



- (51) International Patent Classification:
G06F 17/30 (2006.01)
- (21) International Application Number:
PCT/CN2015/098147
- (22) International Filing Date:
21 December 2015 (21.12.2015)
- (25) Filing Language: English
- (26) Publication Language: English
- (72) Inventor; and
- (71) Applicant : ZHENG, Zhaohui [CN/CN]; Fenglinlvzhou
Apartment 8-15B, Datun Road, Chaoyang District, Beijing
100101 (CN).
- (74) Agent: HANHOW INTELLECTUAL PROPERTY
PARTNERS; ZHU, Haibo, W1-1111, F/11, Oriental
plaza, No.1 East Chang An Avenue, Dongcheng District,
Beijing 100738 (CN).
- (81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,

BZ, CA, CH, CL, 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, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— with international search report (Art. 21(3))

(54) Title: METHOD AND SYSTEM FOR EXPLORING A PERSONAL INTEREST SPACE

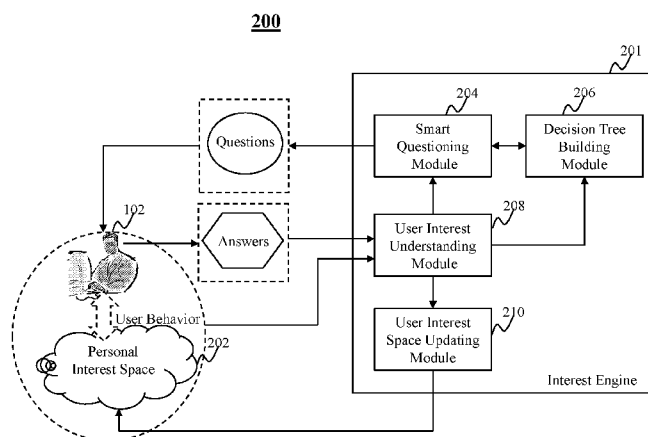


FIG. 2

(57) Abstract: The present teaching relates to exploring a personal interest space based on a decision tree. In one example, information related to a personal interest space associated with a user is obtained. The personal interest space includes at least one channel of content that the user is interested in. A decision tree created based on information related to a plurality of users with respect to a plurality of channels of content is obtained. Each node of the decision tree is associated with at least one of the plurality of channels of content. A first node of the decision tree is identified based on the information related to the personal interest space associated with the user. A question for the user in connection with the first node of the decision tree is determined. An answer to the question is received from the user. A second node of the decision tree is identified based on the answer. The personal interest space associated with the user is updated based on the at least one channel of content associated with the second node of the decision tree.

WO 2017/107026 A1

METHOD AND SYSTEM FOR EXPLORING A PERSONAL INTEREST SPACE**BACKGROUND**

1. Technical Field

[0001] The present teaching relates to methods, systems, and programming for Internet service. More specifically, the present teaching relates to methods, systems, and programming for exploring a personal interest space.

2. Discussion of Technical Background

[0002] The development of the Internet leads to information explosion, where the information and data online are unorganized and have a huge volume. Accordingly, it is difficult to identify useful or interesting information for a particular user from the unorganized and abundant information. Efforts have been made to filter online content provided to a user based on the user's interests.

[0003] For example, FIG. 1 depicts a prior art search engine that estimates a user 102's interests based on search queries. A search engine 104 may include a search module 106 that searches documents for specified queries and returns a list of documents where the queries are found. The search engine 104 cannot provide content to the user 102 without a query submitted by the user 102. Some search engines also include a user interest understanding module 108 that estimates the user 102's interests based on the keywords specified in the search queries. The estimated user interests may be used by the search engine 104 to filter and refine search results returned by the search module 106. However, the interest estimation is usually not precise and comprehensive as it depends completely on what the user 102 inputs in the search query. Moreover, the search keywords are a reflection of the user 102's current short-term interests and usually do not indicate the long-term interests of the user 102.

[0004] In another example illustrated in FIG. 1, a recommendation engine 110 tries to estimate the user 102's interests by analyzing the user 102's behavior with respect to online content in the content space 112. Some known recommendation engines 110 include a user interest understanding module 114, which uses algorithms such as collaborative filtering to "guess" the possible interests of the user 102. The content recommendation module 116 then retrieves content that falls into the possible interests of the user 102 and recommends them to the user 102. However, prediction accuracy of the traditional recommendation engine 110 is largely relied on the amount of

user 102's past behavior data that the recommendation engine 110 can obtain and also the specific algorithm it applies, which oftentimes may not be accurate.

[0005] Therefore, there is a need to provide an effective solution for exploring a user's personal interest space to avoid the above-mentioned drawbacks.

SUMMARY

[0006] The present teaching relates to methods, systems and programming for Internet service. More specifically, the present teaching relates to methods, systems, and programming for exploring a personal interest space based on a decision tree.

[0007] In one example, a method, implemented on at least one computing device each of which has at least one processor, storage, and a communication platform connected to a network for exploring a personal interest space is presented. Information related to a personal interest space associated with a user is obtained. The personal interest space includes at least one channel of content that the user is interested in. A decision tree created based on information related to a plurality of users with respect to a plurality of channels of content is obtained. Each node of the decision tree is associated with at least one of the plurality of channels of content. A first node of the decision tree is identified based on the information related to the personal interest space associated with the user. A question for the user in connection with the first node of the decision tree is determined. An answer to the question is received from the user. A second node of the decision tree is identified based on the answer. The personal interest space associated with the user is updated based on the at least one channel of content associated with the second node of the decision tree.

[0008] In another example, a method, implemented on at least one computing device each of which has at least one processor, storage, and a communication platform connected to a network for creating a decision tree is presented. A plurality of pieces of information related to a plurality of users with respect to a plurality of channels of content are obtained. Each of the plurality of pieces of information is indicative of a level at which a respective one of the plurality of users is interested in each of the plurality of channels. A decision tree is created based on the plurality of pieces of information in accordance with a training model. Each node of the decision tree is associated with at least one of the plurality of channels of content.

[0009] In a different example, a system for exploring a personal interest space is presented. The system includes a user understanding unit, a decision tree selecting unit, a question initiating unit, a question selecting unit, a decision tree traversing unit, and a decision tree traversing

unit. The user understanding unit is configured for obtaining information related to a personal interest space associated with a user. The personal interest space includes at least one channel of content that the user is interested in. The decision tree selecting unit is configured for obtaining a decision tree created based on information related to a plurality of users with respect to a plurality of channels of content. Each node of the decision tree is associated with at least one of the plurality of channels of content. The question initiating unit is configured for identifying a first node of the decision tree based on the information related to the personal interest space associated with the user. The question selecting unit is configured for determining a question for the user in connection with the first node of the decision tree. The decision tree traversing unit is configured for identifying a second node of the decision tree based on an answer to the question from the user. A decision tree traversing unit is configured for updating the personal interest space associated with the user based on the at least one channel of content associated with the second node of the decision tree.

[0010] Other concepts relate to software for implementing the present teaching on exploring a personal interest space. A software product, in accord with this concept, includes at least one non-transitory, machine-readable medium and information carried by the medium. The information carried by the medium may be executable program code data, parameters in association with the executable program code, and/or information related to a user, a request, content, or information related to a social group, etc.

[0011] In one example, a non-transitory, machine-readable medium having information recorded thereon for exploring a personal interest space is presented. The recorded information, when read by the machine, causes the machine to perform a series of processes. Information related to a personal interest space associated with a user is obtained. The personal interest space includes at least one channel of content that the user is interested in. A decision tree created based on information related to a plurality of users with respect to a plurality of channels of content is obtained. Each node of the decision tree is associated with at least one of the plurality of channels of content. A first node of the decision tree is identified based on the information related to the personal interest space associated with the user. A question for the user in connection with the first node of the decision tree is determined. An answer to the question is received from the user. A second node of the decision tree is identified based on the answer. The personal interest space associated with the user is updated based on the at least one channel of content associated with the second node of the decision tree.

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] The methods, systems, and/or programming described herein are further described in terms of exemplary embodiments. These exemplary embodiments are described in detail with reference to the drawings. These embodiments are non-limiting exemplary embodiments, in which like reference numerals represent similar structures throughout the several views of the drawings, and wherein:

[0013] FIG. 1 depicts a prior art search engine that estimates a user's interests based search queries and a prior art recommendation engine that estimates a user's interests based on user behavior;

[0014] FIG. 2 is a high level exemplary system diagram of a system for exploring a personal interest space, according to an embodiment of the present teaching;

[0015] FIG. 3 is a flowchart of an exemplary process performed by the system for exploring a personal interest space, according to an embodiment of the present teaching;

[0016] FIGs. 4 and 5 are depictions of an exemplary process of exploring a personal interest space based on a decision tree, according to an embodiment of the present teaching;

[0017] FIG. 6 is a depiction of exemplary nodes of a decision tree, according to an embodiment of the present teaching;

[0018] FIG. 7 illustrates an exemplary diagram of a smart questioning module, according to an embodiment of the present teaching;

[0019] FIG. 8 is a flowchart of an exemplary process of creating a personal decision forest for a user, according to an embodiment of the present teaching;

[0020] FIG. 9 is a flowchart of an exemplary process of determining timings for asking questions, according to an embodiment of the present teaching;

[0021] FIG. 10 is a flowchart of an exemplary process of smart questioning based on a personal decision forest, according to an embodiment of the present teaching;

[0022] FIG. 11 is a depiction of exemplary training data for decision tree training, according to an embodiment of the present teaching;

[0023] FIG. 12 is a depiction of an exemplary process of building decision trees based on user subset training data, according to an embodiment of the present teaching;

[0024] FIG. 13 is a depiction of an exemplary process of building decision trees based on channel subset training data, according to an embodiment of the present teaching;

[0025] FIG. 14 illustrates an exemplary diagram of a decision tree building module, according to an embodiment of the present teaching;

[0026] FIG. 15 is a flowchart of an exemplary process of building a decision tree, according to an embodiment of the present teaching;

[0027] FIG. 16 is a high level depiction of an exemplary networked environment for exploring a personal interest space, according to an embodiment of the present teaching;

[0028] FIG. 17 is a high level depiction of another exemplary networked environment for exploring a personal interest space, according to an embodiment of the present teaching;

[0029] FIG. 18 depicts the architecture of a mobile device which can be used to implement a specialized system incorporating the present teaching; and

[0030] FIG. 19 depicts the architecture of a computer which can be used to implement a specialized system incorporating the present teaching.

DETAILED DESCRIPTION

[0031] In the following detailed description, numerous specific details are set forth by way of examples in order to provide a thorough understanding of the relevant teachings. However, it should be apparent to those skilled in the art that the present teachings may be practiced without such details. In other instances, well-known methods, procedures, components, and/or circuitry have been described at a relatively high-level, without detail, in order to avoid unnecessarily obscuring aspects of the present teachings.

[0032] The present teaching describes methods, systems, and programming aspects of exploring a user's personal interest space based on smart, personalized, and adaptive questioning in accordance with a decision forest, realized as a specialized and networked system by utilizing one or more computing devices (e.g., mobile phone, personal computer, etc.) and network communications (wired or wireless). From the user's perspective, every time the user answers a question, the content stream may be dynamically updated to reflect the better understanding of the user's interests.

[0033] In one aspect of the present teaching, information related to existing users' interests with respect to channels of content in their personal interest spaces are utilized as training data to create a decision forest (having one or more decision trees) for exploring a user's personal interest space. The information used for training the decision forest is reliable and accurate because it comes from existing user's personal interest spaces that reflect the existing users' long-term and stable interests with respect to each channel of content. Thus, the decision forest is more reliable

and accurate than the models used by the known search engines and recommendation engines for estimating users' interests.

[0034] In another aspect of the present teaching, a user's personal interest space is explored (e.g., by refining existing channels of content, adding new channels of content, etc.) in an interactive manner which involves the user's active participation in the process. For example, smart and fruitful questions are asked at the right timings based on the trained decision forest, and the answers from the user are used as a guidance to update the user's personal interest space. To achieve better user experience and efficiency, the questioning process is personalized for each user in the sense that the decision forest (having one or more decision trees) used for generating the questions may be selected for each individual user and the specific question to be asked may be tailored to fit each individual user's profile and known interests. The questioning process is also performed in an adaptive way as the user's answers to the previous questions may affect the next question to be asked.

[0035] Additional novel features will be set forth in part in the description which follows, and in part will become apparent to those skilled in the art upon examination of the following and the accompanying drawings or may be learned by production or operation of the examples. The novel features of the present teaching may be realized and attained by practice or use of various aspects of the methodologies, instrumentalities and combinations set forth in the detailed examples discussed below.

[0036] FIG. 2 is a high-level exemplary system diagram of a system for exploring a personal interest space, according to an embodiment of the present teaching. The system 200 in this embodiment includes an interest engine 201 that creates and updates a personal interest space 202 for a user 102. The system 200 is a technology-driven, interest-oriented content portal, "interest portal" in short, which facilitates users to efficiently define and explore their personal interest spaces. The interest portal may be supported by the interest engine 201 which seamlessly integrates the cutting-edge technologies of search and recommendation to enable a unique and personalized content consumption experience for the user 102. The interest engine 201 may evolve the user 102's interests by continuously learning from the user 102's interactions with the recommended content as well as the social circles, and algorithmically deliver the most personally relevant content with respect to the user 102's interests.

[0037] The interest engine 201 in this embodiment can organize online information into different channels of content. Each channel includes content items that are related to a common topic or category or have a common feature, e.g., popular items, trending items, etc. The

interest engine 201 may periodically crawl online for news or other content from the Internet, and classify the crawled content items into different channels, according to their respective topics. This operation may not need any involvement of the user 102. A channel may also include one or more sub-channels, each of which includes content items associated with a sub-topic of the topic associated with the channel. For example, a channel associated with topic “sports” may include sub-channels associated with topics like “tennis,” “basketball,” “swimming,” “running,” etc. A channel may be correlated to another channel, based on a correlation between their respective topics. For example, channel “travel” may be correlated with channel “flight ticket,” and channel “children” may be correlated with channel “minivan.” Similarly, a sub-channel may be correlated with another sub-channel.

[0038] The interest engine 201 may store all channels assigned to the user 102 into the personal interest space 202 associated with the user 102. In addition to information about the channels, the personal interest space 202 may also include information about, for example, correlations between the channels, other users connected to the user 102, one or more channels shared by the user 102 with other users, channel weights associated with the channels, etc. A channel weight associated with a channel in a personal interest space 202 may indicate how important the channel is to the user 102 or how interesting the channel is to the user 102. Different users can have different personal interest spaces.

[0039] Accordingly, the interest engine 201 may also store different content channels associated with metadata. For example, metadata associated with a channel may include information about, for example, dynamic content provided by the interest engine 201 through the channel, sub-channels in the channel, subscribed users of the channel, correlations between the channel and other channels, an assigned advertiser for the channel, etc. The interest engine 201 may assign an advertiser to a channel based on a bidding process performed with respect to a plurality of advertisers. For example, based on a bidding process performed with respect to Apple Inc. and Samsung Group, the interest engine 201 may assign one of them to channel “smartphone” such that the assigned advertiser has authority to put its advertisements into the channel.

[0040] The interest engine 201 may provide dynamic content to a subscribed user 102 through the channel. Dynamic content refers to content within a channel, when the interest engine 201 can keep updating content in the channel based on newly obtained information from the Internet and/or activities of a subscribed user 102. After the interest engine 201 assigns a channel to a personal interest space of a user 102, the channel may become a personal channel for the user 102, such that content provided through the channel to the user 102 may be different from content

provided to other users subscribed to the channel. For example, two users subscribed to channel “electronics” may gradually receive different content items provided by the interest engine, after one user clicks most articles in the channel about smartphones and the other user clicks most articles in the channel about digital cameras.

[0041] A user 102 may subscribe to a channel (or a sub-channel) provided by the interest engine 201. In other words, the interest engine 201 may assign a channel to a user 102 after the user 102 submits a subscription request for the channel. In addition, the interest engine 201 may assign one or more channels to a user 102 based on, e.g., a query submitted by the user 102 for channels, a channel previously assigned to the user 102, a channel subscribed to by the user 102, content items browsed or viewed by the user 102, channels associated with other users, or channels selected randomly for testing the user 102’s interest. Details of an interest engine 201 are provided, for example, in PCT Patent Application No. PCT/CN2015/089288, filed September 9, 2015, entitled “METHOD AND SYSTEM FOR PROVIDING ORGANIZED CONTENT,” which is incorporated herein by reference in its entirety.

[0042] In this embodiment, the interest engine 201 includes a smart questioning module 204, a decision tree building module 206, a user interest understanding module 208, and a user interest space updating module 210. The modules mentioned above are used for exploring the user 102’s personal interest space 202 via an interactive process (e.g., questions and answers) based on a decision forest having one or more decision trees.

[0043] The smart questioning module 204 in this embodiment ensures that for each specific user 102, the right questions are asked at the right timings in the right context so that the interest engine 201 can explore the user 102’s person interest space 202 and better understand the user 102’s interests. In one embodiment, the growth of the user 102’s personal interest space 202 may reach a plateau/bottle neck, for example, when the user 102’s interaction with the personal interest space 202 decreases. This may be a right timing for the smart questioning module 204 to ask a question for the user 102 in order to encourage the user 102 to actively express her/his interests and further explore the personal interest space 202. In another embodiment, when a new user 102 signs-up for the interest engine 201, it may be a right timing for the smart questioning module 204 as well to encourage the new user 102 to actively express her/his interests and start to interact with the interest engine 201 in order to build up the personal interest space 202. On the other hand, for better user experience, the smart questioning module 204 may also control the timer intervals between asking each question. For example, based on each user’s historical behavior with respect to questions, a threshold time interval (e.g., once per day, once per week, etc.) may be set up.

[0044] The smart questioning module 204 in this embodiment, at the right timings, selects the right questions that are deemed to be friendly to the user 102 and can gain a better understanding of the user's 102 interests. Based on information of the user 102, e.g., the user 102's personal interest space 202 and personal profile, the smart questioning module 204 may identify one or more nodes in one or more decision trees created by the decision tree building module 206. Each node is associated with a channel of content. Based on the node and the channel associated therewith, the smart questioning module 204 may select, from a plurality of preset questions associated with the node/channel, a question that fits for the user 102. When selecting the question, the smart questioning module 204 may consider the user 102's personal profile to avoid duplicated or contradicting questions. In order to attract the user 102's to answer the question, the smart questioning module 204 may also consider the user 102's known interests and personal profile to select the most interesting question for the user 102.

[0045] The decision tree building module 206 in this embodiment serves as a backend module to support the smart questioning module 204. Based on existing users' personal interest spaces, the decision tree building module 206 can generate a channel vector for each existing user as a piece of training data. Each feature in the channel vector corresponds to the existing user's level of interest with respect to a respective channel. The number of features (channels) may be predefined, e.g., including 10,000 channels. The values of each channel feature may be determined based on the existing user's behavior with respect to the channels in its personal interest space. For example, if the existing user has subscribed to a particular channel, then the value of this channel feature may be "1" in a binary value scale. And if the existing user not only has subscribed to the channel but also actively consumed the content recommended in the channel, then a high positive value may be assigned, e.g., 0.9 in a (-1 to 1) value scale. In another example, if the existing user has explicitly expressed that she/he is not interested in a channel, e.g., by declining the recommendation of the channel or unsubscribing the channel, then a value of "0" or a low negative value may be assigned to this channel feature depending on the value scales used. In still another example, if there is no indication as to whether the existing user is interested in the channel or not, then an "unassigned" value may be assigned to the corresponding channel feature initially. Channel vectors of multiple existing users may be combined to form a training data matrix.

[0046] Referring now to FIG. 11, an exemplary training data matrix 1102 is illustrated. The training data matrix 1102 is formed by channel vectors $1-n$ of users $1-n$ as shown in FIG. 11. Each channel vector comprises values of m channel features ($1-m$). In this example, a binary value scale is used, so that the value of each channel feature may be "1," "0," or "?"

(unassigned)” depending on whether the user is interested in the corresponding channel. It is understood that in other embodiments, the value may reflect the level of interest with respect to the channel in non-binary value scales. As shown in FIG. 11, each row of the training data matrix 1102 represents one channel vector for a specific user with respect to all the channel features used (1- m), and each column of the training data matrix 1102 represents all users (1- n)’ levels of interest with respect to a specific channel. For example, the training data matrix 1102 may show that user i has subscribed to channels 5 and 7, declined to subscribe to channel 8, and not shown whether she/he is interested in channel 8. In this embodiment, the sequence of channel subscription may be stored with the training data matrix 1102 as well. For example, the training data matrix 1102 may show that user i subscribed to channel 7 after she/he subscribed to channel 5. Similarly, for example, the training data matrix 1102 may show that user k has subscribed to channels 8, 6, and 5 in sequence and not shown whether she/he is interested in channel 7.

[0047] Returning to FIG. 2, the decision tree building module 206 then uses the training data matrix formed of channel vectors of a number of existing users to train a decision tree. In this embodiment, expectation-maximization (EM) models may be used to train the decision tree based on the training data matrix, and the details will be discussed later. Each node of the decision tree is associated with a channel of content and a user segment. The granularity of the decision tree becomes finer as traversing down the decision tree.

[0048] It is understood that the decision tree is created based on the specific data matrix used for training. Thus, in some embodiments, the decision tree building module 206 may create multiple decision trees to form a decision forest based on different training data matrices. In one example, the decision tree building module 206 may select a subset of existing users sharing the same or similar profiles (e.g., a male user subset, a young professional user subset, etc.) and train a decision tree that is considered to be more relevant to the specific user group. In another example, the decision tree building module 206 may select a subset of channel features (e.g., channel features related to sports) and train a decision tree that is considered to be more relevant to the specific channel group. Accordingly, the decision tree building module 206 can create a generic decision tree and/or various specific decision trees according to different training data sets.

[0049] Referring now to FIG. 12, a decision tree may be trained based on a training user set with all the existing users. The training user set may be partitioned into m training user subsets (1- m). In this example, the partitioning may be performed based on the users’ age groups. For example, training user subset 1 includes existing users below 30, and training user subset m includes existing users above 60. For each training user subset, the corresponding training data

matrix may be formed and used to train a decision tree 1- m . Each decision tree 1- m may be considered to be more relevant to the users in the corresponding user age group in this example. In the future, in order to explore a user's personal interest space, a decision tree that matches with the user's age group may be selected from the decision trees 1- m and used to determine questions to be asked for the user.

[0050] Similarly, in FIG. 13, a decision tree may be trained based on a channel set with a relatively large number of channel features (e.g., a generic channel set). The channel set may be partitioned into n channel subsets (1- n). For example, channel subset 1 includes health-related channels, and channel subset n includes sports-related channels. For each channel subset, the corresponding training data matrix may be formed and used to train a decision tree 1- n . Each decision tree 1- n may be considered to be more relevant to the channels in the corresponding channel subset in this example. In the future, in order to explore a user's personal interest space, a decision tree that matches with the user's currently-subscribed channels may be selected from the decision trees 1- m and used to determine questions to be asked for the user. It is understood that, in some embodiments, in order to explore the user's known interests, a decision tree that does not match with the user's currently-subscribed channels may be intentionally chosen and used to determine questions to be asked for the user. It is also understood that the partition of a training data set may be performed in both the user and channel dimensions in some embodiments.

[0051] Returning to FIG. 2, the user interest understanding module 208 in this embodiment is responsible for understanding the user 102's interests based on answers from the user 102 in response to the questions asked by the smart questioning module 204. The user interest understanding module 208 may also estimate the user's interests by analyzing user behavior with respect to the personal interest space 202. An answer from the user 102 in response to a question is guidance on how to traverse the decision tree and find out the corresponding nodes and their associated channels of content. The identified nodes are provided to the smart questioning module 204 so that the smart questioning module 204 can adaptively choose the next question to be asked according to the decision tree. It is understood that in some embodiments, the answer may not be an explicitly received from the user 102. Instead, the user 102 may ignore the questions, but perform certain actions that can be implied as answering to the question. For example, the question is related to whether the user 102 is interested in drone. Instead of providing an explicit answer, the user 102 may start to search channels related to drone or browser content related to drone, then the user interest understanding module 208 may treat the user's actions as an implicit answer to the question. In some embodiments, if neither an explicit nor an implicit answer (e.g., user actions) is

received from the user 102 for a particular question, the smart questioning module 204 may ask the same question, or a different question associated with the same node, after a certain time interval. If no response has been received after several tries, the smart questioning module 204 may identify a different node in the decision tree and ask a different question.

[0052] The identified channels of content are provided to the user interest space updating module 210 for updating the personal interest space 202 with a better understanding of the user 102's interests. For example, a new channel may be added to the personal interest space 202, an existing channel may be refined to focus on a more specific topic, and multiple channels may be merged to form a more generic channel. It is understood that the personal interest space 202 of the user 102 is dynamically updated upon receiving each answer from the user 102. In some embodiments, the recommended content stream is instantly updated as well according to the dynamically updated personal interest space 202.

[0053] FIG. 3 is a flowchart of an exemplary process performed by the system for exploring a personal interest space, according to an embodiment of the present teaching. Starting at 302, user information is obtained. The user information may include demographic information, such as age, gender, occupation, race, etc., the user's social connections, the user's declaration interests, and the user's historical behavior data. At 304, initial channels are allocated to the user based on the obtained user information. For example, one or more channels may be recommended to the user for subscription. At 306, a personal interest space is generated for the user. The user interest space may include at least some of the initial channels that have been explicitly subscribed to by the user. In some embodiments, the user may actively participate in the process of generating the personal interest space by, for example, searching and subscribing to certain channels of interest even though they are not recommended to the user. At 308, the personal interest space associated with the user is stored. At 310, user behavior with respect to the personal interest space is monitored, and the user behavior information is obtained. The user behavior includes, for example, consumption of content provided in the subscribed channels. At 312, the personal interest space is updated based on the user behavior information. In other words, channels in the personal interest space may be dynamically changed over time.

[0054] On the other hand, at 314, information of multiple existing users' personal interest spaces is collected. Each of the existing users' personal interest spaces may be used to generate a channel vector for the existing user, which is indicative of a level at which the existing user is interested in each of the plurality of channels. At 316, a decision forest having one or more decision trees is built based on the information obtained at 314. For example, the information may

be converted into a training data matrix. A decision tree may be trained based on the training data matrix using an EM model. At 318, a personal decision forest may be determined for the user. The personal decision forest may include a generic decision tree trained based on training data from all existing users and one or more specific decision trees trained based on subsets of training data related to a specific user group or a channel feature group. The specific decision trees may be selected based on the user's information, e.g., the user's demographic information and the user's currently-subscribed channels in the personal interest space. At 320, smart questions are generated and asked for the user based on the user's information and the user's personal decision forest. At 322, answers to the questions are received from the user. The answers may be mapped to one or more channels of content according to the personal decision forest. Continuing to 312, the personal interest space is updated based on the channels associated with the user's answers to the smart questions.

[0055] FIGs. 4 and 5 are depictions of an exemplary process of exploring a personal interest space based on a decision tree, according to an embodiment of the present teaching. As shown in FIG. 4, a user's personal interest space 400 is updated according to a decision tree 401 over times. In this example, the personal interest space 400 initially includes two channels, i.e., a sports channel 402 and a pop music channel 404, which correspond to nodes A and B of the decision tree 401, respectively. If it is observed that the exploration of the user's personal interest space 400 has reached a bottle neck (e.g., no change in the personal interest space 400 for a while), questions may be asked based on nodes A and B associated with the channels 402, 404, respectively. In this example, the question related to node A may be used to get a better understanding of the user's specific interests within the sports channel 402. If the answer indicates that the user is only interested in basketball among all sports, then the user's sports channel 402 is refined to become a more specific channel within the sports channel 402 – a basketball channel 408, which is associated with node C. Similarly, the question related to node B may be used to find other channels that are related to the pop music channel 404. If the answer indicates that the user is also interested in classical music, then a new classical music channel 410 is added to the user's personal interest space 400, which is associated with node D of the decision tree 401.

[0056] Content items are then recommended to the user based on the updated personal interest space 400, which now includes the basketball channel 408, the pop music channel 404, and the classical music channel 410. In this example, the exploration of the user's personal interest space 400 may reach another bottle neck after a while, which invokes the smart questioning process again based on the decision tree 401. Nodes C and D are used to determine the questions to

be asked at this time. If the user's answer to the question associated with node C indicates that the user is also interested in fantasy games (node E), then a new fantasy games channel 412 is added. Similarly, the user's answer to the question associated with node D may indicate that the user is in fact interested in music in general. As a result, a music channel 414 (node F) is generated by merging the existing pop music channel 404 and classical music channel 410.

[0057] As shown in FIG. 5, a decision tree 502 is created by the decision tree building module 206. Each node of the decision tree 502 is associated with a channel of content. Traversing down the decision tree 502 from the root node 1, each node is connected to two low-level nodes. For each node of the decision tree 502, the smart questioning module 204 may select one question to ask the user 102. For example, question 1 related to the root node 1 is asked for the user 102, and the user 102's answer 1 guides the decision tree 502 to be traversed down to node 2. Subsequently, question 2 related to node 2 is asked for the user 102, and the user 102's answer 2 further guides the path of traversing down the decision tree 502 towards node 3. As the decision tree 502 is traversed down from the root node 1 based on the user 102's answers to each question, a better understanding of the user 102's interests is achieved.

[0058] FIG. 6 is a depiction of exemplary nodes of a decision tree, according to an embodiment of the present teaching. Three exemplary nodes 602, 604, 606 are extracted from a decision tree. As the decision tree is created by training a matrix including a plurality of channel vectors for a plurality of users, each node of the decision tree is associated with one of the plurality of channels and a user segment of the plurality of users. For example, node 602 is associated with a technology channel and a user segment 610. The technology channel is one of the features of the training vectors. The user segment 610 includes users who have exhibited a certain level of interest with respect to the technology channel (e.g., subscribed to the technology channel in her/his personal interest space). In this example, the node 602 is associated with a preset question 608. The question 608 may be, "What will you bring with you if you are put on the Moon alone?" with the choices of "A: Xbox 360" and "B: iPod." For better user experience, a question that directly asks for a user's interests may be avoided. Instead, questions that are considered to be interesting and funny may be used. It is also understood that more than two choices may be provided to the user. In other words, each node of a decision tree may be connected to more than two low-level nodes. It is also understood that the user may be allowed to provide more than one answer to a question. That is, more than one node may be identified when traversing down the decision tree based on each of the user's answers.

[0059] In this example, answers A and B to the question 608 lead to node 604 and node 606, respectively. Node 604 is associated with a gaming channel, and node 606 is associated with a music channel. Node 604 is associated with a user segment 612, which includes users who have exhibited a certain level of interest with respect to the technology channel and the gaming channel. Similarly, node 606 is associated with a user segment 614, which includes users who have exhibited a certain level of interest with respect to the technology channel and the music channel. As mentioned above, each node 604, 606 may be associated with one or more questions as well so that the decision tree can be further traversed down from the node 604 and/or 606 to the low-level nodes.

[0060] FIG. 7 illustrates an exemplary diagram of a smart questioning module, according to an embodiment of the present teaching. The smart questioning module 204 in this embodiment works in conjunction with the decision tree building module 206 and the user interest understanding module 208 to ask the user 102 the right questions at the right timings in order to explore the user 102's personal interest space 202. The smart questioning module 204 in this embodiment includes a questioning timing unit 702, an interest growth analyzing unit 704, a question initiating unit 706, a user understanding unit 708, a decision tree selecting unit 710, a question selecting unit 712, and a decision tree traversing unit 714.

[0061] The questioning timing unit 702 in this embodiment is configured to determine the timing of asking each question. The timing may be determined based on various factors, for example, the growth of the user 102's personal interest space 202 and the time intervals between questions. The interest growth analyzing unit 704 may be responsible for analyzing the growth of the personal interest space 202 to see if it reaches a plateau/bottle neck. For example, the personal interest space 202 may have not been updated for a while because the user 102 has not interacted with the personal interest space 202 by consuming recommended content or searching and subscribing to any channels. A timer 716 may be used by the questioning timing unit 702 to control the time interval between asking two subsequent questions. A general time interval threshold may be used for all users. In some embodiments, by observing a specific user 102's responses to previous questions, a personalized time interval threshold may be applied for each user. For example, if a user 102 has shown an interest in answering questions asked by the smart questioning module 204 before, then a short time interval threshold may be used for the user 102 in the future as she/he will not likely feel bothered by the questions.

[0062] The decision tree selecting unit 710 in this embodiment is configured to select one or more decision trees from the decision forest 720 to create a personal decision forest

718 for the user 102. As mentioned above, the decision forest 720 may include a plurality of decision trees trained by the decision tree building module 206 including, for example, a general decision tree database 720-1, a channel subset based decision tree database 720-2, and a user subset based decision tree database 720-3. The selection made by the decision tree selecting unit 710 may be based on the understanding of the user 102 performed by the user understanding unit 708. The user understanding unit 708 may provide information related to the user 102's subscribed channels in the personal interest space 202 and information related to the user profile 722. In one embodiment, the decision tree selecting unit 710 may select one or more user subset based decision trees based on the user profile 722. For example, if the user 102 is a young, male professional, then the decision tree selecting unit 710 may choose the decision trees that are trained based on training data of existing users who are also young, male professionals. In another embodiment, the decision tree selection unit 710 may select one or more channel subset based decision trees based on the user 102's currently-subscribed channels. For example, if the user 102 has subscribed to a number of channels related to politics, then the decision tree selecting unit 710 may select the decision trees that are trained based on training data with the channel features related to politics as well. In this embodiment, the decision tree selecting unit 710 may select the generic decision tree as well.

[0063] The question initiating unit 706 in this embodiment is configured to identify an initiate node on a decision tree in the personal decision forest 718 based on the information related to the user 102 that is provided by the user understanding unit 708. In one embodiment, if the user 102 is a new user, and very limited information related to the user 102 has been provided, then the question initiating unit 706 may set the root node of a decision tree as the starting point. Otherwise, the question initiating unit 706 may identify one or more nodes of a decision tree based on a matching between the user 102's subscribed channels and the channels associated with each node of the decision tree. Moreover, other user information, such as the user profile 722, may be used to help the question initiating unit 706 to identify the suitable initiate node. For example, if there is a large confidence that the user 102 is a male, then any node that is associated with a female-related channel may be ruled out as being an initiate node by the question initiating unit 706.

[0064] The question selecting unit 712 in this embodiment is configured to determine a question for the user 102 in connection with the node identified by the question initiating unit 706. As mentioned above, each node of a decision tree is associated with a channel of content. For each channel of content, a plurality of relevant questions may be predetermined and stored in the question database 724. Although multiple questions may be associated with the same node/channel, each of them may be designed to be asked for particular type of users. The question

selecting unit 712 may first identify a plurality of questions in connection with a particular node identified by the question initiating unit 706 and then based on the information of the user 102 provided by the user understanding unit 708, further select one of those questions that the user 102 is most likely to answer. In other words, for example, even for the same node, different questions may be asked for a female user and a male user. Additional rules may be applied by the question selecting unit 712 to select the right question for the user 102. For example, a new question cannot contradict with previously asked questions or be duplicated with the previously asked questions.

[0065] In this embodiment, on one hand, the answers to the questions are provided by the user 102 to the user interest understanding module 208 in order to update the personal interest space 202. On the other hand, the decision tree traversing unit 714 in this embodiment is configured to identify additional nodes on the decision tree based on the answers to the previous questions provided by the user 102. In one embodiment, the additional nodes are identified by traversing down the decision tree from the initiate node based on the answers to the previous questions. The identified additional nodes and channels associated therewith are provided to the question selecting unit 712 for selecting corresponding questions. Each question is asked at a timing determined by the questioning timing unit 702. It is understood that the granularity of the decision tree becomes finer as the decision tree is traversed down. Thus, in other words, generic questions are asked before the fine and specific questions are asked.

[0066] FIG. 8 is a flowchart of an exemplary process of creating a personal decision forest for a user, according to an embodiment of the present teaching. Starting at 802, information related to a user is obtained. The information may include the user profile. At 804, a user subset based decision tree is selected based on the user's information. At 806, the personal interest space of the user is retrieved. At 808, channels subscribed by the user are determined from the user's personal interest space. At 810, a channel subset based decision tree is selected based on the user's subscribed channels. At 812, a generic decision tree is retrieved. At 814, a personal decision forest including the generic decision tree, the user subset based decision tree, and the channel subset based decision tree is created for the user.

[0067] FIG. 9 is a flowchart of an exemplary process of determining timings for asking questions, according to an embodiment of the present teaching. Starting at 902, the growth of a user's personal interest space is analyzed. At 904, whether the growth of the user's personal interest space has reached a plateau/bottle neck is determined. If the answer is negative, the process returns to 902. Otherwise, at 906, the time interval from asking the last question is checked. At 908, whether the time interval is above the time interval threshold is determined. If the answer is

negative, the process returns to 902. Otherwise, at 910, the user's current channel vector is determined based on the user's subscribed channels in the personal interest space. The value of each channel feature is indicative of the level of interest of the user with respect to the corresponding channel. At 912, one or more nodes in the user's personal decision forest are identified based on the user's current channel vector. At 914, one or more questions are selected based on the identified one or more nodes.

[0068] FIG. 10 is a flowchart of an exemplary process of smart questioning based on a personal decision forest, according to an embodiment of the present teaching. Starting at 1002, a user's current channel vector is determined. The channel vector may be determined based on the user's subscribed channels in the personal interest space. The value of each channel feature is indicative of the level of interest of the user with respect to the corresponding channel. At 1004, a node in the user's personal decision forest is identified based on the user's current channel vector. At 1006, information of the user, e.g., the user profile, is obtained. At 1008, whether there is any contradiction between the user's information and the identified node is checked. If the answer is positive, the process returns to 1004 to identify a different node. Otherwise, at 1010, a question is selected based on the node and the information of the user. At 1012, an answer to the question is obtained from the user. At 1014, another node in the personal decision forest is identified based on the answer, for example, by traversing down the decision tree from the previous node. At 1016, another question is selected based on the node identified at 1014 and the information of the user.

[0069] FIG. 14 illustrates an exemplary diagram of a decision tree building module, according to an embodiment of the present teaching. The decision tree building module 206 in this embodiment includes a user channel vector building unit 1402, a training data partitioning unit 1404, a decision tree training unit 1406, and a decision tree pruning unit 1408. The user channel vector building unit 1402 is configured to retrieve information of existing user's personal interest spaces 1410 and generate channel vectors for each of the existing users based on the levels of interest of the existing user with respect to each of a plurality of channels. As mentioned above in detail, the value of each channel feature in the channel vector may be a numeric value indicative of the level of interest of the existing user with respect to the channel if the existing user has interacted with the channel (e.g., by searching and subscribing to a channel or by declining to subscribe to a recommended channel). If the existing user has not interacted with the channel, i.e., the level of interest with respect to the channel is unknown for the existing use, then an "unassigned" value may be initially assigned in the channel vector. All the existing user's channel vectors are stored in the user channel vector database 1412.

[0070] In this embodiment, the training data partitioning unit 1404 may separate the data in the user channel vector database 1412 into various data subsets in the user dimension and/or the channel dimension. For example, predetermined user subsets 1414 (e.g., age groups, genders, occupations, etc.) and channel subsets 1416 (e.g., sports, music, politics, etc.) may be used by the training data partitioning unit 1404 for dividing the training data set. A training data matrix 1418 is generated based on the divided training data set. It is understood that the training data set may not be divided by the training data partitioning unit 1404 in some embodiments so that the training data matrix 1418 represents all existing user's channel vectors with respect to all channels used by the interest engine 201.

[0071] The decision tree training unit 1406 in this embodiment is responsible for creating a decision tree 1422 based on the training data matrix 1418 in accordance with a training model 1420. The training model 1420 may be, for example, an EM model. In this example, for all "unassigned" values in the training data matrix 1418, a value of "1" or "0" may be randomly assigned. An EM algorithm may be used to iteratively train a decision tree using the training data matrix 1418 until the termination condition is met (converged). At the end, the decision tree 1422 is obtained, and each channel feature of each user has either an observed or assigned value. In some embodiments, the decision tree 1422 may be a probability based decision tree.

[0072] In one embodiment, one EM algorithm includes the following steps:

- (1) Randomly assign 1/0 (or simply assign 0) to each <user, channel feature> pair of which the channel feature value is missing.
- (2) Build a root-mean-square-error (RMSE) decision tree with the above data set. The RMSE loss function is the weighted combination of (a) true label loss function and (b) assigned label loss function, where (a) should have a higher weight than (b).
- (3) For each user, assign values to all her/his missing channel features according to the leaf node predictions (in case of binary representation, the values would be the $[(\# \text{ of node mates with positive value} / \# \text{ of node mates with negative value}) / \text{global ratio}] > \text{threshold?} 1:0$). The default threshold could be 1, but subject to empirical tuning.
- (4) Recalculate the channel feature probability distribution on the entire data set and update the global ratio.
- (5) Repeat the above steps 2, 3, and 4 until the termination condition is met.

[0073] In another embodiment, another EM algorithm includes the following steps:

- (1) According to a predetermined probability, randomly assign 1/0 to each <user, channel feature> pair of which the channel feature value is missing.

- (2) Build a RMSE decision tree with the above data set. The RMSE loss function is the weighted combination of (a) true label loss function, (b) assigned positive loss function, and (c) assigned negative loss function, where (a) should have a higher weight than (b) and (c), and (b) and (c) have an equal weight.
- (3) For each user, assign values to all her/his missing channel features according to its leaf node predictions (in case of binary representation, the values would be the (# of node mates with positive value > # of node mates with negative value?1:0).
- (4) Recalculate the channel feature probability distribution on the entire data set, and accordingly maintain the predetermined probability by reassigning weights to assigned labels.
- (5) Repeat the above steps 2, 3 and 4 until the termination condition is met.

[0074] The decision tree pruning unit 1408 in this embodiment may adjust the trained decision tree 1422. For example, duplicated nodes in the decision tree 1422 may be merged. The adjustment may be performed manually based on human knowledge or automatically or semi-automatically based on preset rules and conditions. The pruned decision tree is stored in the decision forest 720, which is to be used by the smart questioning module 204.

[0075] FIG. 15 is a flowchart of an exemplary process of building a decision tree, according to an embodiment of the present teaching. Starting at 1502, information about multiple existing user's personal interest spaces is collected. At 1504, training data of channel vectors for the existing users is generated based on the collected information. At 1506, the training data is partitioned into user subsets and/or channel subsets. At 1508, matrices are created for each partition of the training data. At 1510, an EM model is obtained for each matrix. At 1512, a decision tree is trained base on each matrix using the obtained EM model. At 1514, the trained decision tree is pruned. At 1516, the decision tree is stored for future use.

[0076] FIG. 16 is a high-level depiction of an exemplary networked environment for exploring a personal interest space, according to an embodiment of the present teaching. In FIG. 16, the exemplary networked environment 1600 includes one or more users 102, a network 1602, the interest engine 201, a search/recommendation engine 1604, one or more advertisers 1608, and content sources 1606. The network 1602 may be a single network or a combination of different networks. For example, the network 1602 may be a local area network (LAN), a wide area network (WAN), a public network, a private network, a proprietary network, a Public Telephone Switched Network (PSTN), the Internet, a wireless network, a virtual network, or any combination thereof.

[0077] Users 102 may be of different types such as users connected to the network 1602 via desktop computers 102-4, laptop computers 102-3, a built-in device in a motor vehicle

102-2, or a mobile device 102-1. In one embodiment, users 102 may be connected to the network 1602 and able to access and interact with online content provided by the interest engine 201 through wired or wireless technologies and related operating systems implemented within user-wearable devices (e.g., glasses, wrist watch, etc.). A user 102 may send a request for online content to the interest engine 201 and receive content organized in channels through the network 1602. The user 102 may also receive a question from the interest engine 201 and send the answer to the interest engine 201 via the network 1602.

[0078] The interest engine 201 in this embodiment may serve as a backend for an interest portal associated with an entity, which may be an individual, a firm, or an organization, such as a television station, a newspaper issuer, a web page host, an online service provider, an app manager, or a game server. For example, the interest engine 201 can be provided by an organization such as USPTO.gov, a content provider such as CNN.com and google.com, or a content-feed source such as tweeter or blogs. In one embodiment, the interest engine 201 may be provided by entities that develop, support and/or provide online content via mobile applications (e.g., installed on smartphones, tablet devices, etc.). Content provided by the interest engine 201 may include, for example, graphics, data, news, articles, software-based products, and so forth. The interest engine 201 may provide software or app used to access the content.

[0079] Referring to the above example, the content sent to user 102 may be generated by the interest engine 201 based on the content sources 1606. A content source 1606 may correspond to an entity where the content was originally generated and/or stored. For example, a novel may be originally printed in a magazine, but then posted online at a web site controlled by a content provider managing the interest engine 201. The content sources 1606 in the exemplary networked environment 1600 include multiple content sources 1606-1, 1606-2... 1606-3.

[0080] A user 102 may also submit a search query to the interest engine 201 via an interest portal. The interest engine 201 may forward the query to the search/recommendation engine 1604 for performing a search based on the query. In one embodiment, the query may be used to search for a channel provided by the interest engine 201. When there is no channel matching the query, the interest engine 201 may provide an option for the user 102 to create a new channel based on the query and subscribe to the new channel. When the user 102 is entering the query via the interest portal, the interest engine 201 may provide query suggestions each of which is a name of a channel stored at the interest engine 201. In another embodiment, the query may be used to search for content items within one or more channels provided by the interest engine 201. The interest engine 201 may also cooperate with the search/recommendation engine 1604 to

recommend a channel or a content item within a channel to a user 102. The recommendation may be based on the user 102's assigned channels, the user 102's activities, or other information related to the user 102. It is understood that the search/recommendation engine 1604 may also perform regular search and recommendation operations based on unorganized content online. It can also be understood that in some embodiments, the search engine and the recommendation engine are separate from each other.

[0081] When content is sent to the user 102, one or more advertising opportunities may be available for one or more advertisements to be presented with the content to the user 102, through a same channel, on a same web page, via a same application, or in a same user session. For an available advertising opportunity, the interest engine 201 may send a request to advertisers 1608 to ask for advertisements. In one embodiment, the request may include a bid request, such that the interest engine 201 can sell the advertising opportunity by soliciting bids from the advertisers 1608. An advertiser 1608 may correspond to an entity, whether an individual, a firm, or an organization, doing or planning to do (or otherwise involved in) advertising business. As such, an advertiser 1608 may be an entity that provides product(s) and/or service(s), and itself handles the advertising process for its own product(s) and/or service(s) at a platform (e.g., websites, mobile applications, etc.). For example, advertisers 1608-1...1608-2 may include companies like General Motors, BestBuy, or Disney. In some other cases, however, an advertiser 1608 may be an entity that only handles the advertising process for product(s) and/or service(s) provided by another entity. In one embodiment, the advertisers 1608 have stored their advertisements at an advertisement database connected to the interest engine 201.

[0082] FIG. 17 is a high-level depiction of another exemplary networked environment for exploring a personal interest space, according to an embodiment of the present teaching. The exemplary networked environment 1700 in this embodiment is similar to the exemplary networked environment 1600 in FIG. 16, except that the search/recommendation engine 1604 serves as a backend system for the interest engine 201.

[0083] FIG. 18 depicts the architecture of a mobile device which can be used to realize a specialized system implementing the present teaching. In this example, a device of the user 102 used for interacting with the interest engine 201 may be a mobile device 1800, including, but is not limited to, a smart phone, a tablet, a music player, a handheld gaming console, a global positioning system (GPS) receiver, and a wearable computing device (e.g., eyeglasses, wrist watch, etc.), or in any other form. The mobile device 1800 in this example includes one or more central processing units (CPUs) 1802, one or more graphic processing units (GPUs) 1804, a display 1806, a

memory 1808, a communication platform 1810, such as a wireless communication module, storage 1812, and one or more input/output (I/O) devices 1814. Any other suitable component, such as but not limited to a system bus or a controller (not shown), may also be included in the mobile device 1800. As shown in FIG. 18, a mobile operating system 1816, e.g., iOS, Android, Windows Phone, etc., and one or more applications 1818 may be loaded into the memory 1808 from the storage 1812 in order to be executed by the CPU 1802. The applications 1818 may include a web browser or any other suitable mobile apps used for interacting with the interest engine 201. Execution of the applications 1818 may cause the mobile device 1800 to perform some processing as described in the present teaching. For example, user inputs may be received via the I/O devices 1814 and sent to the interest engine 201 via the communication platform 1810. Presentation of the recommended content to the user 102 may be made by the GPU 1804 in conjunction with the display 1806.

[0084] To implement the present teaching, computer hardware platforms may be used as the hardware platform(s) for one or more of the elements described herein. The hardware elements, operating systems, and programming languages of such computers are conventional in nature, and it is presumed that those skilled in the art are adequately familiar therewith to adapt those technologies to implement the processing essentially as described herein. A computer with user interface elements may be used to implement a personal computer (PC) or other type of workstation or terminal device, although a computer may also act as a server if appropriately programmed. It is believed that those skilled in the art are familiar with the structure, programming, and general operation of such computer equipment and as a result the drawings should be self-explanatory.

[0085] FIG. 19 depicts the architecture of a computing device which can be used to realize a specialized system implementing the present teaching. The computer may be a general-purpose computer or a special purpose computer. This computer 1900 can be used to implement any components of the personal interest space exploration architecture as described herein. Different components of the system, e.g., as depicted in FIG. 2, can all be implemented on one or more computers such as computer 1900, via its hardware, software program, firmware, or a combination thereof. Although only one such computer is shown, for convenience, the computer functions relating to medical record completion may be implemented in a distributed fashion on a number of similar platforms, to distribute the processing load.

[0086] The computer 1900, for example, includes COM ports 1902 connected to and from a network connected thereto to facilitate data communications. The computer 1900 also includes a CPU 1904, in the form of one or more processors, for executing program instructions.

The exemplary computer platform includes an internal communication bus 1906, program storage and data storage of different forms, e.g., disk 1908, read only memory (ROM) 1910, or random access memory (RAM) 1912, for various data files to be processed and/or communicated by the computer, as well as possibly program instructions to be executed by the CPU 1904. The computer 1900 also includes an I/O component 1914, supporting input/output flows between the computer and other components therein such as user interface elements 1916. The computer 1900 may also receive programming and data via network communications.

[0087] Hence, aspects of the method of personal interest space exploration, as outlined above, may be embodied in programming. Program aspects of the technology may be thought of as “products” or “articles of manufacture” typically in the form of executable code and/or associated data that is carried on or embodied in a type of machine readable medium. Tangible non-transitory “storage” type media include any or all of the memory or other storage for the computers, processors or the like, or associated modules thereof, such as various semiconductor memories, tape drives, disk drives and the like, which may provide storage at any time for the software programming.

[0088] All or portions of the software may at times be communicated through a network such as the Internet or various other telecommunication networks. Such communications, for example, may enable loading of the software from one computer or processor into another. Thus, another type of media that may bear the software elements includes optical, electrical, and electromagnetic waves, such as used across physical interfaces between local devices, through wired and optical landline networks and over various air-links. The physical elements that carry such waves, such as wired or wireless links, optical links or the like, also may be considered as media bearing the software. As used herein, unless restricted to tangible “storage” media, terms such as computer or machine “readable medium” refer to any medium that participates in providing instructions to a processor for execution.

[0089] Hence, a machine readable medium may take many forms, including but not limited to, a tangible storage medium, a carrier wave medium or physical transmission medium. Non-volatile storage media include, for example, optical or magnetic disks, such as any of the storage devices in any computer(s) or the like, which may be used to implement the system or any of its components as shown in the drawings. Volatile storage media include dynamic memory, such as a main memory of such a computer platform. Tangible transmission media include coaxial cables; copper wire and fiber optics, including the wires that form a bus within a computer system. Carrier-wave transmission media can take the form of electric or electromagnetic signals, or

acoustic or light waves such as those generated during radio frequency (RF) and infrared (IR) data communications. Common forms of computer-readable media therefore include for example: a floppy disk, a flexible disk, hard disk, magnetic tape, any other magnetic medium, a CD-ROM, DVD or DVD-ROM, any other optical medium, punch cards paper tape, any other physical storage medium with patterns of holes, a RAM, a PROM and EPROM, a FLASH-EPROM, any other memory chip or cartridge, a carrier wave transporting data or instructions, cables or links transporting such a carrier wave, or any other medium from which a computer can read programming code and/or data. Many of these forms of computer readable media may be involved in carrying one or more sequences of one or more instructions to a processor for execution.

[0090] Those skilled in the art will recognize that the present teachings are amenable to a variety of modifications and/or enhancements. For example, although the implementation of various components described above may be embodied in a hardware device, it can also be implemented as a software only solution- e.g., an installation on an existing server. In addition, the personal interest space exploration system and its components as disclosed herein can be implemented as a firmware, firmware/software combination, firmware/hardware combination, or a hardware/firmware/software combination.

[0091] While the foregoing has described what are considered to constitute the present teachings and/or other examples, it is understood that various modifications may be made therein and that the subject matter disclosed herein may be implemented in various forms and examples, and that the teachings may be applied in numerous applications, only some of which have been described herein. It is intended by the following claims to claim any and all applications, modifications and variations that fall within the true scope of the present teachings.

WE CLAIM:

1. A method, implemented on a machine having at least one processor, storage, and a communication platform connected to a network for exploring a personal interest space, the method comprising:

obtaining information related to a personal interest space associated with a user, wherein the personal interest space comprises at least one channel of content that the user is interested in;

obtaining a decision tree created based on information related to a plurality of users with respect to a plurality of channels of content, wherein each node of the decision tree is associated with at least one of the plurality of channels of content;

identifying a first node of the decision tree based on the information related to the personal interest space associated with the user;

determining a question for the user in connection with the first node of the decision tree;

receiving an answer to the question from the user;

identifying a second node of the decision tree based on the answer; and

updating the personal interest space associated with the user based on the at least one channel of content associated with the second node of the decision tree.

2. The method of claim 1, further comprising:

analyzing a growth of the personal interest space associated with the user;

determining a timing of asking the question based on the growth of the personal interest space associated with the user; and

asking the user the question at the determined timing.

3. The method of claim 1, wherein the question for the user is selected from a plurality of questions in connection with the first node based on the information related to the personal interest space and/or a profile of the user.

4. The method of claim 1, wherein the second node is identified by traversing down the decision tree from the first node based on the answer to the question.

5. The method of claim 1, wherein the information related to the plurality of users with respect to the plurality of channels of content comprises a plurality of vectors, each of which is

indicative of a level at which a respective one of the plurality of users is interested in each of the plurality of channels.

6. The method of claim 5, wherein the decision tree is created by a training model based on the plurality of vectors.

7. The method of claim 1, wherein the decision tree is selected from a plurality of decision trees based on the information related to the personal interest space and/or a profile of the user.

8. A system for exploring a personal interest space, comprising:
a user understanding unit configured for obtaining information related to a personal interest space associated with a user, wherein the personal interest space comprises at least one channel of content that the user is interested in;
a decision tree selecting unit configured for obtaining a decision tree created based on information related to a plurality of users with respect to a plurality of channels of content, wherein each node of the decision tree is associated with at least one of the plurality of channels of content;
a question initiating unit configured for identifying a first node of the decision tree based on the information related to the personal interest space associated with the user;
a question selecting unit configured for determining a question for the user in connection with the first node of the decision tree;
a decision tree traversing unit configured for identifying a second node of the decision tree based on an answer to the question from the user; and
a user interest updating module configured for updating the personal interest space associated with the user based on the at least one channel of content associated with the second node of the decision tree.

9. The system of claim 8, further comprising:
an interest growth analyzing unit configured for analyzing a growth of the personal interest space associated with the user; and
a questioning timing unit configured for determining a timing of asking the question based on the growth of the personal interest space associated with the user.

10. The system of claim 8, wherein the question for the user is selected from a plurality of questions in connection with the first node based on the information related to the personal interest space and/or a profile of the user.

11. The system of claim 8, wherein the second node is identified by traversing down the decision tree from the first node based on the answer to the question.

12. The system of claim 8, wherein the information related to the plurality of users with respect to the plurality of channels of content comprises a plurality of vectors, each of which is indicative of a level at which a respective one of the plurality of users is interested in each of the plurality of channels.

13. The system of claim 12, wherein the decision tree is created by a training model based on the plurality of vectors.

14. The system of claim 8, wherein the decision tree is selected from a plurality of decision trees based on the information related to the personal interest space and/or a profile of the user.

15. A non-transitory, machine-readable medium having information recorded thereon for exploring a personal interest space, wherein the information, when read by a machine, causes the machine to perform the steps of:

obtaining information related to a personal interest space associated with a user, wherein the personal interest space comprises at least one channel of content that the user is interested in;

obtaining a decision tree created based on information related to a plurality of users with respect to a plurality of channels of content, wherein each node of the decision tree is associated with at least one of the plurality of channels of content;

identifying a first node of the decision tree based on the information related to the personal interest space associated with the user;

determining a question for the user in connection with the first node of the decision tree;

receiving an answer to the question from the user;

identifying a second node of the decision tree based on the answer; and

updating the personal interest space associated with the user based on the at least one channel of content associated with the second node of the decision tree.

16. The medium of claim 15, wherein the information, when read by the machine, further causes the machine to perform the steps of:

- analyzing a growth of the personal interest space associated with the user;
- determining a timing of asking the question based on the growth of the personal interest space associated with the user; and
- asking the user the question at the determined timing.

17. The medium of claim 15, wherein the question for the user is selected from a plurality of questions in connection with the first node based on the information related to the personal interest space and/or a profile of the user.

18. The medium of claim 15, wherein the second node is identified by traversing down the decision tree from the first node based on the answer to the question.

19. The medium of claim 15, wherein the information related to the plurality of users with respect to the plurality of channels of content comprises a plurality of vectors, each of which is indicative of a level at which a respective one of the plurality of users is interested in each of the plurality of channels.

20. The medium of claim 19, wherein the decision tree is created by a training model based on the plurality of vectors.

21. The medium of claim 15, wherein the decision tree is selected from a plurality of decision trees based on the information related to the personal interest space and/or a profile of the user.

22. A method, implemented on a machine having at least one processor, storage, and a communication platform connected to a network for creating a decision tree, the method comprising:
obtaining a plurality of pieces of information related to a plurality of users with respect to a plurality of channels of content, wherein each of the plurality of pieces of information is indicative

of a level at which a respective one of the plurality of users is interested in each of the plurality of channels; and

creating a decision tree based on the plurality of pieces of information in accordance with a training model, wherein each node of the decision tree is associated with at least one of the plurality of channels of content.

23. The method of claim 22, wherein

each of the plurality of pieces of information includes a vector having values of a plurality of features determined based on the level at which the respective user is interested in the respective channel of content.

24. The method of claim 22, wherein each node of the decision tree is associated with a subset of the plurality of users who share a same value of at least one of the plurality of features of their vectors.

25. The method of claim 22, wherein the training model includes an expectation-maximization model.

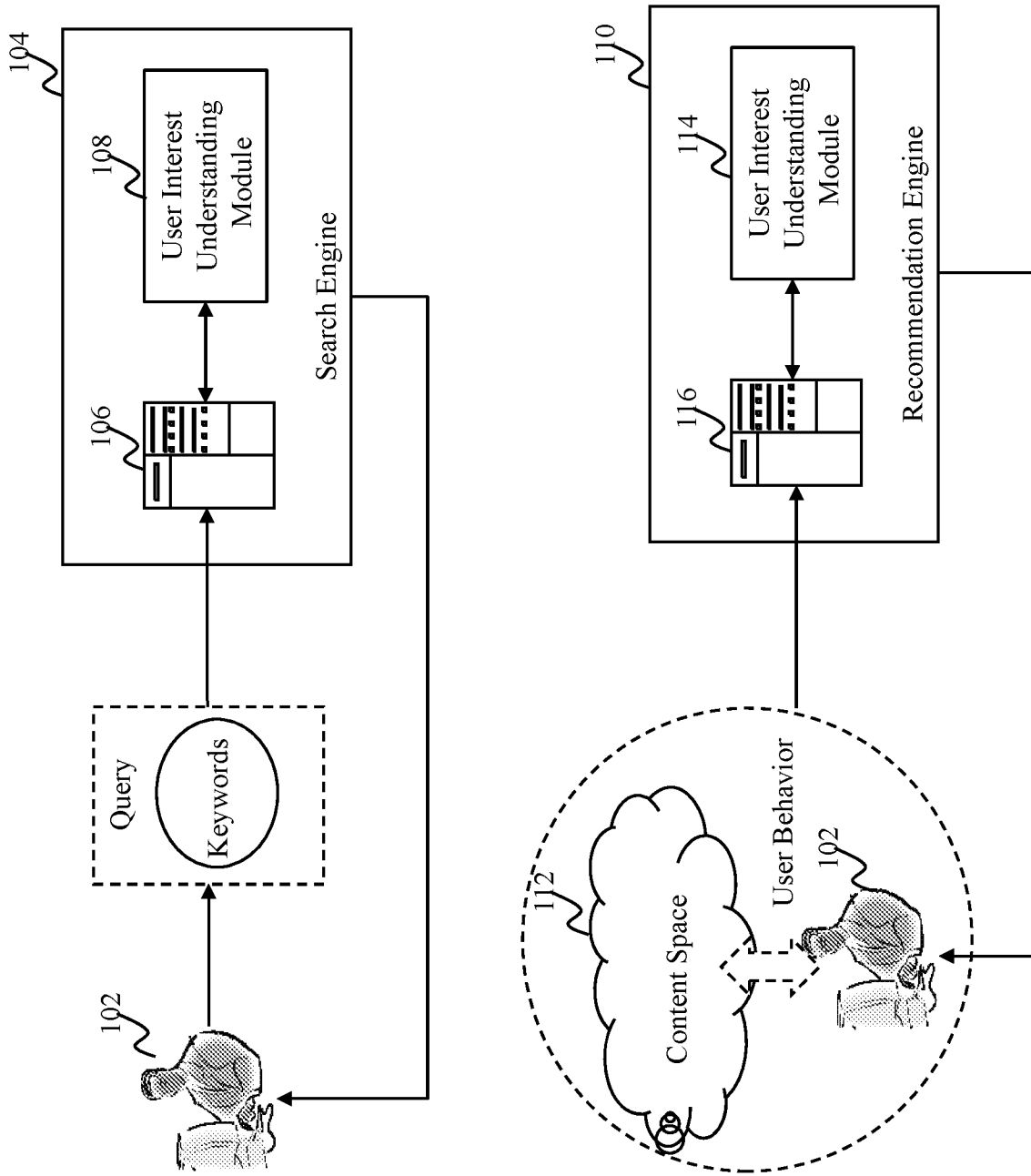


FIG. 1 (PRIOR ART)

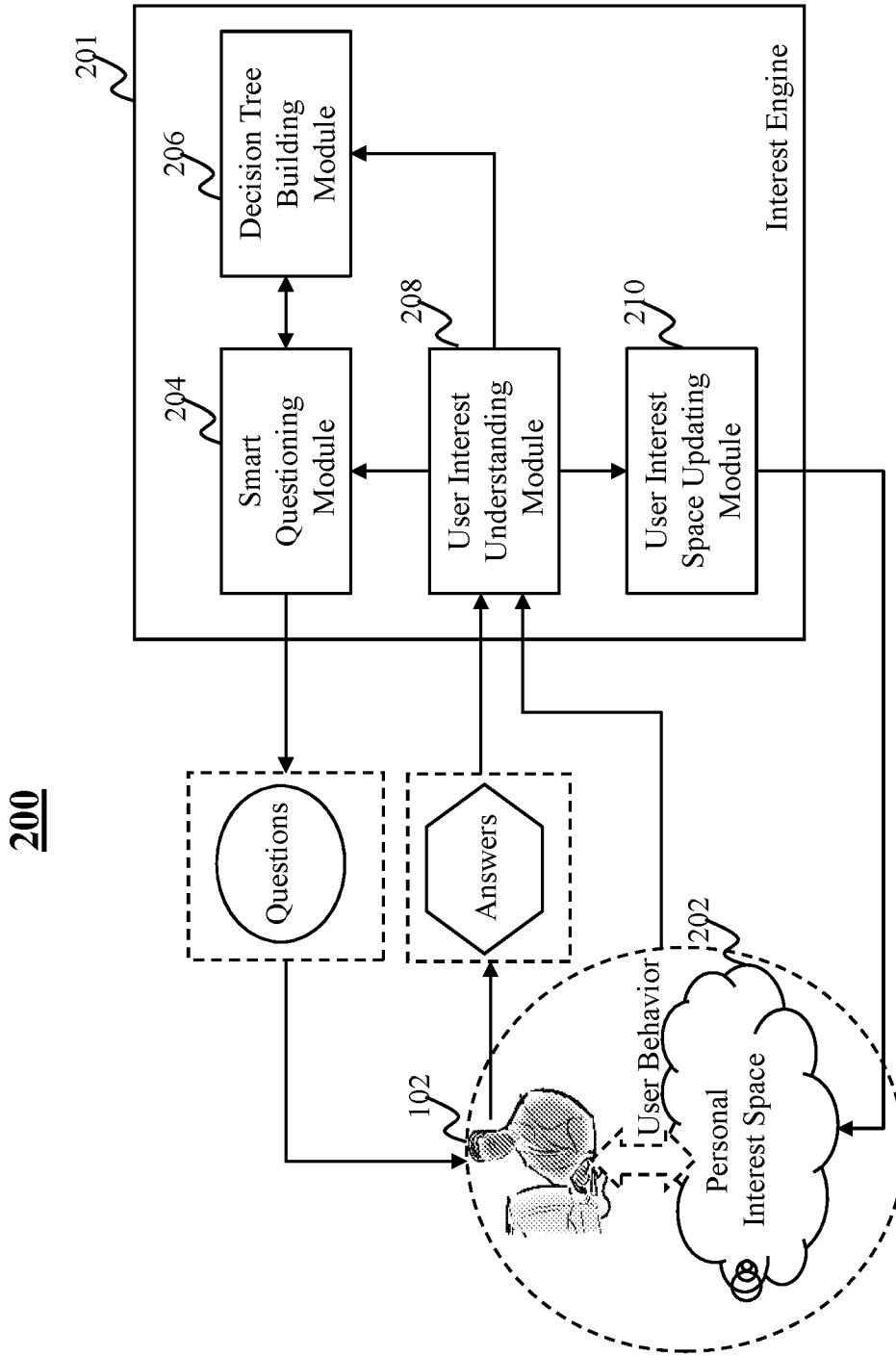


FIG. 2

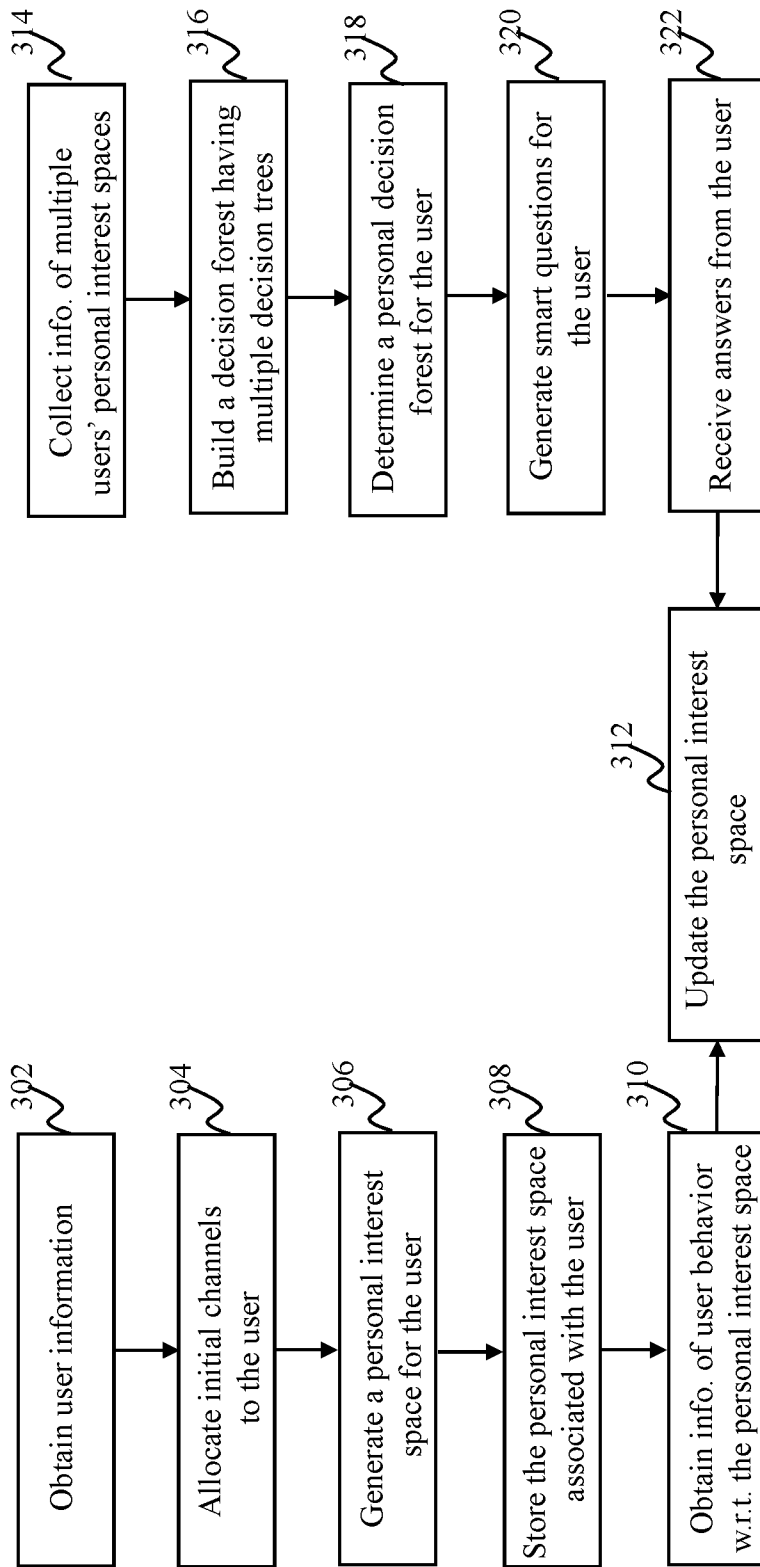


FIG. 3

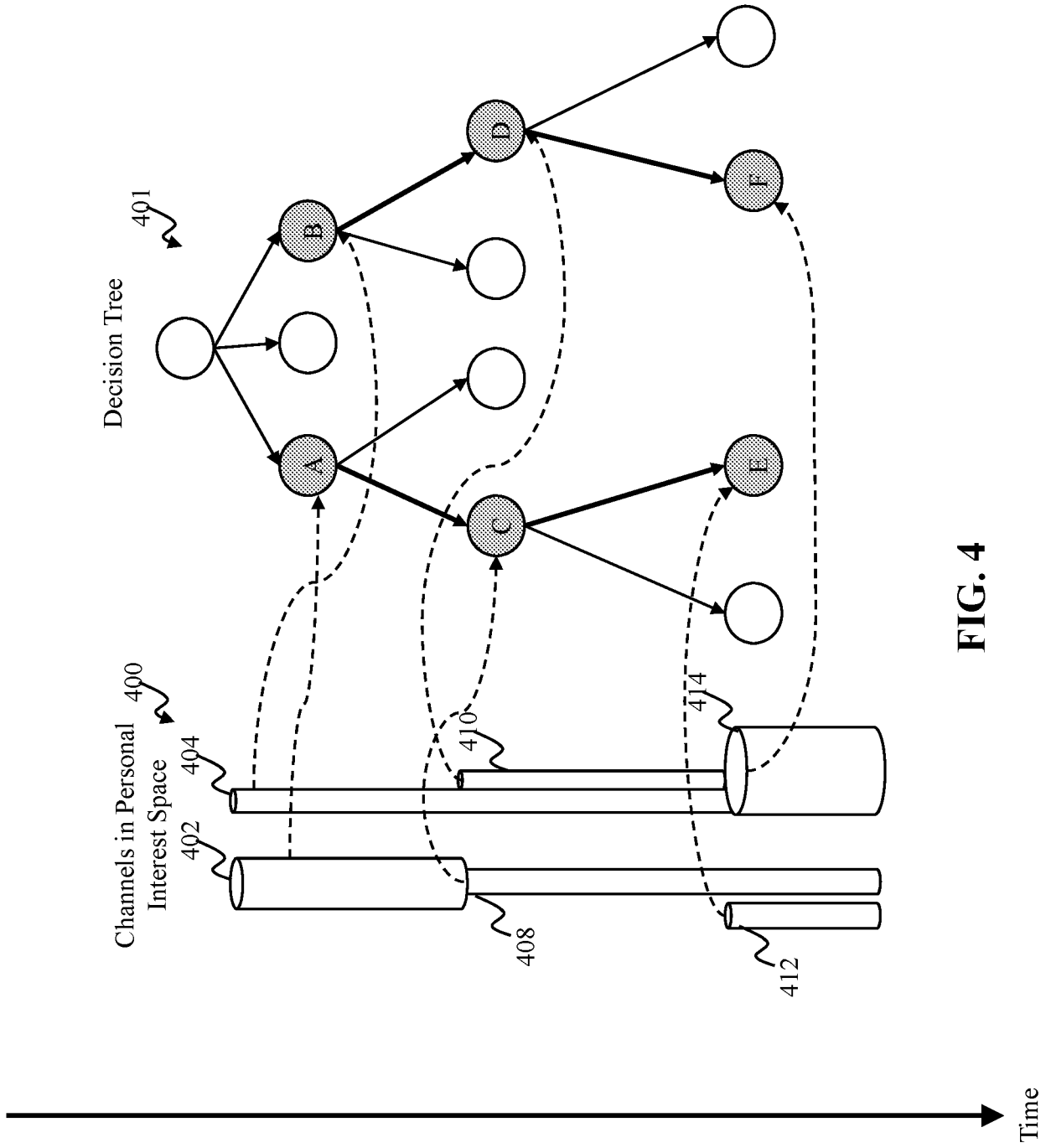


FIG. 4

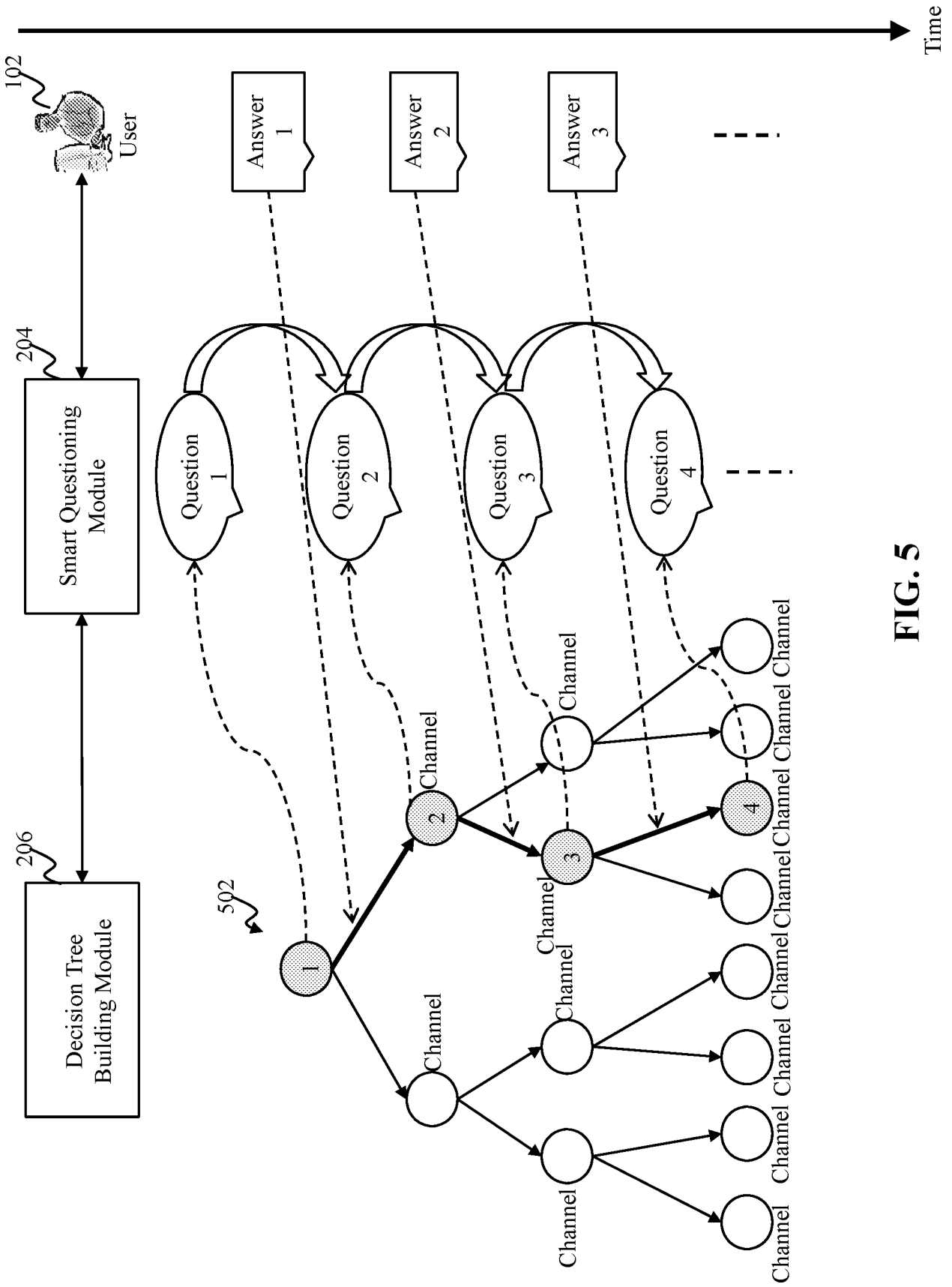


FIG. 5

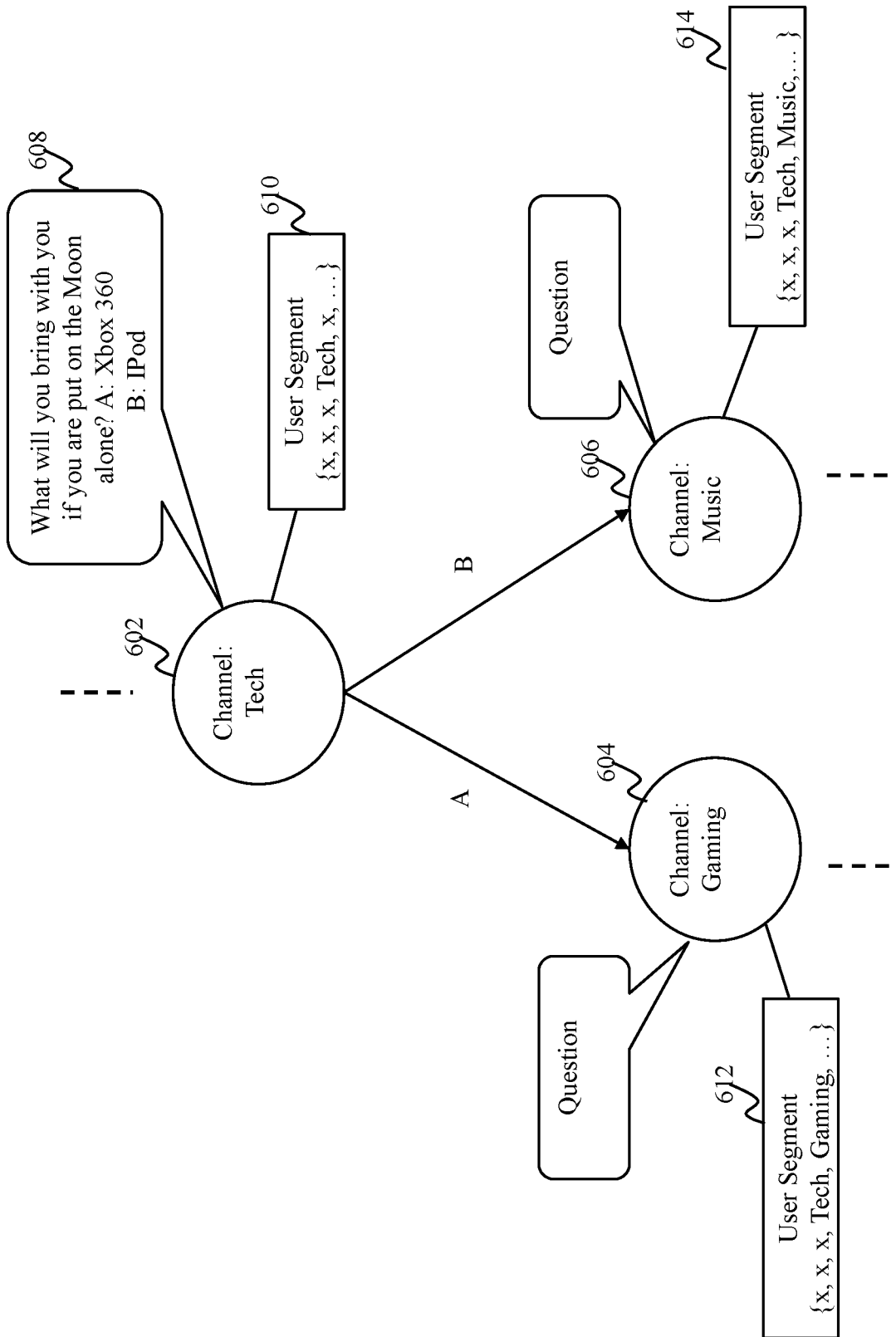


FIG. 6

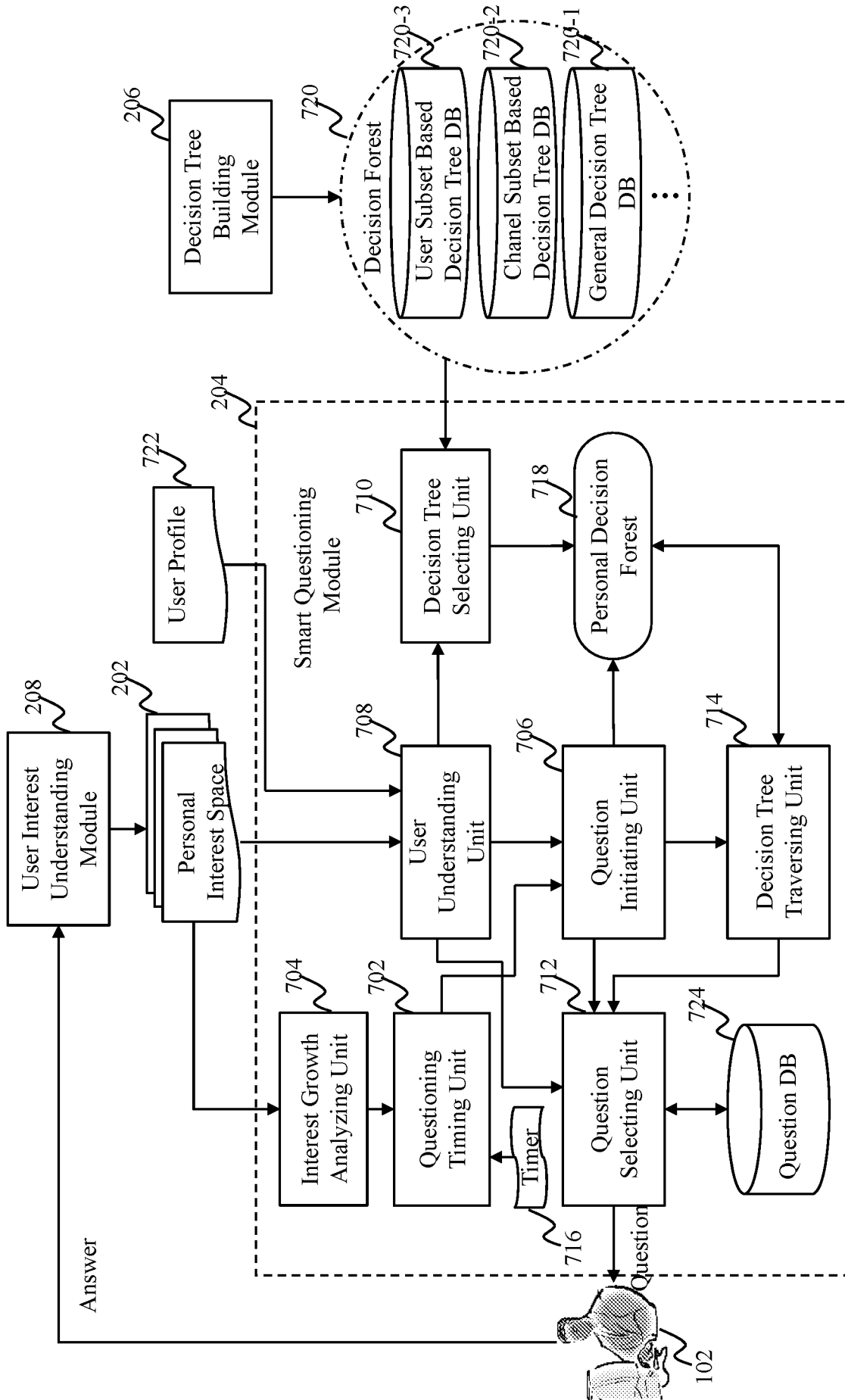


FIG. 7

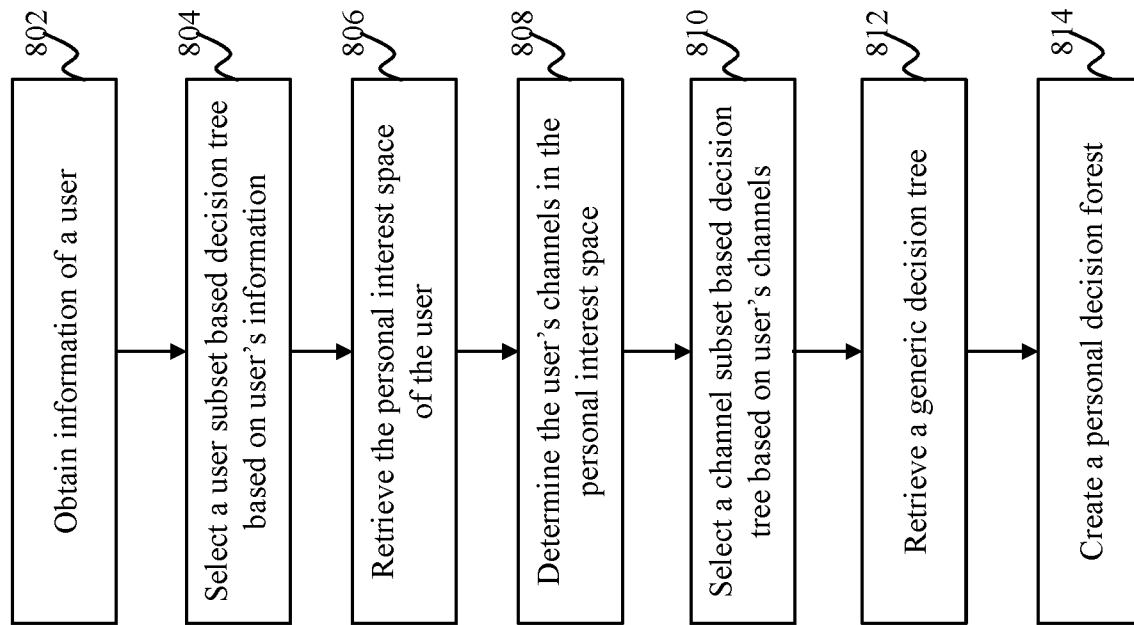


FIG. 8

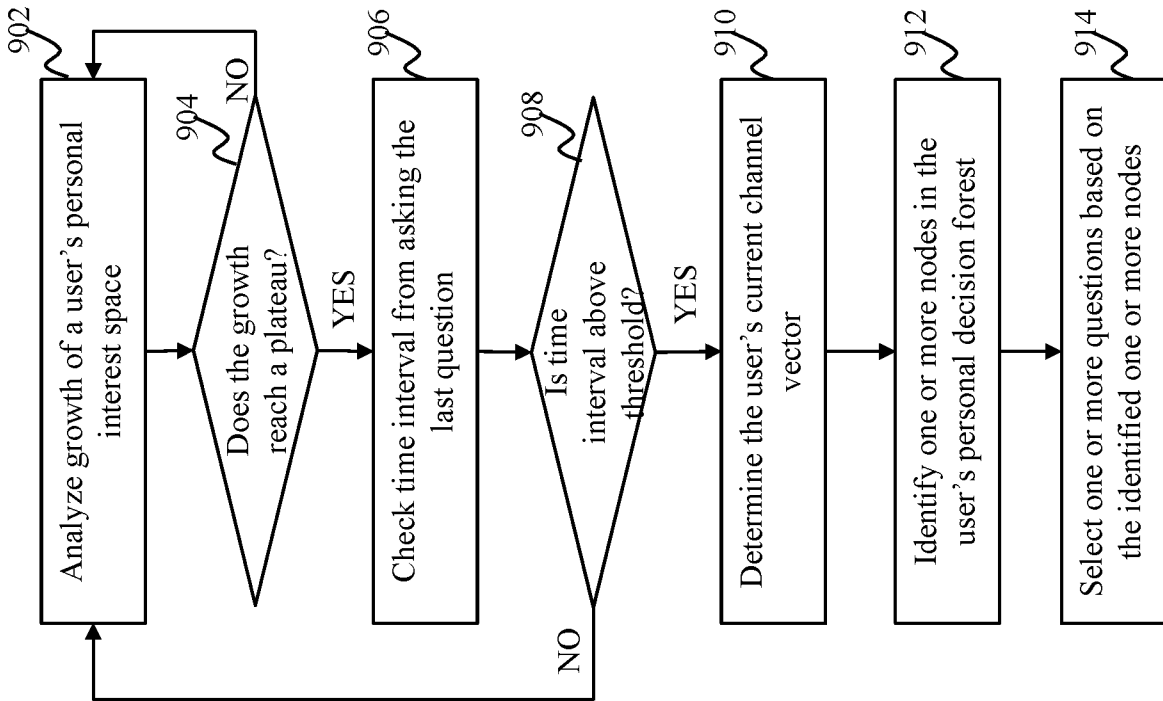


FIG. 9

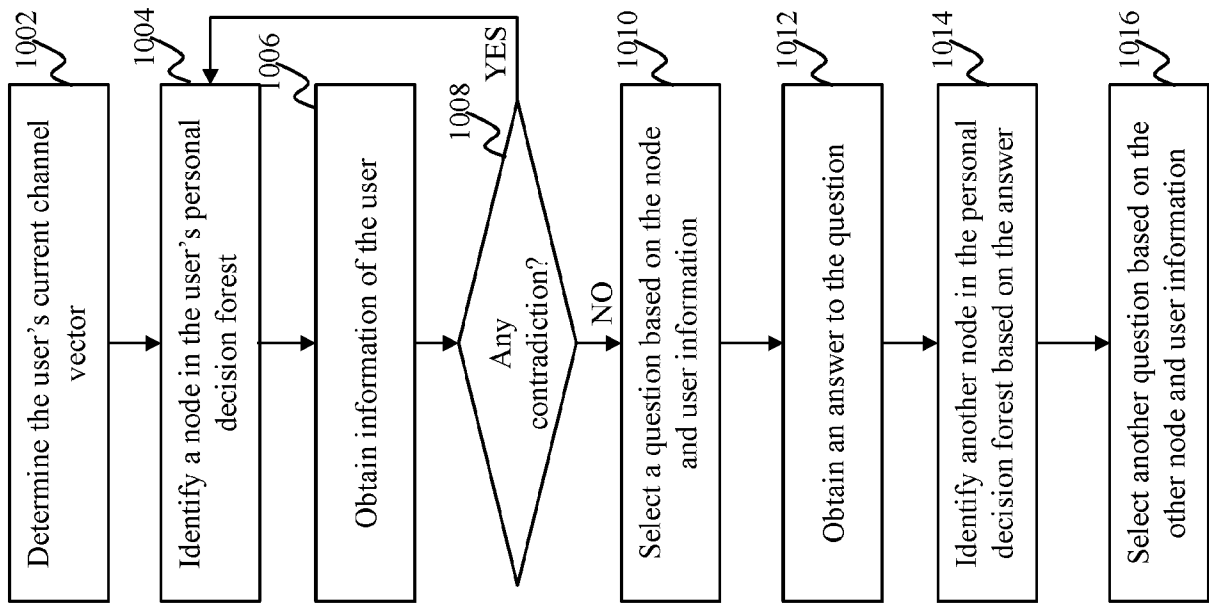


FIG. 10

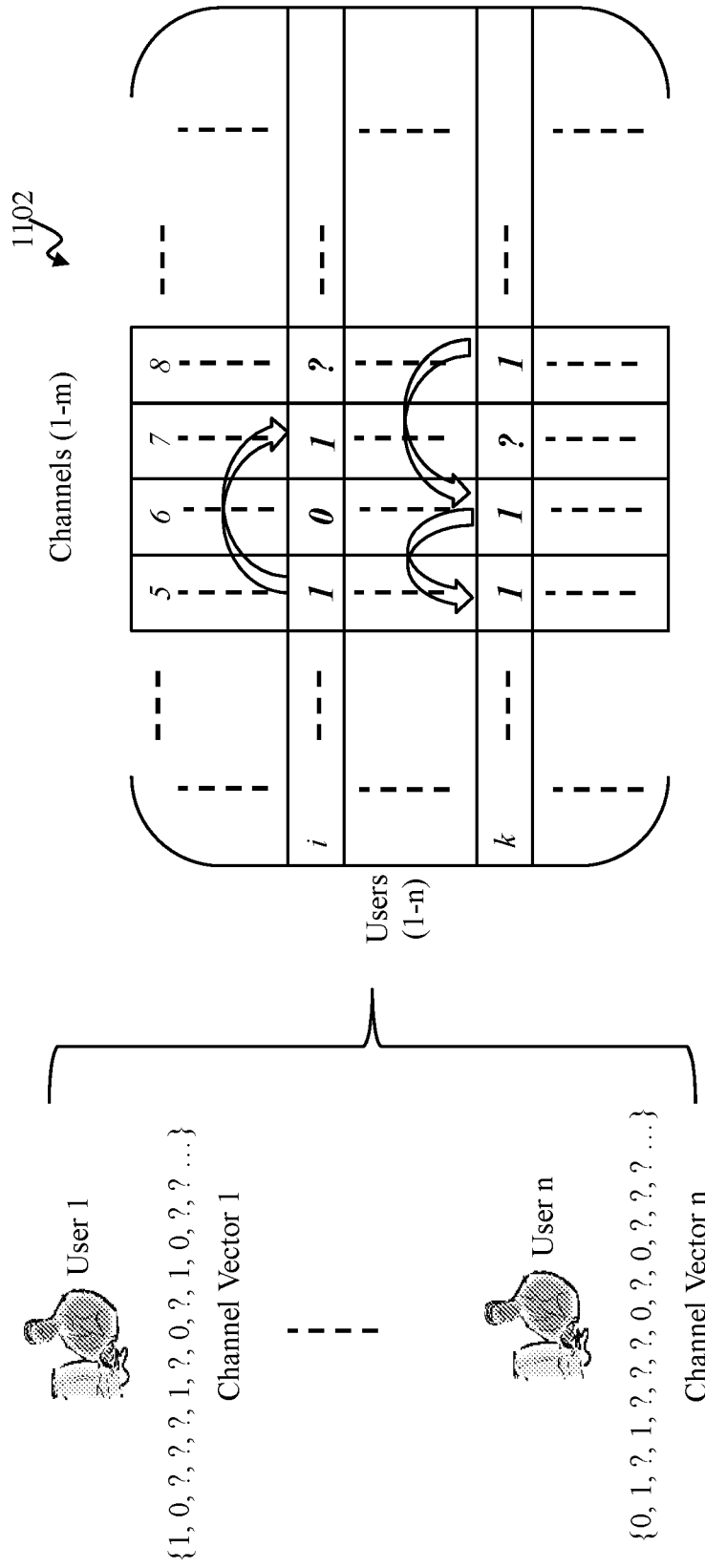


FIG. 11

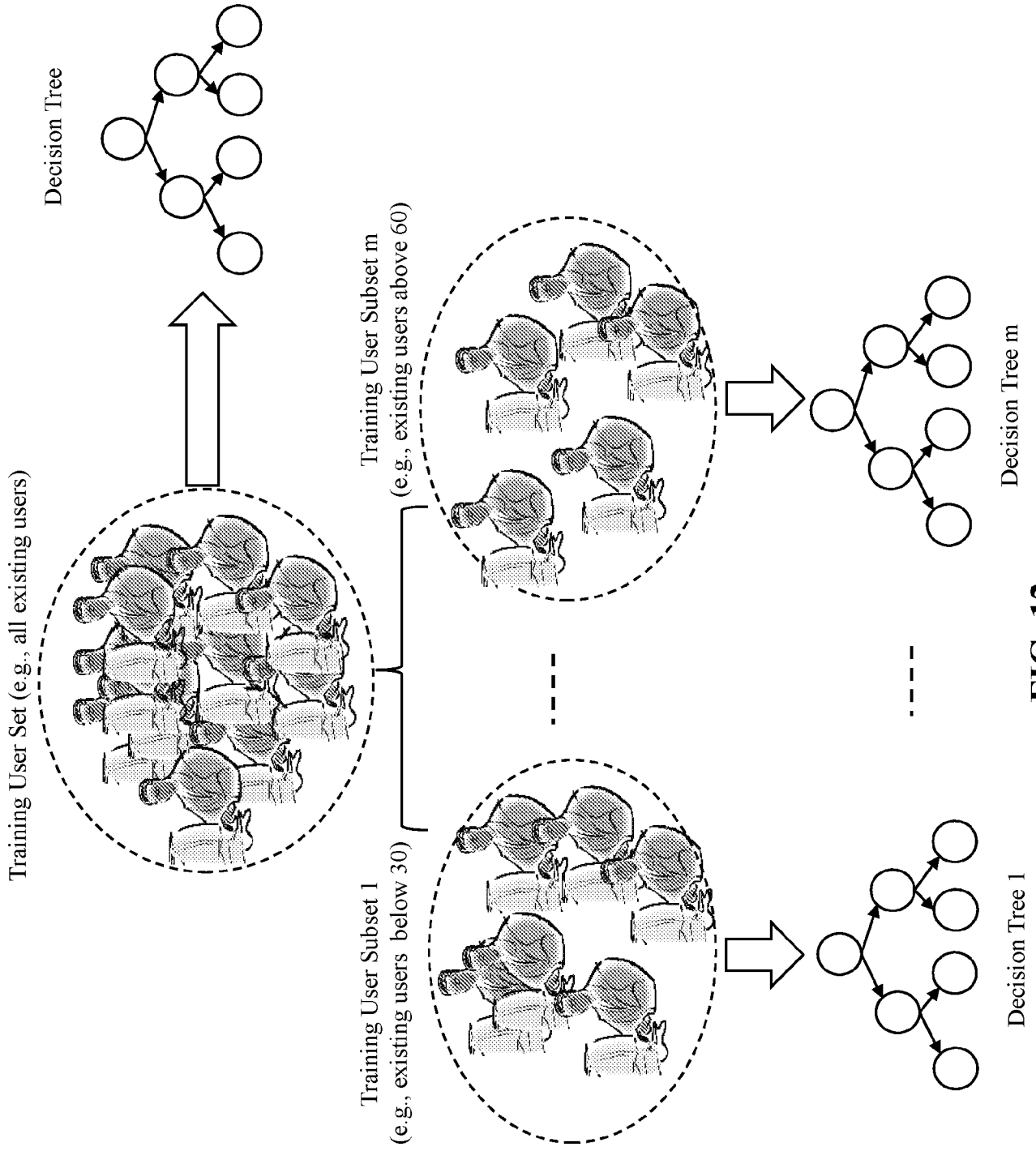


FIG. 12

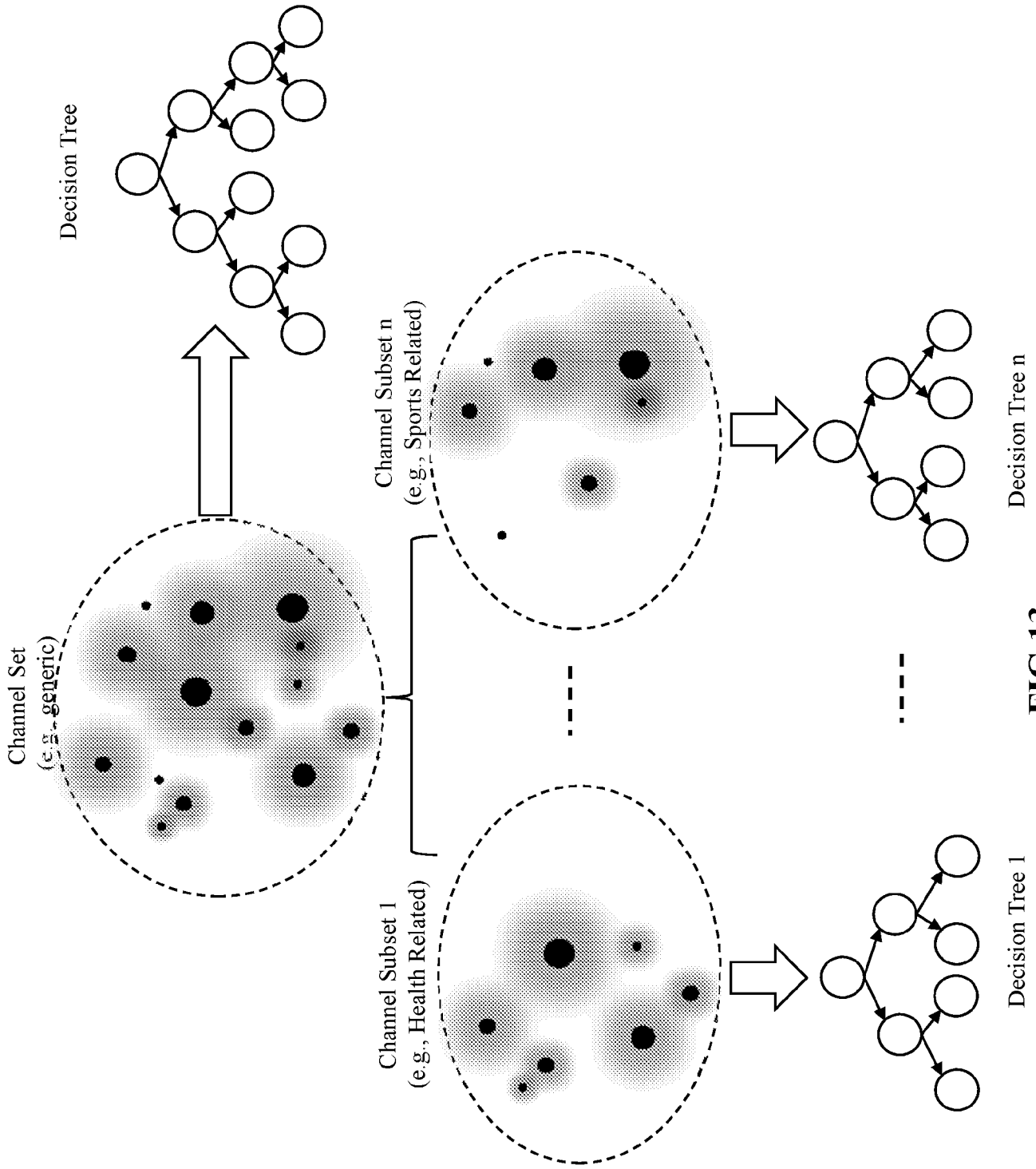


FIG.13

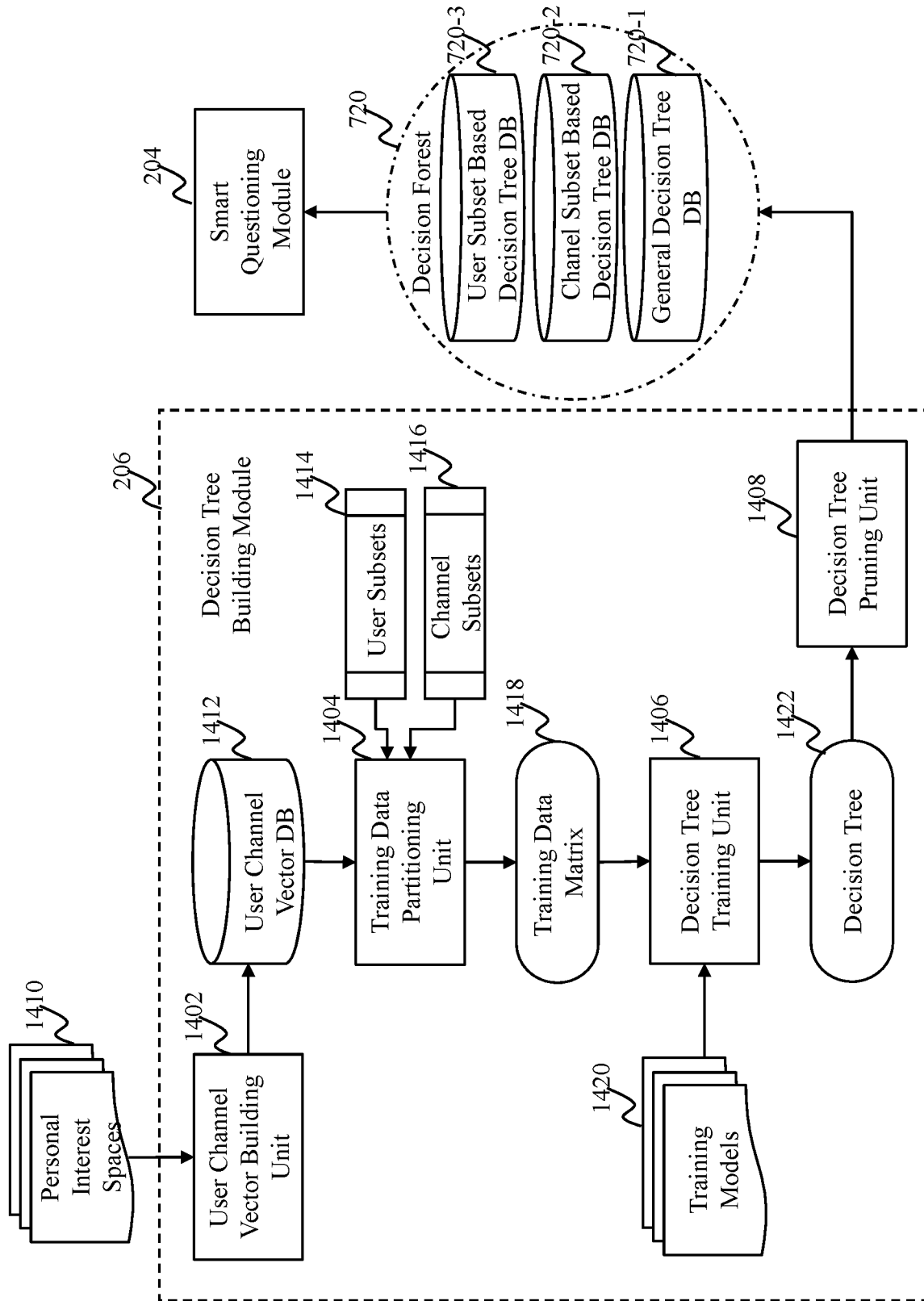


FIG. 14

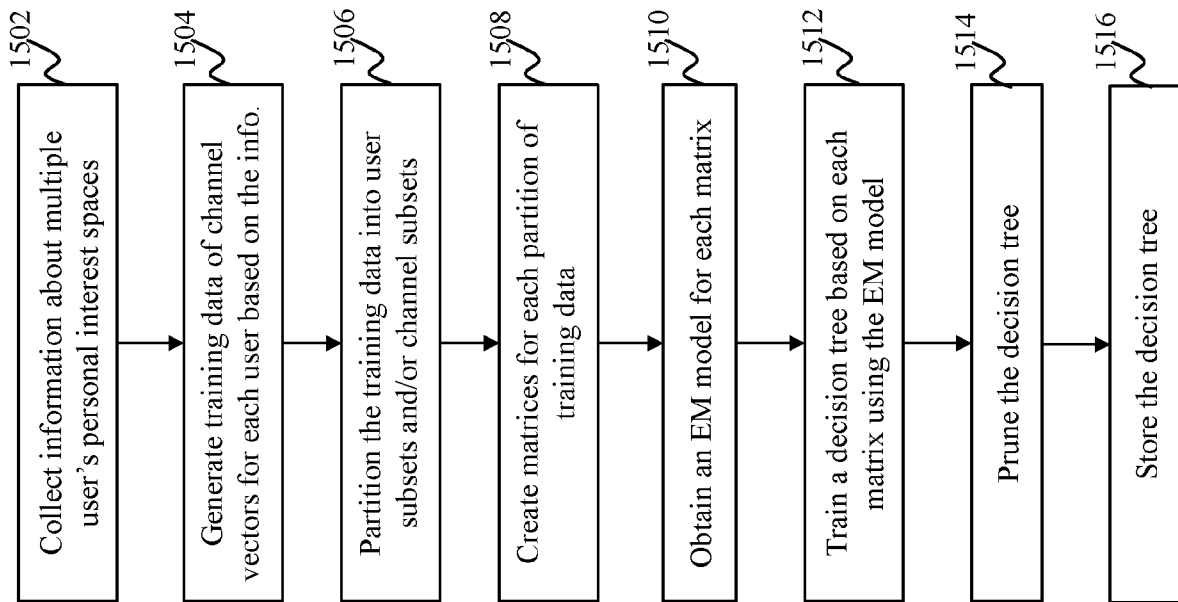


FIG. 15

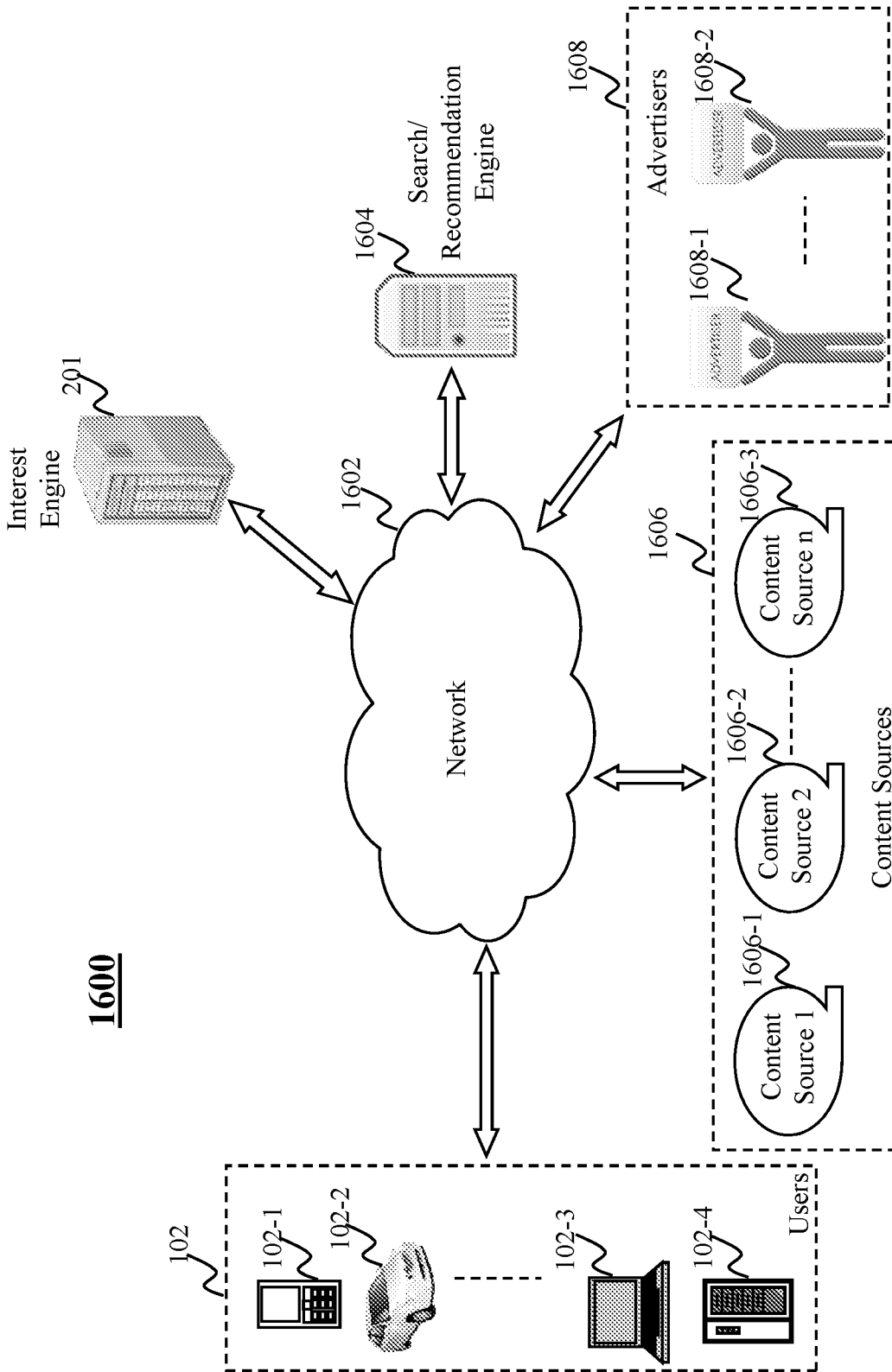
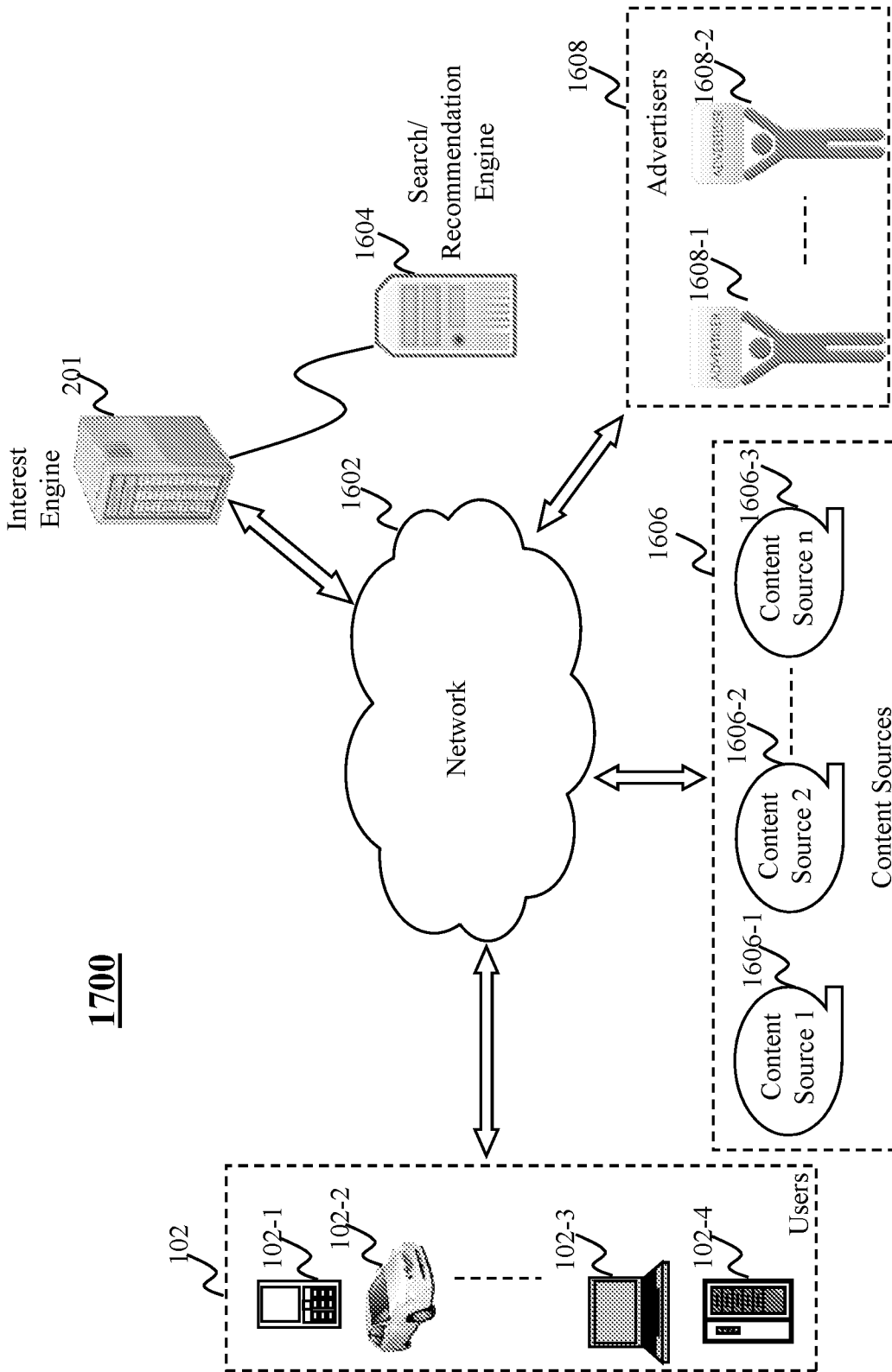
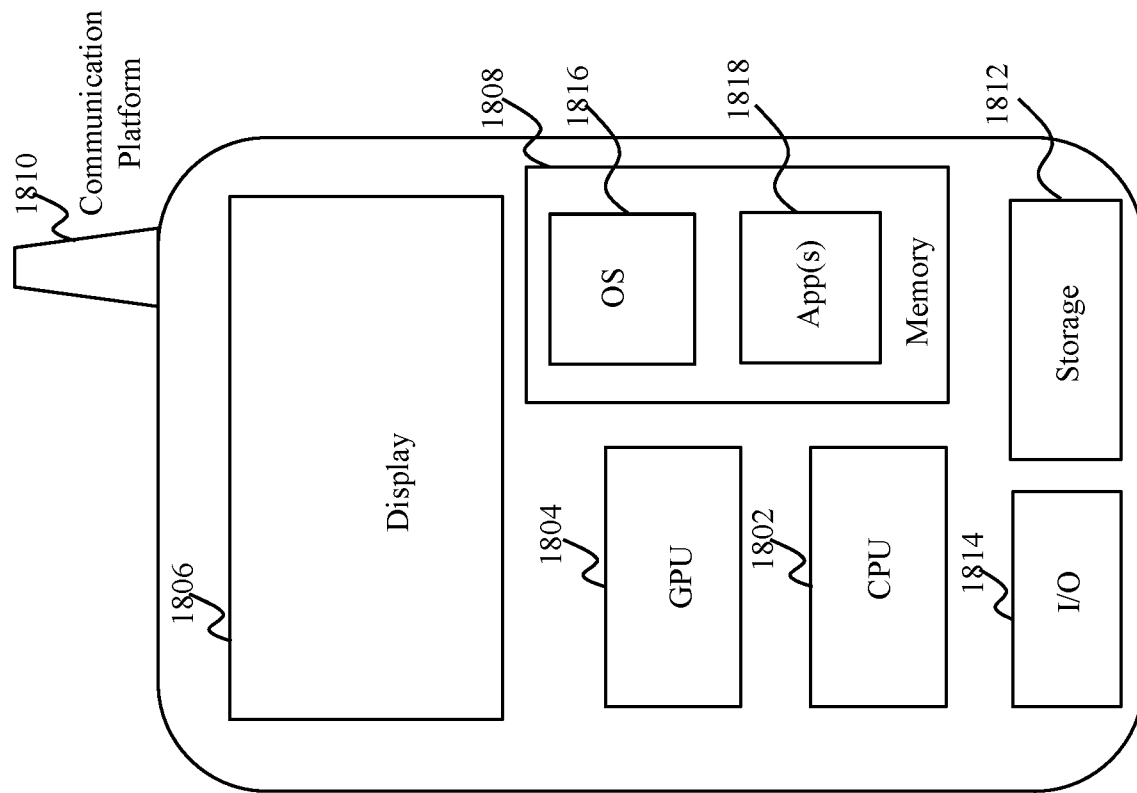


FIG. 16



1700

FIG. 17



1800

FIG. 18

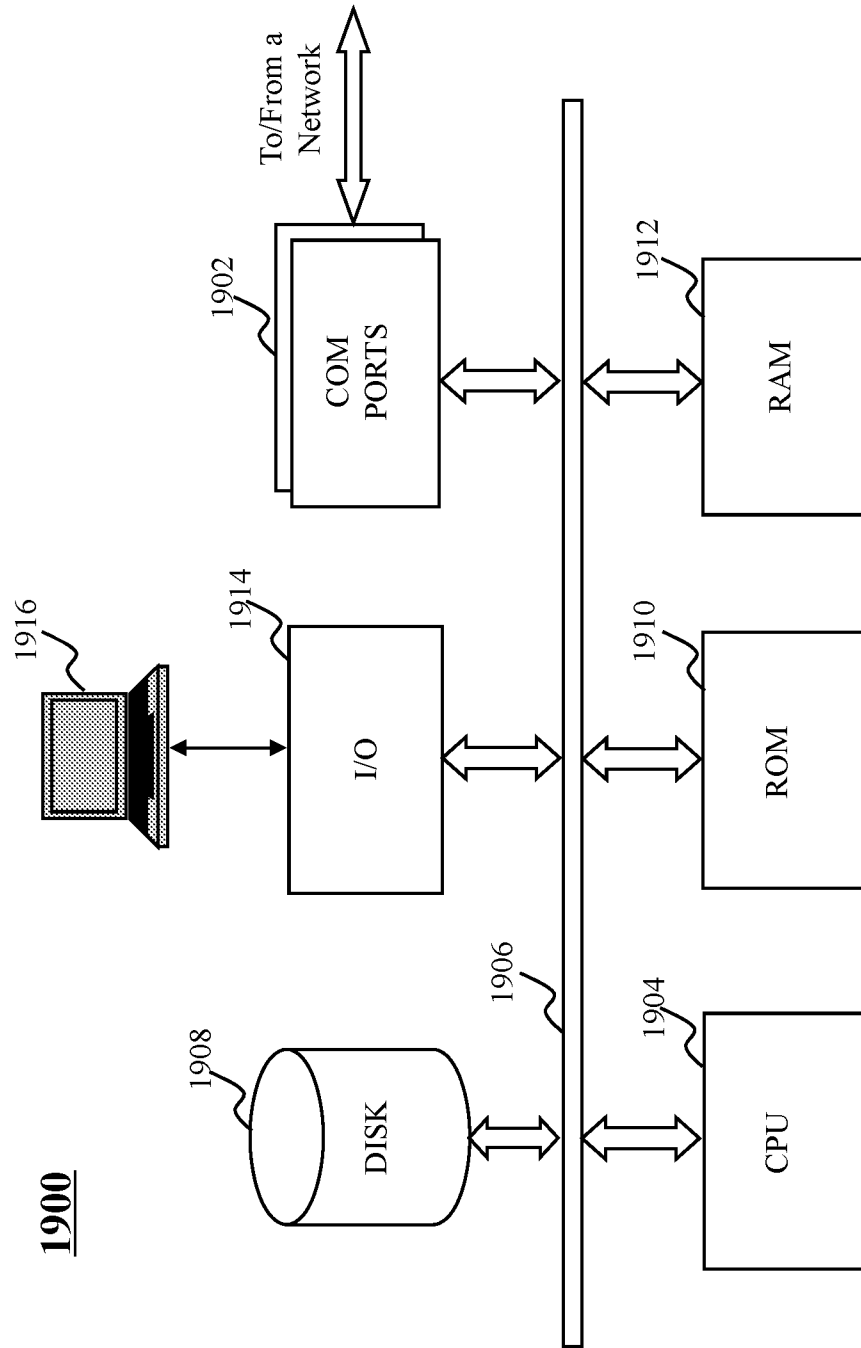


FIG. 19

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2015/098147

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F; G06Q; H04L; H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, EPODOC, WPI, CNKI, IEEE: interest, tree, node, question, query, ask, answer, personal, characteristic, channel, class, content, network, update, decision tree

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 102760128 A (EAST CHINA NORMAL UNIVERSITY) 31 October 2012 (2012-10-31) description, paragraphs [0031]-[0054], [0062]-[0096], figures 1-4, and 7	1-25
A	CN 104254851 A (BEIJING HAIPU WANGJU TECHNOLOGY LTD.) 31 December 2014 (2014-12-31) the whole document	1-25
A	CN 104395901 A (BEIJING YIDIAN WANGJU TECHNOLOGY CO., LTD.) 04 March 2015 (2015-03-04) the whole document	1-25
A	CN 103235823 A (SHANGHAI HEGUANG INFORMATION TECHNOLOGY CO., LTD.) 07 August 2013 (2013-08-07) the whole document	1-25
A	US 2014325567 A1 (MICROSOFT CORPORATION) 30 October 2014 (2014-10-30) the whole document	1-25

 Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

31 August 2016

Date of mailing of the international search report

21 September 2016

Name and mailing address of the ISA/CN

STATE INTELLECTUAL PROPERTY OFFICE OF THE
P.R.CHINA
6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing
100088, China

Authorized officer

WANG,Jian

Facsimile No. (86-10)62019451

Telephone No. (86-10)61648279

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2015/098147

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	102760128	A	31 October 2012	None			
CN	104254851	A	31 December 2014	WO	2013138969	A1	26 September 2013
				US	2015112918	A1	23 April 2015
CN	104395901	A	04 March 2015	US	2015206183	A1	23 July 2015
				HK	1207704	A1	05 February 2016
				WO	2014043849	A1	27 March 2014
CN	103235823	A	07 August 2013	None			
US	2014325567	A1	30 October 2014	CN	105210373	A	30 December 2015
				EP	2989802	A1	02 March 2016
				KR	20160003801	A	11 January 2016
				WO	2014176383	A1	30 October 2014