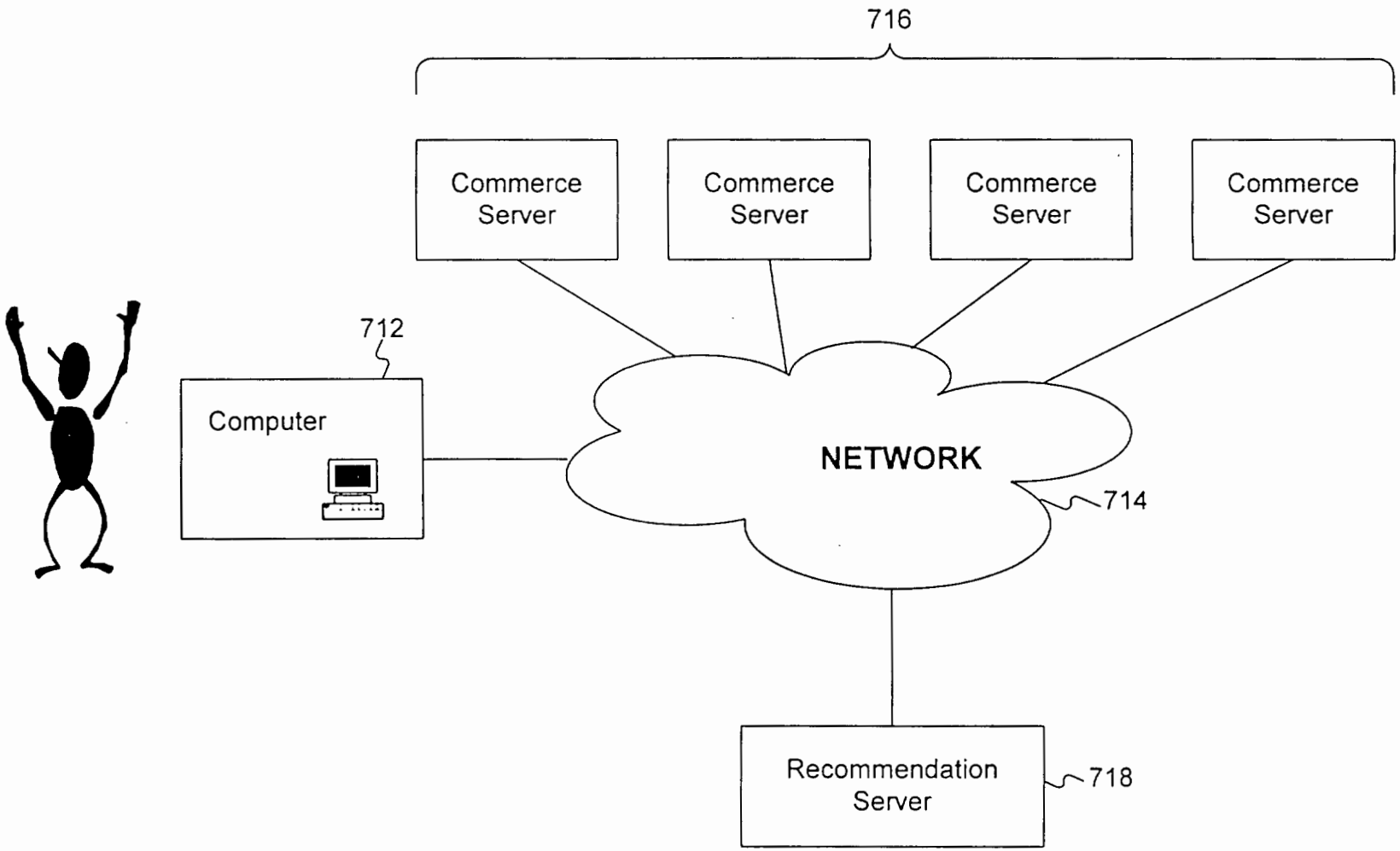
716
Commerce Server
Commerce Server
Commerce Server
Commerce Server
712
Computer
NETWORK
714
Recommendation Server
718

FIG. 7B