

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
7 February 2008 (07.02.2008)

PCT

(10) International Publication Number
WO 2008/016523 A2

(51) International Patent Classification:
H04J 1/00 (2006.01)

(74) Agent: **ROGITZ, John, L.**; Rogitz & Associates, 750 B Street, Suite 3120, San Diego, CA 92101 (US).

(21) International Application Number:
PCT/US2007/016727

(22) International Filing Date: 24 July 2007 (24.07.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/835,020 1 August 2006 (01.08.2006) US
11/602,566 21 November 2006 (21.11.2006) US

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(71) Applicants (for all designated States except US): **SONY CORPORATION** [JP/JP]; 1-7-1 Konan, Minato-ku, Tokyo, 108-0075 (JP). **SONY ELECTRONICS INC.** [US/US]; 1 Sony Drive, Park Ridge, NJ 07656 (US).

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

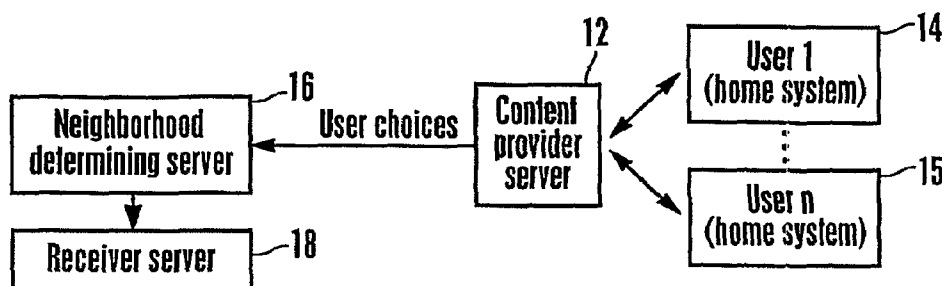
(72) Inventors; and

(75) Inventors/Applicants (for US only): **GEORGIS, Nikolaos** [GR/US]; 11588 Windcrest Lane, Apt. 1123, San Diego, CA 92128 (US). **HWANG, Paul, Jin** [US/US]; 554 E. San Jose Avenue, #205, Burbank, CA 91501 (US). **LI-DE LIN, Frank** [—/US]; 28322 Glenmeade Way, Escondido, CA 92026 (US).

Published:

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

(54) Title: SYSTEM AND METHOD FOR NEIGHBORHOOD OPTIMIZATION FOR CONTENT RECOMMENDATION



(57) Abstract: A cost function is stochastically optimized using, e.g., simulated annealing to render a neighborhood of entities based on which content recommendations can be provided to a user of a home entertainment system (14). The cost function represents a normalized sum of rating similarity scores from entities (15) of the neighborhood that are related to content items viewed by the user.



WO 2008/016523 A2

SYSTEM AND METHOD FOR NEIGHBORHOOD OPTIMIZATION FOR CONTENT RECOMMENDATION

This application claims priority from U.S. provisional patent application serial no. 60/835,020, filed August 1, 2006, incorporated herein by reference.

I. Field of the Invention

The present invention relates generally to systems and methods for content recommendation.

2. Background of the Invention

Systems and methods have been developed to recommend content to users of home entertainment systems based on ratings of the content from neighbors (e.g., other home systems in a cable TV system or other entertainment network). It will readily be appreciated that to recommend content to a particular user, selecting the neighborhood members is important, because if the wrong neighborhood is used, the recommendations may not be very useful to the particular person to whom they are made.

As understood herein, current methods of neighborhood selection are not as optimum as they might be. For example, in collaborative filtering (CF), opinions from users in the form of ratings of items are collected, and when the system is asked for a recommendation, the system identifies similar users based on, e.g., similarity of demographics to suggest the items these users have liked in the past. This method is based only on the judgments of the user neighborhood.

User similarity can be estimated using cosine-based similarity between two users, but as recognized herein, relying solely on this method, the number of neighbors typically must be defined in advance without any good way to know how many neighbors is optimal.

Thus, the number of neighbors typically is fixed in advance without knowledge of the optimum number of neighbors that might benefit an individual user. Complicating the issue is the fact that the total number of ways of forming a reasonably sized neighborhood of one thousand other users is greater than the number of atoms in the known universe, so

that, as recognized herein, at best a pseudo-optimal neighborhood feasibly can be defined. It is to this problem that the present invention is directed.

SUMMARY OF THE INVENTION

A method is disclosed for recommending content to a home system for display of the content on a display device of the home system. The method includes defining a neighborhood of other users using a cost function described below, and based on the neighborhood, providing at least one content recommendation to a user of the home system.

In some embodiments one term of the cost function represents a sum of rating similarity scores from entities of the neighborhood and related to content items viewed by the user. The cost function may be normalized. If desired, the rating similarity scores can be cosine-based similarity scores, and the cost function can be optimized using a stochastic method such as simulated annealing. In some embodiments the simulated annealing can be undertaken using operators described below and referred to as "move", "swap", "swap-swap", and "swap-move".

In another aspect, a system receives ratings from entities that collectively comprise potential neighbors of a subject user in a neighborhood. The system returns content recommendations to the subject user based thereon. The system includes at least one server that is programmed to stochastically establish a pseudo-optimum neighborhood based on the ratings, and to return content recommendations based on the neighborhood.

In yet another aspect, a computer iteratively computes a cost function representing a sum of rating similarity scores from entities of a neighborhood of a home entertainment system and related to content items viewed by a user of the home entertainment system. The neighborhood is used to recommend content to the user of the home entertainment system.

The details of the present invention, both as to its structure and operation, can best be understood in reference to the accompanying drawings, in which like reference numerals refer to like parts, and in which:

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a block diagram of a non-limiting system in accordance with the present invention; and

Figure 2 is a flow chart of the present logic.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

Referring initially to Figure 1, a system is shown, generally designated 10, that includes a content provider server 12 such as but not limited to a cable head end server that can provide content in an intelligent way to a primary user home system 14. The primary home system 14 may include one or more multimedia display devices such as televisions and/or computers and one or more multimedia or content data stores such as DVRs and disk players such as digital video disk (DVD) or Blu-Ray players, etc. that can supply content for display on the display device. Without limitation, the user system interface may be implemented by a set-top box. In lieu of or in addition to a STB, the user system interface may also be implemented by an Internet connection device such as a wired or wireless modem or other type of wide area network connection. Thus, communication between the primary home system 14 and servers described below may be via Internet and/or TV cable and/or broadcast links, both terrestrial and satellite, and can be two-way, e.g., the primary home system 14 can receive content from the server 12 and communicate back to the server 12 ratings for particular pieces of content.

As shown in Figure 1, not only the primary home system 14 but neighbor home systems 15 can communicate with the server 12. Like the primary home system 14, the neighbor systems 15 can receive content from the server 12 and communicate back to the server 12 ratings for particular pieces of content. The purpose of this invention to define a near-optimal ("pseudo-optimal") subset of neighbor systems 15 as a "neighborhood" of the primary home system 14, the ratings from which can be used to recommend content to the primary home system 14. It is to be understood that the same process can be employed for each neighbor system 15, wherein the neighbor system 15 would be the primary home system in the below-described algorithms.

In the non-limiting implementation shown, the content server 12 can communicate with a neighborhood determining server 16 that essentially is a computer which executes the logic below to determine a near-optimal neighborhood. The near-optimal neighborhood can be sent to a recommendation server 18 that can return content recommendations to the primary home system 14 using the neighborhood in accordance with principles known in the art. While three servers 12, 16, 18 are shown, greater or fewer servers can be used.

Turning now to Figure 2, at block 20 a cost function C is established. In accordance with present principles, the preferred cost function is given in equation (1). The cost function is used to stochastically establish a pseudo-optimum neighborhood based on ratings from the neighborhood systems 15. The right-most sum in equation (1) represents the sum of rating similarity scores (which can use cosine-based methods or other similarity methods) from the neighborhood systems 15 that are related to content items viewed by a user of the primary home entertainment system 14, and the middle sum essentially is a normalization factor. Stated differently, the cost function C to be optimized is the sum of scores for all content items already in the primary system 14 user's bin for a given neighborhood N , and its range (due to normalization) is from zero to one, with one indicating that an ideally optimal neighborhood has been found.

$$(1) \quad C = f(N) = \frac{1}{|N|} \sum_{s \in \tilde{S}} \left(\frac{1}{\sum_{c' \in \tilde{C}} |sim(c, c')|} \times \sum_{c' \in \tilde{C}} sim(c, c') \times r_{c', s} \right)$$

The above equation in words is C (to be maximized) = $1/N$ times (the sum over all movies s that have been rated) $\{1/(\text{sum over all candidate neighbors } 15 \text{ in the current neighborhood of the absolute value of the similarity between the rating from each candidate neighbor } 15 \text{ and the rating from the primary system } 14) \text{ times the } [(\text{sum over all candidate neighbors } 15 \text{ in the current neighborhood of the similarity between the rating from each candidate neighbor } 15 \text{ and the rating from the primary system } 14) \text{ times the rating from the candidate neighbor for the movie being summed}]\}$.

Once the cost function has been defined, the logic moves to block 22 to define an initial subset of all the neighbor systems 15 as an initial "neighborhood" for the primary system 14. A subset may be selected at random. Proceeding to block 24, the cost function above is calculated a first time.

At block 26, an operator is selected with which the neighborhood is perturbed. In one non-limiting implementation, one of four operators may be selected at block 26. A non-neighborhood one of the systems 15 can be into the neighborhood ("move"), with a "move" also encompassing removing a system 15 from the current neighborhood. The second operator ("swap") involves swapping a system 15 who is not in the current neighborhood with a system 15 who is, while a third operator ("swap-swap") involves swapping a system 15 who is not in the current neighborhood with a system 15 who is in the neighborhood twice in succession. A fourth operator ("swap-move") requires executing a swap then a move in immediate succession.

Proceeding to block 28, after perturbation by an operator at block 26 the cost function is recalculated. Proceeding to decision diamond 30, it is determined whether the value of the cost function increased over the prior value, and if it did the perturbation is accepted at block 32. If however the cost function did not increase at decision diamond 30, the logic moves to decision diamond 34 to accept or not the perturbation with a probability defined by equation (2). Based on the probability and a random draw to compare a number with the probability, the perturbation may be accepted at block 32 or not accepted at block 36 in essence backed out, so that the neighborhood resumes the membership it had prior to perturbing it at block 26.

$$(2) \quad P = e^{-\frac{\Delta C}{T}}$$

$P = e^{-(\text{change in } C)/T}$, where "T" is an empirically determined annealing "temperature."

Decision diamond 38 determines whether the probability is sufficiently low (e.g., below an empirically-determined value) to indicate that a near-optimum neighborhood has

been attained. If so, the process ends at state 40 to output the current neighborhood as that on which to base recommendations to the primary system 14. Otherwise, the logic loops back to block 26.

If desired, incremental cost function evaluation can be used, as can lookup tables. Different operators than those set forth herein may also be used, and fast annealing can be used. Moreover, a standard user-based cost function solution can be used to define the initial neighborhood, and other cost functions may be used.

It may now be appreciated that the present invention can be used as an enhancement to current engines, e.g., to current cosine-based similarity engines, without the need for pair-wise comparison for user/item similarity estimations. Near optimal (also referred to herein as "pseudo-optimal") solutions can be obtained using the present stochastic algorithms. Differently sized neighborhoods advantageously can be used for different users to obtain better solutions can be obtained than existing methods.

As an alternative to simulated annealing, genetic algorithms, so-called ant colony optimization, and other methods may be used.

While the particular SYSTEM AND METHOD FOR NEIGHBORHOOD OPTIMIZATION FOR CONTENT RECOMMENDATION is herein shown and described in detail, it is to be understood that the subject matter which is encompassed by the present invention is limited only by the claims.

WHAT IS CLAIMED IS:

1. A method for recommending content to a home system (14) for display of the content on a display device of the home system (14), comprising:

defining a neighborhood of other users using the cost function:

$$C = f(N) = \frac{1}{|N|} \sum_{s \in S} \left(\frac{1}{\sum_{c' \in \hat{C}} |sim(c, c')|} \times \sum_{c' \in \hat{C}} sim(c, c') \times r_{c', s} \right)$$

based on the neighborhood, providing at least one content recommendation to a user of the home system (14).

2. The method of Claim 1, wherein one term of the cost function represents a sum of rating similarity scores from entities of the neighborhood and related to content items viewed by the user.

3. The method of Claim 1, wherein the cost function is normalized.

4. The method of Claim 2, wherein the rating similarity scores are cosine-based similarity scores.

5. The method of Claim 1, wherein the cost function is optimized using a stochastic method.

6. The method of Claim 5, wherein the stochastic method is simulated annealing.

7. The method of Claim 6, wherein the simulated annealing is undertaken using at least two operators selected from the group consisting of moving a non-neighborhood entity into the neighborhood ("move"), moving a neighborhood entity out of the neighborhood ("move"), swapping an entity who is not in the neighborhood with an entity who is in the neighborhood ("swap"), swapping an entity who is not in the neighborhood with an entity who is in the neighborhood twice in succession ("swap-swap"), and executing a swap then a move in immediate succession.

8. A system receiving ratings from entities that collectively comprise potential neighbors (15) of a subject user in a neighborhood and returning content recommendations to the subject user at least partially based thereon, comprising:

at least one server (12, 16, 18) programmed to stochastically establish a pseudo-optimum neighborhood based at least in part on the ratings, and to return content recommendations at least partially based on the neighborhood.

9. The system of Claim 8, wherein the server (12, 16, 18) defines the neighborhood using a cost function, wherein one term of the cost function represents a sum of rating similarity scores from entities of the neighborhood and related to content items viewed by the user.

1/1

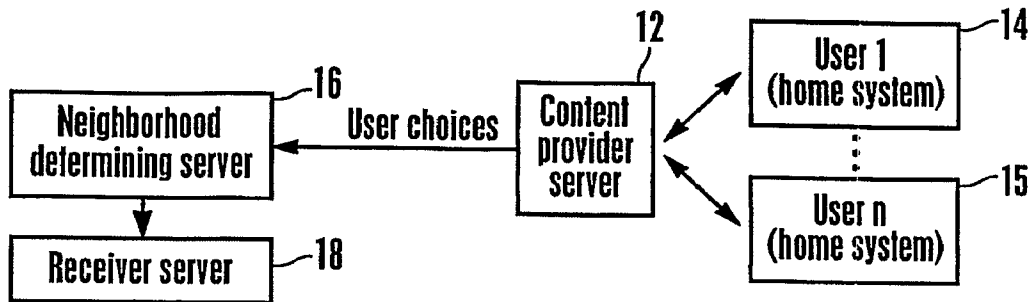


Figure 1

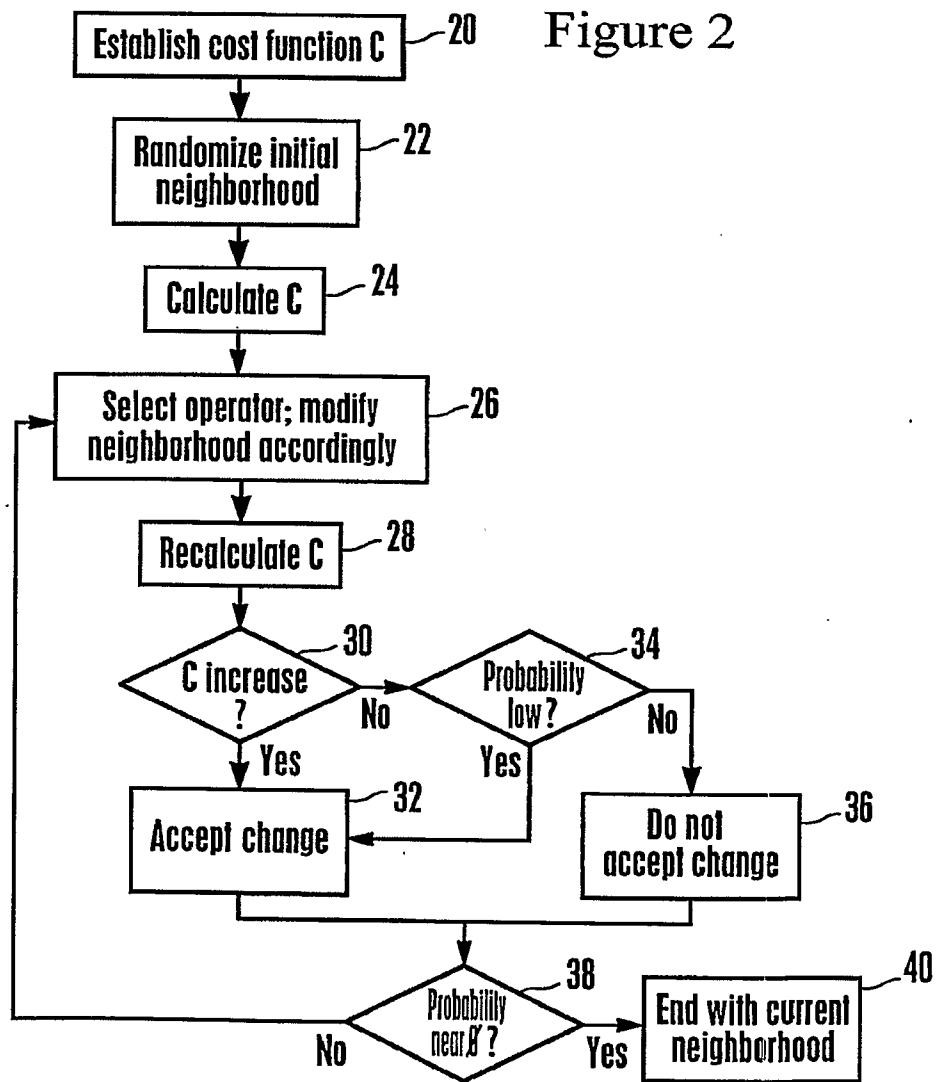


Figure 2