



- (51) International Patent Classification:
G06F 17/30 (2006.01)
- (21) International Application Number:
PCT/CN2015/082512
- (22) International Filing Date:
26 June 2015 (26.06.2015)
- (25) Filing Language: English
- (26) Publication Language: English
- (71) Applicant: **HEWLETT-PACKARD DEVELOPMENT COMPANY, L.P.** [US/US]; 11445 Compaq Center Drive West, Houston, Texas 77070 (US).
- (72) Inventors; and
- (71) Applicants (*for US only*): **YU, Xiao-Feng** [CN/CN]; Building 20, Universal Business Park, 10 Jiu XianQiao Road, Chaoyang District, Beijing 100015 (CN). **XIE, Jun-Qing** [CN/CN]; Building 20, Universal Business Park, 10 Jiu XianQiao Road, Chaoyang District, Beijing 100015 (CN).
- (74) Agent: **DEQI INTELLECTUAL PROPERTY LAW CORPORATION**; 7/F, Xueyuan International Tower, No. 1 Zhichun Road, Haidian District, Beijing 100083 (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).

[Continued on next page]

(54) Title: LATENT TOPIC DETERMINATION

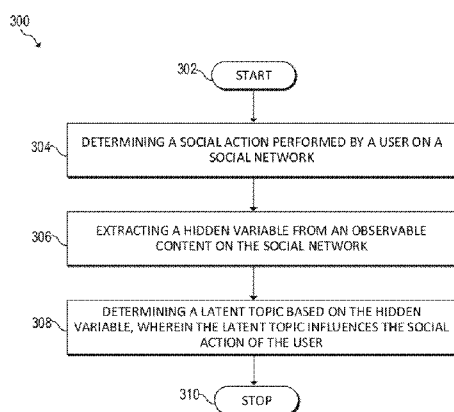


FIG. 3

(57) Abstract: In one example in accordance with the present disclosure, a method for latent topic determination includes determining a social action performed by a user in a social network. The method includes extracting a hidden variable from an observable content on the social network. The hidden variable corresponds to a user characteristic. The method also includes determining a latent topic based on the hidden variable, wherein the latent topic influences the social action of the user.

Published:

— *with international search report (Art. 21(3))*

LATENT TOPIC DETERMINATION

BACKGROUND

[0001] Social networks provide an environment for users to perform a variety of social actions such as generating content, communicating with other users, reviewing products, publishing papers, etc. The exponential growth of social media and online social networking services worldwide has generated increased importance in predicting and understanding users' behaviors and activities in socially connected networks. Social action prediction can provide insights into the behaviors and experiences for users as well as market demand analysis and product improvement.

BRIEF DESCRIPTION OF THE DRAWINGS

[0002] The following detailed description references the drawings, wherein:

[0003] FIG. 1 is a block diagram of an example computing environment in which latent topic determination may be useful;

[0004] FIG. 2 is a graphical representation of an example latent topic graph;

[0005] FIG. 3 is a flowchart of an example method for latent topic determination;

[0006] FIG. 4 is a flowchart of an example method for latent topic determination;

[0007] FIG. 5 is a flowchart of an example method for latent topic determination;

[0008] FIG. 6 is a block diagram of an example system for latent topic determination; and

[0009] FIG. 7 is a block diagram of an example system for latent topic determination.

DETAILED DESCRIPTION

[0010] Social actions performed on social networks may be greatly influenced by user characteristics defining user preferences and user interests. Accordingly, modeling users' topics of interest may play an important role in predicting social behaviors. As used herein, a *topic* may be defined as a set of fine-grained expressions that essentially captures a user's preferences or interests. For example, in an academic social network, the topic {*AI, intelligence, artificial, machine*} intrinsically illustrates a user's interest of AI related areas. A user interested in AI may be more likely to publish AI related papers than to publish software engineering papers. As another example, in a consumer reviewing site, the topic {*Software, Hardware, PC*} intrinsically illustrates a user's interest of PC related areas. A user interested in PCs may be more likely to review and purchase computer related products.

[0011] As shown in the preceding examples, multiple topics can be utilized to explore a user's diverse range of interests. In addition, topics can be partially overlapped to exhibit related interests. These topics are usually implicit and/or hidden, meaning that such topics are unobservable and unavailable in the majority of currently popular and predominant online social networking sites and are thus referred to as *latent topics*. Because the latent topics are hidden, the latent topics are not labeled. Labeling a latent topic may include associating a keyword or other descriptor to identify the latent topic. Labeling, or annotating, latent topics manually requires considerable engineering and human labor effort, which is time-consuming and ineffective.

[0012] Example systems for latent topic determination address these challenges by utilizing a latent topic graph (LTG) model to automatically

learn intermediate interdependencies and dynamics between users' social action and observable content. Social actions may include any action taken on a social network, such as generating content, communicating with other users, reviewing products, publishing papers, etc. Social networks may include any platforms that allow a user to generate content viewable by other users, such as a communication network, an academic network, a product review network, etc. Observable content is content generated by a social action of the user, such as a paper, a post, a product review, a message, etc. Observable content may be any data (text, images, video, etc.) accessible on the social network. variables in terms of latent topics.

[0013] The intermediate interdependencies and dynamics between users' social action and observable content are expressed in terms of latent topics. The latent topics are represented in the LTG by a set of hidden variables. Through efficient training and inference algorithms, LTG captures flexible interactions between the set of hidden variables and users' social action variables to make these hidden variables interpretable for meaningful latent topic discovery. LTG can also incorporate a variety of factors and evidences for accurate social action prediction. In addition, the LTG model allows natural incorporation of unlabeled data for semi-supervised learning and scalable parallel and distributed computing of big social data.

[0014] An example method for determining latent topics may include determining, by a processor, a social action performed by a user in a social network. The method may also include extracting, by the processor, a hidden variable, corresponding to a user characteristic, from an observable content on the social network. The method may also include determining, by the processor, a latent topic based on the hidden variable, wherein the latent topic influences the social action of the user.

[0015] FIG. 1 is an example environment 100 in which various examples may be implemented as a latent topic determination system 110. Latent topic determination system 110 may comprise various components, including a social action engine 112, an observable content engine 114, a hidden variable engine 116, a latent topic engine 118, a social action engine 120, and/or

other components. Environment 100 may also include various components including a server computing device 130 and a client computing device 132. The client computing device 132 may communicate requests to and/or receive responses from the server computing device 130. The server computing device 130 may receive and/or respond to requests from the client computing device 132. The client computing device 132 may be any type of computing device providing a user interface through which a user can interact with a software application. For example, the client computing device 132 may include a laptop computing device, a desktop computing device, an all-in-one computing device, a tablet computing device, a mobile phone, and/or other electronic device suitable for displaying a user interface and processing user interactions with the displayed interface. While the server computing device 130 is depicted as a single computing device, the server computing device 130 may include any number of integrated or distributed computing devices serving at least one software application for consumption by the client computing device 132.

[0016] The various components (e.g., components 130 and/or 132) depicted in FIG. 1 may be coupled to at least one other component via a network 134. Network 134 may comprise any infrastructure or combination of infrastructures that enable electronic communication between the components. For example, the network 134 may include at least one of the Internet, an intranet, a PAN (Personal Area Network), a LAN (Local Area Network), a WAN (Wide Area Network), a SAN (Storage Area Network), a MAN (Metropolitan Area Network), a wireless network, a cellular communications network, a Public Switched Telephone Network, and/or other network. According to various implementations, latent topic determination system 110 and the various components described herein may be implemented in hardware and/or a combination of hardware and programming that configures hardware. Furthermore, in FIG. 1 and other Figures described herein, different numbers of components or entities than depicted may be used.

[0017] As described above, latent topic determination system 110 may comprise various components, including a social action engine 112, an observable content engine 114, a hidden variable engine 116, a latent topic

engine 118, a social action engine 120, and/or other components. The components of the latent topic determination system 110 as described herein, may refer to a hardware or a combination of hardware and instructions that performs a designated function. As is illustrated with respect to FIG. 7, the hardware of the various components of latent topic determination system 110, for example, may include one or both of a processor and a machine-readable storage medium, while the instructions are code stored on the machine-readable storage medium and executable by the processor to perform the designated function.

[0018] Social action engine 112 may determine a social action performed by a user in a social network. Example social actions may include publishing a paper, citing a paper, creating a review, creating a user-user relationship, etc. Social networks may include any platforms that allow a user to generate content viewable by other users, such as a communication network, an academic network, a product review network, etc.

[0019] Observable content engine 114 may identify an observable content on the social network. The observable content may be generated by the social action of the user, such as a paper, a post, a product review, a message, etc. Observable content may be any data (text, images, video, etc.) accessible on the social network.

[0020] Hidden variable engine 116 may extract a hidden variable from the observable content on the social network. The hidden variable may correspond to a user characteristic.

[0021] The user characteristics may include user preferences corresponding to a user of a social network, such as topics, products, etc. A topic may be defined as a set of fine-grained expressions that essentially captures a user's preferences or interests. Example topics may include topics for research papers, topics for online product reviews, etc. User characteristics may also include prior impressions from a first user of a social network on a second user of the social network. For example, a first user may have influence on a second user in creating a relationship, creating content, etc. In other examples, the user characteristics may include additional and/or alternate elements associated with a user. Hidden variable engine 116 may extract a key

word from an observable content on the social network, associate the key word with the latent topic and create a group of related latent topics. The hidden variable engine 116 may incorporate one or more models, formulas, equations and/or algorithms in extracting the hidden variable. For example, the hidden variable engine 116 may incorporate a latent topic graph (LTG) model.

[0022] FIG. 2 shows a graphical representation of a model 200 of a latent topic graph (LTG). The model 200 includes a social action portion 202, a user portion 204 and a hidden variable portion 206. Although the model 200 illustrates five social actions, five users and three hidden variables, this is for example only and any number of social actions, users and/or hidden variables may be represented in the LTG model. The social action portion 202 captures dependencies $P(y|h)$ between hidden variables h and social action variables y by feature functions $g(\cdot)$. The user portion 204 captures dependencies $P(h|s)$ between observable social data s and hidden variables h by feature functions $f(\cdot)$.

[0023] In an attribute augmented social network $s = (\mathcal{U}, \mathcal{E}, \mathcal{R})$ representing $|\mathcal{U}| = N$ users $u_i \in \mathcal{U}$, $1 \leq i \leq N$, $|\mathcal{E}| = M$ dyads or relationships between users $e_{ij} \in \mathcal{E}$, $\mathcal{E} \subset \mathcal{U} \times \mathcal{U}$ and $|\mathcal{R}| = L$ and attributes $r_l \in \mathcal{R}$, $1 \leq l \leq L$ for all users and dyads, $y = \{y_i\}_{i=1}^N$, $y_i \in \mathcal{Y}^N$ may represent N users' behaviors and activities, and $h = \{h_j\}_{j=1}^n$, $h_i \in \mathcal{H}^n$ may represent n latent topics of users' interests. Hidden variable engine 116 may optimize and find the most likely types of social actions y^* that have the maximum a posteriori (MAP) probability $P(y|s)$ given the social network s . Hidden variable engine 116 may also determine the most likely latent topics h^* such that $P(h|s)$ is optimized.

[0024] The model 200 uses a set of hidden variables $h = \{h_j\}_{j=1}^n$, $h_i \in \mathcal{H}^n$ as n latent topics of users' interests, since h are not observable in social data s , where $\mathcal{H}^n$ is the finite set of all possible hidden variable values. In one example, the number of latent topics n may vary from 10 to 50 with an incremental step of 10, although this is only a sample value and a variety of values and increments may be used for the number of latent topics. A hidden dynamic probabilistic model may be expressed as:

[0025] Formula (1):

$$P(y|s) = \sum_h P(y|h)P(h|s)$$

[0026] In formula (1), $P(h|s)$ models probabilistic dependencies between observable social data s and hidden variables h , while $P(y|h)$ models probabilistic dependencies between hidden variables h and social action variables y . For the concrete modeling, distributions may be specified for these dependencies. The LTG may be defined based on exponential-family random graph models (ERGMs). If $\{\phi_c\}$ and $\{\psi_d\}$ are two sets of factors in s , the dependencies can be factorized as a product of factors over all cliques as:

Formula (2):

$$P(y|h) \sim \prod_c \phi_c(y_c, h_c), P(h|s) \sim \prod_d \psi_d(h_d, s_d)$$

[0027] Each factor is instantiated as the exponential family of an inner product over sufficient statistics (feature functions) and corresponding parameters. The dependencies can be expressed as:

[0028] Formula (3):

$$P(y|h) = \frac{1}{Z(h)} \exp\left\{\sum_{k=1}^l \mu_k g_k(y, h)\right\}$$

[0029] Formula (4):

$$P(h|s) = \frac{1}{Z(s)} \exp\left\{\sum_{k=1}^m \lambda_k f_k(h, s)\right\}$$

[0030] In formulas (3) and (4), $g_k(y, h)$ and $f_k(h, s)$ are feature functions for user behaviors, hidden variables, and observable social data. These feature functions capture the hidden-dynamic dependencies in LTG. y is a member of $\mathcal{Y}^N$ and h is a member of $\mathcal{H}^n$, respectively. The vector of model parameters is $\theta = \{\{\lambda_{k=1}^m\}, \{\mu_{k=1}^l\}\}$. $Z(h)$ and $Z(s)$ are corresponding normalization functions defined as:

[0031] Formula (5):

$$Z(h) = \sum_y \exp \left\{ \sum_{k=1}^l \mu_k g_k(y, h) \right\}$$

[0032] Formula (6):

$$Z(s) = \sum_h \exp \left\{ \sum_{k=1}^m \lambda_k f_k(h, s) \right\}$$

[0033] The model 200 captures intermediate interdependencies and dynamics between observable data and users' social action variables in terms of hidden variables for latent topics of users' preferences. A benefit of LTG is its ability to represent rich and dynamic interdependencies and interactions between observable social data and hidden variables, and between hidden variables and social actions.

[0034] Although exploiting an intermediate layer of hidden variables could model topics of users' interests, it also makes the likelihood function non-convex. Noticeably, to avoid bad locally optimal solutions and yield flexible and interpretable interactions between hidden variables and desired social actions, an entropy based regularization term may be used in the LTG model. Since ℓ_1 -regularization typically leads the objective function to be nondifferentiable, ℓ_2 -regularization may be more suitable for probabilistic models such as LTG. In view of such considerations, the objective function in LTG may be defined to optimize the model parameters θ as:

[0035] Formula (7):

$$\begin{aligned} \mathcal{O}(\theta) &= \mathbb{E}[P(y|s)] - \frac{\|\theta\|^2}{2\sigma^2} - \eta \mathbb{E}_{P(y|h)}[P(y|h)], \\ &= \sum_{(y,s)} \tilde{P}(y,s) \log P_\theta(y|s) - \frac{\|\theta\|^2}{2\sigma^2} - \eta \mathbb{E}_{P(y|h)}[P(y|h)] \end{aligned}$$

[0036] In formula (7), the first term $\mathbb{E}[P(y,s)]$ is the expectation of $P(y,s)$ given the data distribution in the input network. The second term is a ℓ_2 -

regularization term to avoid over-fitting. $\tilde{P}(y,s)$ and $P_\theta(y|s)$ represent empirical distribution and conditional distribution, respectively. The third term $\mathbb{E}_{P(y|h)}[P(y|h)]$ aims to minimize the conditional entropy between hidden variables and social action labels, and η is a balance parameter to weight $\mathbb{E}_{P(y|h)}[P(y|h)]$. The balance parameter may be set to a variety of values, such as $\{0, 0.001, 0.005, 0.01, 0.05, 0.1\}$. The conditional entropy is the conditional probability distribution of social action labels given hidden variables. More importantly, the dynamics and flexibility are ensured between hidden variables and social action labels in order to make the hidden variables explainable to discover meaningful latent topics of users' interests. Conditional entropy may be defined as:

[0037] Formula (8):

$$\mathbb{E}_{P(y|h)}[P(y|h)] = - \sum_{\dot{y}} \sum_h P_\theta(\dot{y}, h) \log P_\theta(\dot{y}, h)$$

[0038] The model parameters θ can be efficiently learned by maximizing the objective function. This can be expressed as:

[0039] Formula (9): $\theta^* = \arg \max_{\theta} O(\theta)$

[0040] In the training set consisting of N users' behavior labels $y = \{y_i\}_{i=1}^N, y_i \in \mathcal{Y}^N$, of social graph s , s_i *simplifies the notation and* denotes all information related to user u_i , including all other users and dyads connected to u_i in s . In other words, s_i is a subgraph of s characterized by user u_i . Since the LTG model contains a hidden layer, the objective function $O(\theta)$ is no longer convex, and a closed form solution may not exist because optimization procedures typically only find local optima. Taking the partial derivatives of the objective function with respect to the parameters λ_k and μ_k yields:

[0041] Formula (10):

$$\begin{aligned} \frac{\partial O(\theta)}{\partial \lambda_k} = & \sum_{i=1..N} \tilde{P}(y_i, s_i) \sum_{h'} P_\theta(h'|y_i, s_i) f_k(h', s_i) \\ & - \sum_{i=1..N} \tilde{P}(s_i) \sum_{y', h'} P_\theta(y', h'|s_i) f_k(h', s_i) - \frac{\lambda_k}{\sigma^2} \end{aligned}$$

[0042] Formula (11):

$$\begin{aligned} \frac{\partial \mathcal{O}(\theta)}{\partial \mu_k} = & \sum_{i=1..N} \tilde{P}(y_i, s_i) \sum_{h'} P_\theta(h'|y_i, s_i) g_k(y_i, h') \\ & - \sum_{i=1..N} \tilde{P}(s_i) \sum_{y', h'} P_\theta(y', h'|s_i) g_k(y_i, h') - \frac{\mu_k}{\sigma^2} \\ & + \eta P(\dot{y}|\dot{h})(\mu_k - \sum_{y''} P(y''|\dot{h}) \mu_{k(y'', \dot{h})}) \end{aligned}$$

[0043] In the above formulas, both $\tilde{P}(y_i, s_i)$ and $\tilde{P}(s_i)$ are empirical distributions given the training data. y' is a member of $\mathcal{Y}^N$ and h' is a member of $\mathcal{H}^n$, and similarly for y'' and $\dot{h}$. In formula (11), $k(y'', \dot{h})$ equals to the unique $\hat{k}$ given $\dot{y}$ and $\dot{h}$. Both formula (10) and formula (11) require computing two different kinds of marginal probabilities: $P_\theta(h'|y_i, s_i)$ and $P_\theta(y', h'|s_i)$.

[0044] In order to compute these two probabilities, the factor theory of ERGMs and the graphical structure of LTGs can be used (e.g., repeated structure and parameter tying). Note that both formula (10) and formula (11) contain expectations of the fully-observed gradient, where the expectations are taken over the hidden variables h' . If $\{\phi_w\}$ is a set of factors, $P_\theta(h'|y_i, s_i)$ is exactly a marginal distribution of the clamped LTG defined as:

$$\begin{aligned} [0045] \quad \text{Formula (12):} \quad P(h|y, s) &= \frac{\prod_w \Phi_w(y_w, h_w, s_w)}{\sum_h \prod_w \Phi_w(y_w, h_w, s_w)} \\ &= \frac{\prod_c \phi_c(y_c, h_c) \prod_d \psi_d(h_d, s_d)}{\sum_h \prod_c \phi_c(y_c, h_c) \prod_d \psi_d(h_d, s_d)} \\ &= \frac{\exp\{\sum_{k=1}^l \mu_k g_k(y, h) + \sum_{k=1}^m \lambda_k f_k(h, s)\}}{\sum_h \exp\{\sum_{k=1}^l \mu_k g_k(y, h) + \sum_{k=1}^m \lambda_k f_k(h, s)\}} \end{aligned}$$

[0046] Another marginal probability $P_\theta(y', h'|s_i)$ is the same marginal probability required in a fully-observed LTG and thus can be computed in a similar fashion. In summary, the marginal probabilities can be calculated as:

[0047] Formula(13):

$$P_{\theta}(h'|y_i, s_i) = \frac{\exp(\sum_{k=1}^l \mu_k g_k(y_i, h') + \sum_{k=1}^m \lambda_k f_k(h', s_i))}{\sum_{h'} \exp(\sum_{k=1}^l \mu_k g_k(y_i, h') + \sum_{k=1}^m \lambda_k f_k(h', s_i))}$$

[0048] Formula(14):

$$P_{\theta}(y', h'|s_i) = \frac{\exp(\sum_{k=1}^l \mu_k g_k(y_i, h') + \sum_{k=1}^m \lambda_k f_k(h', s_i))}{\sum_{y', h'} \exp(\sum_{k=1}^l \mu_k g_k(y_i, h') + \sum_{k=1}^m \lambda_k f_k(h', s_i))}$$

[0049] Now the gradient of the objective function could be calculated via formula (10) and formula (11), and $O(\theta)$ can be maximized by standard techniques such as L-BFGS (Limited Memory Broyden–Fletcher–Goldfarb–Shanno) and loopy belief propagation (LBP). However, L-BFGS may not be efficient for large social graphs, since L-BFGS needs to scan the entire training set for normalization function computation and parameter update. On the other hand, the approximation technique LBP is unstable and may cause convergence problems. To this end, stochastic gradient descent (SGD) may be exploited for optimization, which is much more computationally efficient than both L-BFGS and LBP. In addition, SGD can be decomposed into sessions and parallelized using one or more frameworks for distributed processing, such as the Hadoop® MapReduce framework.

[0050] Labeling users' behaviors, however, may be infeasible and expensive in large-scale networks. Accordingly, semi-supervised learning can be of great practical value. The LTG model allows natural incorporation of unlabeled data for this technology. In the supervised learning setting described above, the objective function $O(\theta)$ is only optimized by using the labeled data. By involving unlabeled data, the computation of empirical distributions and conditional probabilities can be naturally extended. Given the independently and identically distributed (iid) social data, the expectations of feature functions $f_k(h, s)$ and $g_k(y, h)$ can be computed on the labeled data and are a good estimation of expectations of feature functions based on the whole dataset. To do this, the social action labeled data may be used to infer the unknown social action labels. If y^{ja} represents the labeling configuration inferred from the known social action labels y , and $y^{all} = y \cup y^{ja}$, the first term in $O(\theta)$ can be rewritten as:

[0051] Formula (15): $\sum_{(y^{all}, s)} \tilde{P}(y^{all}, s) \log P_{\theta}(y^{all}|s)$

[0052] Scalability issue in dealing with large-scale social networks may also present a challenge. This challenge may be addressed through the distributed implementation of the SGD learning algorithm based on a framework for distributed processing, such as the Hadoop® MapReduce framework for big social data. Since SGD computes the gradient over a few training instances, it is straightforward to implement it in a parallel manner under the master-slave architecture of such a framework. The gradient of the objective function can be computed distributively and independently on each computing data node in the map stage, and all gradient values and update model parameters on the name node can be merged in the reduce stage. The two stages are repeated until convergence.

[0053] Turning back to FIG. 1, social action engine 120 may predict a future social action based on the hidden variable and/or latent topic. As discussed above, one problem of social action prediction is to find the most probable social action labels y^* that have the maximum a posteriori (MAP) probability $P_{\theta}(y|s)$ for the new testing social data s . This could be inferred via maximizing the conditional model straightforwardly as:

[0054] Formula (16): $y^* = \arg \max_y P_{\theta}(y|s)$

[0055] In formula (16), the parameters θ are estimated from the training dataset. Since the conditional probability $P_{\theta}(y|s)$ is defined in formula (1), formula (16) can be rewritten as:

[0056] Formula (17): $y^* = \arg \max_y \sum_h P_{\theta}(y|h)P_{\theta}(h|s)$

[0057] Both $P_{\theta}(y|s)$ and $P_{\theta}(h|s)$ can be estimated using formula (3) and formula (4), respectively. Thus, the most likely social action labels y^* can be computed using formula (17).

[0058] Latent topic determination system 110 may also include a latent topic engine 118. The latent topic engine 118 may determine a latent topic based on the hidden variable. The latent topic may influence the social action of the user. The latent topics may also demonstrate a user's interests

for social action prediction. The most likely latent topics h^* can be obtained via maximizing the conditional probability $P_\theta(h|s)$:

[0059] Formula (18):
$$h^* = \arg \max_h P_\theta(h|s)$$

[0060] Since LTG learns an intermediate interdependencies and dynamics between social data s and social actions y in terms of latent topics h , the explainable latent topics should have flexible and reasonable dependencies with users' social actions to capture users' interests or preferences. This nature of LTG modeling allows latent topics $\hat{h}$ to be inferred as approximated most likely latent topics h^* . When the social action is predicted, the corresponding values of hidden variables are also inferred. That is, when the most likely social action label y^* is estimated using formula (17), latent topics $\hat{h}$ can be obtained such that $P_\theta(y|s)$ and $P_\theta(h|s)$ are maximized as:

[0061] Formula 19:
$$\hat{h} = \arg \max_y \sum_h P_\theta(y|h)P_\theta(h|s)$$

[0062] FIG. 3 is a flowchart of an example method 300 for latent topic determination. Method 300 may be described below as being executed or performed by a system, for example, system 110 of FIG. 1, system 600 of FIG. 6 or system 700 of FIG. 7. Other suitable systems and/or computing devices may be used as well. Method 300 may be implemented in the form of executable instructions stored on at least one machine-readable storage medium of the system and executed by at least one processor of the system. Alternatively or in addition, method 300 may be implemented in the form of electronic circuitry (e.g., hardware). In alternate embodiments of the present disclosure, one or more steps of method 300 may be executed substantially concurrently or in a different order than shown in FIG. 3. In alternate embodiments of the present disclosure, method 300 may include more or less steps than shown in FIG. 3. In some embodiments, one or more of the steps of method 300 may, at certain times, be ongoing and/or may repeat.

[0063] Method 300 may start at step 302 and continue to step 304 where the method may include determining a social action performed by a user in a social network. The techniques in block 304 may be performed similarly to the techniques described in relation to social action engine 112 discussed above in

reference to FIG. 1. At step 306, the method may include extracting a hidden variable from an observable content on the social network. The techniques in block 306 may be performed similarly to the techniques described in relation to hidden variable engine 116 discussed above in reference to FIG. 1. The hidden variable may correspond to a user characteristic. The observable content may be generated by the social action of the user. At step 308, the method may include determining a latent topic based on the hidden variable. The techniques in block 308 may be performed similarly to the techniques described in relation to latent topic engine 118 discussed above in reference to FIG. 1. The latent topic may influence the social action of the user. Method 300 may eventually continue to step 310, where method 300 may stop.

[0064] FIG. 4 is a flowchart of an example method 400 for latent topic determination. Method 400 may be described below as being executed or performed by a system, for example, system 110 of FIG. 1, system 600 of FIG. 6 or system 700 of FIG. 7. Other suitable systems and/or computing devices may be used as well. Method 400 may be implemented in the form of executable instructions stored on at least one machine-readable storage medium of the system and executed by at least one processor of the system. Alternatively or in addition, method 400 may be implemented in the form of electronic circuitry (e.g., hardware). In alternate embodiments of the present disclosure, one or more steps of method 400 may be executed substantially concurrently or in a different order than shown in FIG. 4. In alternate embodiments of the present disclosure, method 400 may include more or less steps than shown in FIG. 4. In some embodiments, one or more of the steps of method 400 may, at certain times, be ongoing and/or may repeat.

[0065] Method 400 may start at step 402 and continue to step 404 where the method may include determining a first probability that the social action depends on a hidden variable. The techniques in block 404 may be performed similarly to the techniques described in relation to hidden variable engine 116 discussed above in reference to FIG. 1 and may incorporate one or more formulas described in relation to hidden variable engine 116, such as formulas (1), (2) and (3). At step 406, the method may include determining a

second probability that the hidden variable depends on the observable content. The techniques in block 406 may be performed similarly to the techniques described in relation to hidden variable engine 116 discussed above in reference to FIG. 1 and may incorporate one or more formulas described in relation to hidden variable engine 116, such as formulas (1), (2) and (4). At step 408, the method may include determining the latent topic based on the first probability and the second probability. The latent topic may influence the social action of the user. The techniques in block 408 may be performed similarly to the techniques described in relation to latent topic 118 discussed above in reference to FIG. 1 and may incorporate one or more formulas described in relation to latent topic 118, such as formulas (17) and (18). At step 410, the method may include determining a balance parameter to weight the first probability that the social action depends on the hidden variable. The techniques in block 410 may be performed similarly to the techniques described in relation to hidden variable engine 116 discussed above in reference to FIG. 1 and may incorporate one or more formulas described in relation to hidden variable engine 116, such as formula (7). At step 412, the method may include determining a probability that the social action will occur on the social network. For example, the method may include optimizing a posterior probability that the social action will occur on the social network. The posterior probability may be optimized by determining the most likely types of social actions that have the maximum a posterior (MAP) probability in the social network. The techniques in block 412 may be performed similarly to the techniques described in relation to hidden variable engine 116 and/or social action engine 120 discussed above in reference to FIG. 1. The techniques in block 306 may also incorporate one or more formulas described in relation to hidden variable engine 116 and/or social action engine 120, such as formula (16) and formula (17). Method 400 may eventually continue to step 414, where method 400 may stop.

[0066] FIG. 5 is a flowchart of an example method 500 for latent topic determination. Method 500 may be described below as being executed or performed by a system, for example, system 110 of FIG. 1, system 600 of FIG. 6 or system 700 of FIG. 7. Other suitable systems and/or computing devices may

be used as well. Method 500 may be implemented in the form of executable instructions stored on at least one machine-readable storage medium of the system and executed by at least one processor of the system. Alternatively or in addition, method 500 may be implemented in the form of electronic circuitry (e.g., hardware). In alternate embodiments of the present disclosure, one or more steps of method 500 may be executed substantially concurrently or in a different order than shown in FIG. 5. In alternate embodiments of the present disclosure, method 500 may include more or less steps than shown in FIG. 5. In some embodiments, one or more of the steps of method 500 may, at certain times, be ongoing and/or may repeat.

[0067] Method 500 may start at step 502 and continue to step 504 where the method may include labeling the social action to create a labeled social action data. Labeling a latent topic may include associating a keyword or other descriptor to identify the latent topic. The techniques in block 504 may be performed similarly to the techniques described in relation to hidden variable engine 116 discussed above in reference to FIG. 1 and may incorporate one or more formulas described in relation to hidden variable engine 116. At step 506, the method may include determining a first marginal probability that the latent topic influences the labeled social action data on the social network. The techniques in block 506 may be performed similarly to the techniques described in relation to hidden variable engine 116 discussed above in reference to FIG. 1 and may incorporate one or more formulas described in relation to hidden variable engine 116, such as formulas (12) and (13). At step 508, the method may include determining a second marginal probability that the observable content influences the latent topic on the social network. The techniques in block 508 may be performed similarly to the techniques described in relation to hidden variable engine 116 discussed above in reference to FIG. 1 and may incorporate one or more formulas described in relation to hidden variable engine 116, such as formula (14). At step 510, the method may include determining a parameter for the latent topic based on the first marginal probability and the second marginal probability. The techniques in block 510 may be performed similarly to the techniques described in relation to hidden variable engine 116

discussed above in reference to FIG. 1 and may incorporate one or more formulas described in relation to hidden variable engine 116. At step 512, the method may include determining a label for an unlabeled social action. Labels for the unknown social actions can be inferred from the social action labeled data. The label for the unlabeled social action may also be determined using a latent topic graph model, such as the LTG model 200 discussed above in reference to FIG. 2. The techniques in block 512 may be performed similarly to the techniques described in relation to hidden variable engine 116 discussed above in reference to FIG. 1 and may incorporate one or more formulas described in relation to hidden variable engine 116, such as formula (15). An example latent topic graph model is discussed above in reference to FIG. 2. Method 500 may eventually continue to step 514, where method 500 may stop.

[0068] FIG. 6 is a block diagram of an example latent topic determination system 600. System 600 may be similar to system 110 of FIG. 1, for example. In the embodiment of FIG. 6, system 600 includes social action handler 602, observable content identifier 604, hidden variable extractor 606 and latent topic handler 608.

[0069] Social action handler 602 may determine a social action performed by a user in a social network. Social action handler 602 may be implemented in the form of executable instructions stored on at least one machine-readable storage medium of system 600 and executed by at least one processor of system 600. Alternatively or in addition, social action handler 602 may be implemented in the form of one or more hardware devices including electronic circuitry for implementing the functionality of social action handler 602.

[0070] Observable content identifier 604 may identify an observable content on the social network generated by the social action. Observable content identifier 604 may be implemented in the form of executable instructions stored on at least one machine-readable storage medium of system 600 and executed by at least one processor of system 600. Alternatively or in addition, observable content identifier 604 may be implemented in the form of one or more hardware devices including electronic circuitry for implementing the functionality of observable content identifier 604.

[0071] Hidden variable extractor 606 may extract a hidden variable from the observable content on the social network. The hidden variable may correspond to a user characteristic. Hidden variable extractor 606 may be implemented in the form of executable instructions stored on at least one machine-readable storage medium of system 600 and executed by at least one processor of system 600. Alternatively or in addition, hidden variable extractor 606 may be implemented in the form of one or more hardware devices including electronic circuitry for implementing the functionality of hidden variable extractor 606.

[0072] Latent topic handler 608 may determine a latent topic based on the hidden variable. The latent topic may define an interdependency between the social action and the observable content. Latent topic handler 608 may be implemented in the form of executable instructions stored on at least one machine-readable storage medium of system 600 and executed by at least one processor of system 600. Alternatively or in addition, latent topic handler 608 may be implemented in the form of one or more hardware devices including electronic circuitry for implementing the functionality of latent topic handler 608.

[0073] FIG. 7 is a block diagram of an example system 700 for latent topic determination. System 700 may be similar to system 110 of FIG. 1, for example. In the embodiment of FIG. 7, system 700 includes a processor 702 and a machine-readable storage medium 704. Although the following descriptions refer to a single processor and a single machine-readable storage medium, the descriptions may also apply to a system with multiple processors and multiple machine-readable storage mediums. In such examples, the instructions may be distributed (e.g., stored) across multiple machine-readable storage mediums and the instructions may be distributed (e.g., executed by) across multiple processors.

[0074] Processor 702 may be one or more central processing units (CPUs), microprocessors, and/or other hardware devices suitable for retrieval and execution of instructions stored in machine-readable storage medium 704. In the particular embodiment shown in FIG. 7, processor 702 may fetch, decode, and execute instructions 706, 708 and 710 to perform latent topic determination. As an alternative or in addition to retrieving and executing instructions, processor 702

may include one or more electronic circuits comprising a number of electronic components for performing the functionality of one or more of the instructions in machine-readable storage medium 704. With respect to the executable instruction representations (e.g., boxes) described and shown herein, it should be understood that part or all of the executable instructions and/or electronic circuits included within one box may, in alternate embodiments, be included in a different box shown in the figures or in a different box not shown.

[0075] Machine-readable storage medium 704 may be any electronic, magnetic, optical, or other physical storage device that stores executable instructions. Thus, machine-readable storage medium 704 may be, for example, Random Access Memory (RAM), an Electrically-Erasable Programmable Read-Only Memory (EEPROM), a storage drive, an optical disc, and the like. Machine-readable storage medium 704 may be disposed within system 700, as shown in FIG. 7. In this situation, the executable instructions may be “installed” on the system 700. Alternatively, machine-readable storage medium 704 may be a portable, external or remote storage medium, for example, that allows system 700 to download the instructions from the portable/external/remote storage medium. In this situation, the executable instructions may be part of an “installation package”. As described herein, machine-readable storage medium 704 may be encoded with executable instructions for a web technology responsive to mixtures of emotions.

[0076] Referring to FIG. 7, social action instructions 706, when executed by a processor (e.g., 702), may cause system 700 to determine a social action performed by a user in a social network. Hidden variable instructions 708, when executed by a processor (e.g., 702), may cause system 700 to extract a hidden variable from an observable content on the social network, wherein the hidden variable defines a set of expressions that identify a user interest. The observable content may be generated by the social action of the user.

[0077] Latent topic instructions 710, when executed by a processor (e.g., 702), may cause system 700 to determine a latent topic based on the hidden variable. The latent topic may influence the social action of the user. The social action may correspond to the user interest.

[0078] The foregoing disclosure describes a number of examples for latent topic determination. The disclosed examples may include systems, devices, computer-readable storage media, and methods for latent topic determination. For purposes of explanation, certain examples are described with reference to the components illustrated in FIGS. 1-7. The functionality of the illustrated components may overlap, however, and may be present in a fewer or greater number of elements and components. Further, all or part of the functionality of illustrated elements may co-exist or be distributed among several geographically dispersed locations. Further, the disclosed examples may be implemented in various environments and are not limited to the illustrated examples.

[0079] Further, the sequence of operations described in connection with FIGS. 1-7 are examples and are not intended to be limiting. Additional or fewer operations or combinations of operations may be used or may vary without departing from the scope of the disclosed examples. Furthermore, implementations consistent with the disclosed examples need not perform the sequence of operations in any particular order. Thus, the present disclosure merely sets forth possible examples of implementations, and many variations and modifications may be made to the described examples.

What is claimed is:

CLAIMS

1. A method for determining latent topics, the method comprising:
determining, by a processor, a social action performed by a user in a social network;
extracting, by the processor, a hidden variable from an observable content on the social network, wherein the hidden variable corresponds to a user characteristic; and
determining, by the processor, a latent topic based on the hidden variable, wherein the latent topic influences the social action of the user.
2. The method of claim 1 further comprising:
determining, by the processor, a first probability that the social action depends on the hidden variable;
determining, by the processor, a second probability that the hidden variable depends on the observable content; and
determining, by the processor, the latent topic based on the first probability and the second probability.
3. The method of claim 2 further comprising:
determining, by the processor, a balance parameter to weight the first probability that the social action depends on the hidden variable.
4. The method of claim 1 further comprising:
predicting, by the processor, a future social action based on the latent topic.
5. The method of claim 1 further comprising:
extracting, by the processor, a key word from the observable content;
associating, by the processor, the key word with the latent topic; and
creating, by the processor, a group of related latent topics.
6. The method of claim 1 further comprising:

labeling, by the processor, the social action to create a labeled social action data;

determining, by the processor, a first marginal probability that the latent topic influences the labeled social action data on the social network;

determining, by the processor, a second marginal probability that the observable content influences the latent topic on the social network; and

determining, by the processor, a parameter for the latent topic based on the first marginal probability and the second marginal probability.

7. The method of claim 6 further comprising:

determining, by the processor, a label for an unlabeled social action using a latent topic graph model.

8. The method of claim 1 further comprising:

optimizing a posterior probability that the social action will occur on the social network.

9. A system for determining latent topics, the system comprising:

a social action handler to determine a social action performed by a user in a social network;

an observable content identifier to identify an observable content on the social network generated by the social action;

a hidden variable extractor to extract a hidden variable from the observable content on the social network, wherein the hidden variable corresponds to a user characteristic; and

a latent topic handler to determine a latent topic based on the hidden variable, wherein the latent topic defines an interdependency between the social action and the observable content.

10. The system of claim 9 wherein the latent topic handler is further to:

determine a first probability that the social action depends on the hidden variable; determine a second probability that the hidden variable depends on the

observable content; and

determine the latent topic based on the first probability and the second probability.

11. The system of claim 9 further comprising:
a social action predictor to predict a future social action based on the latent topic.

12. A non-transitory machine-readable storage medium comprising instructions executable by a processor of a computing device for determining latent topics, the machine-readable storage medium comprising:

instructions to determine a social action performed by a user in a social network;

instructions to extract a hidden variable from an observable content on the social network, wherein the hidden variable defines a set of expressions that identify a user interest; and

instructions to determine a latent topic based on the hidden variable, wherein the latent topic influences the social action of the user and the social action corresponds to the user interest.

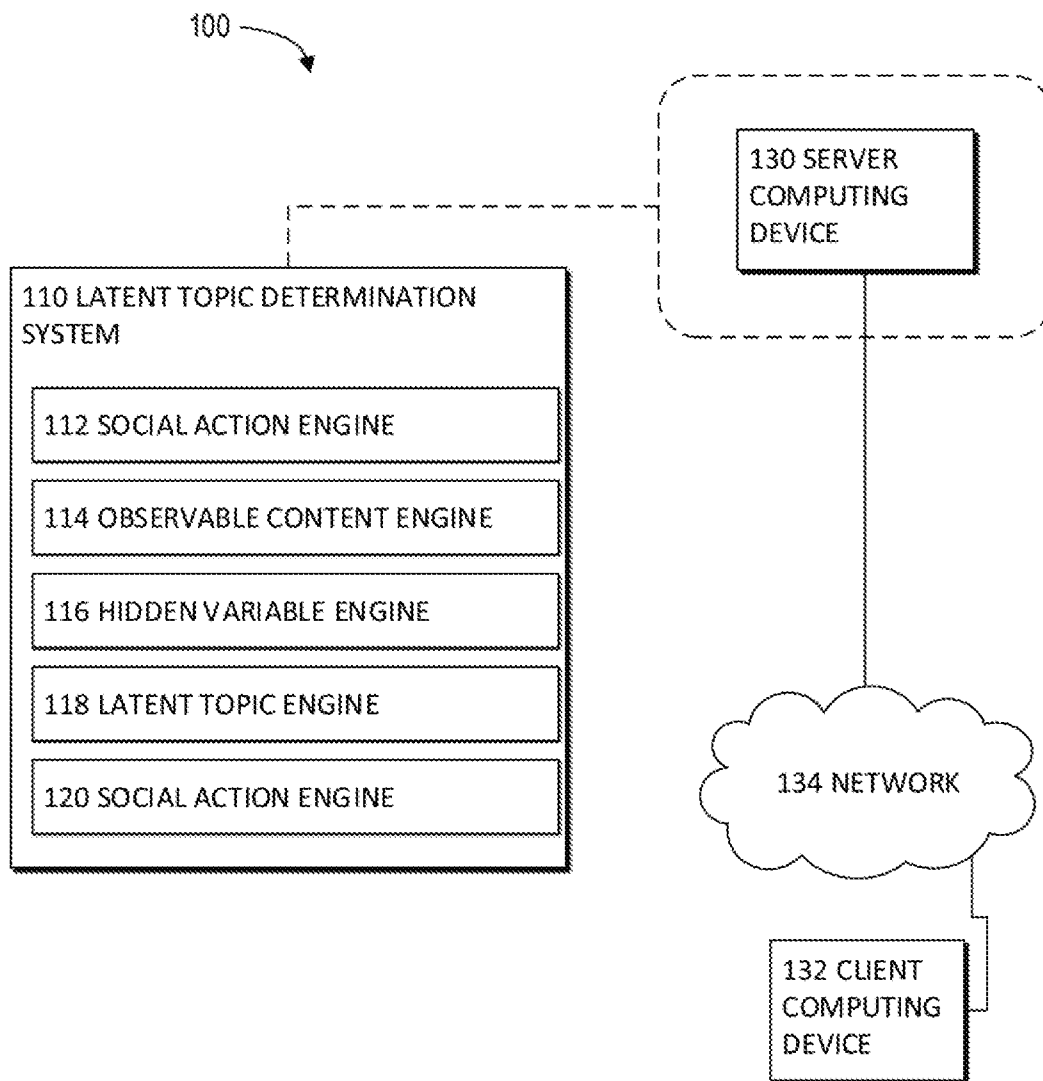
13. The non-transitory machine-readable storage medium of claim 12, wherein the observable content is generated by the social action.

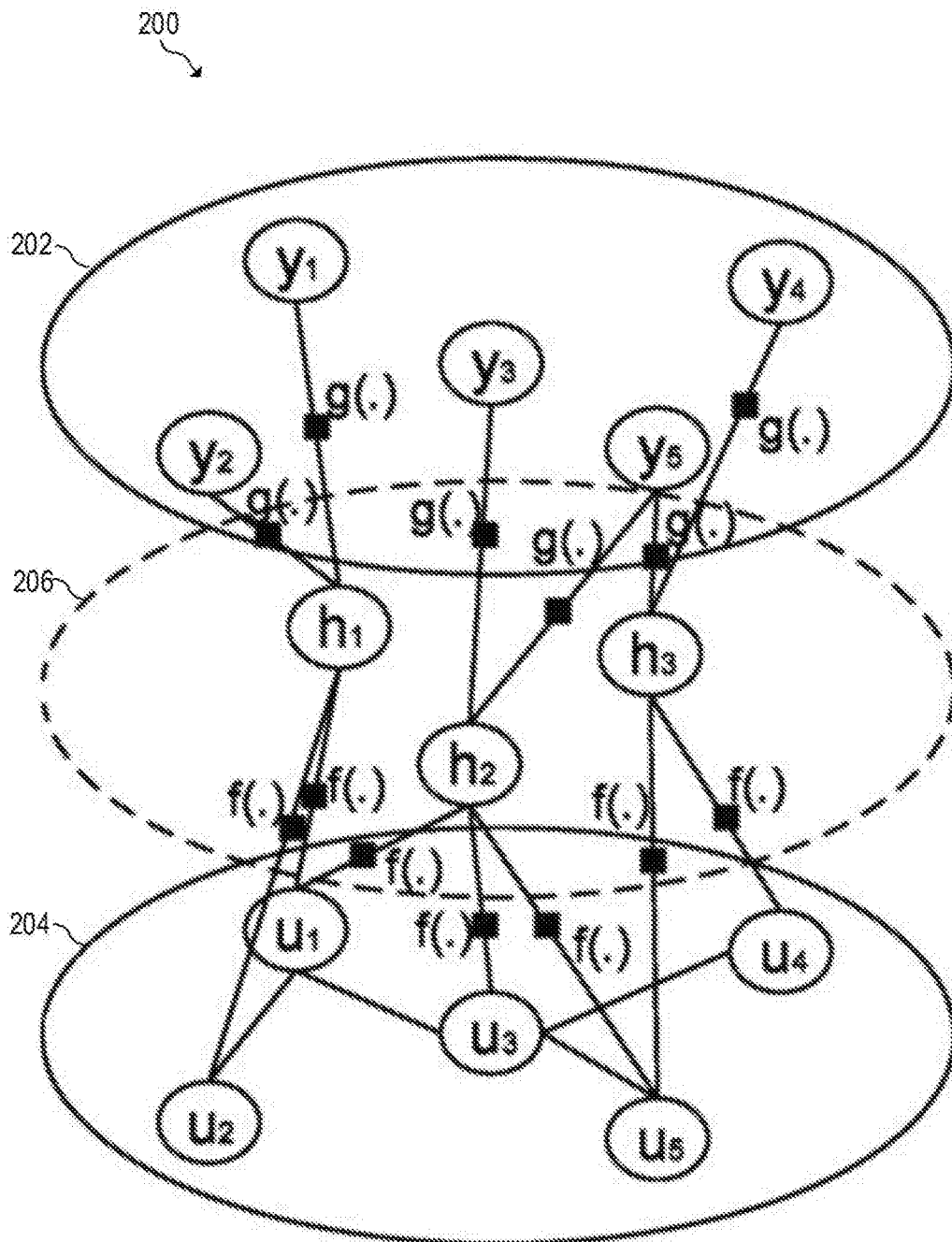
14. The non-transitory machine-readable storage medium of claim 12 further comprising:

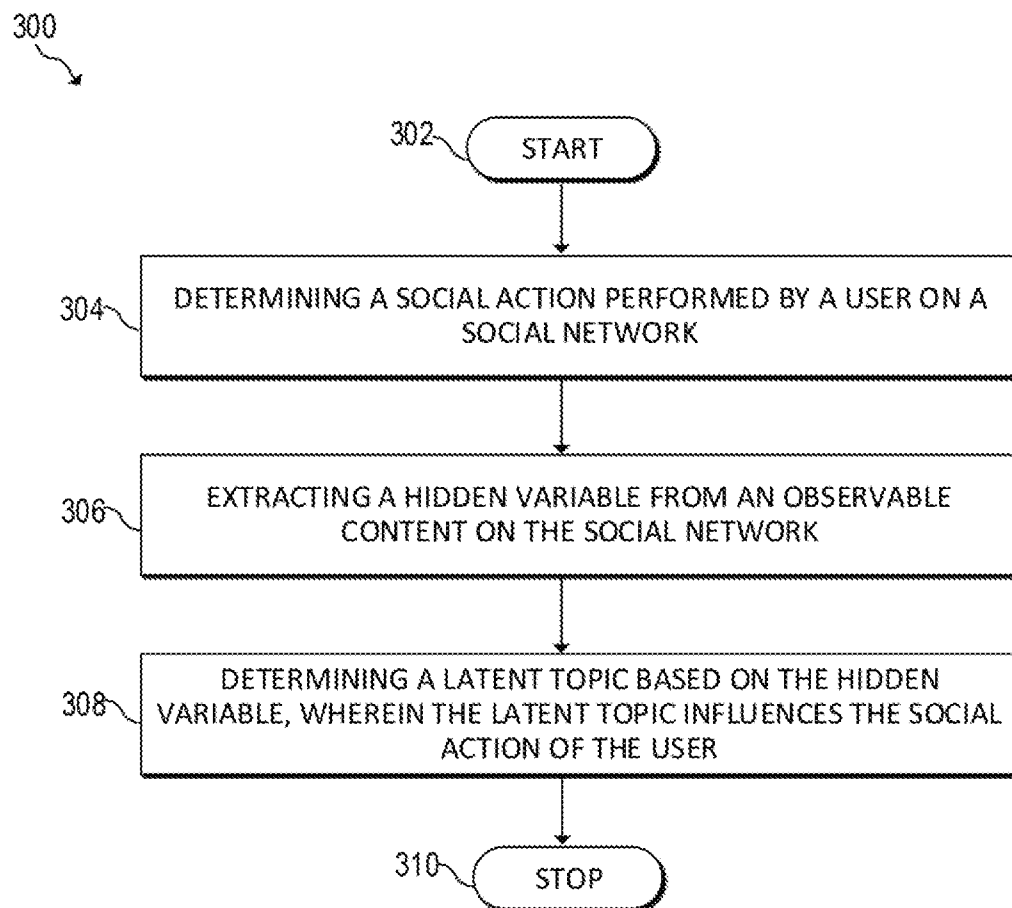
instructions to predict a future social action based on the latent topic.

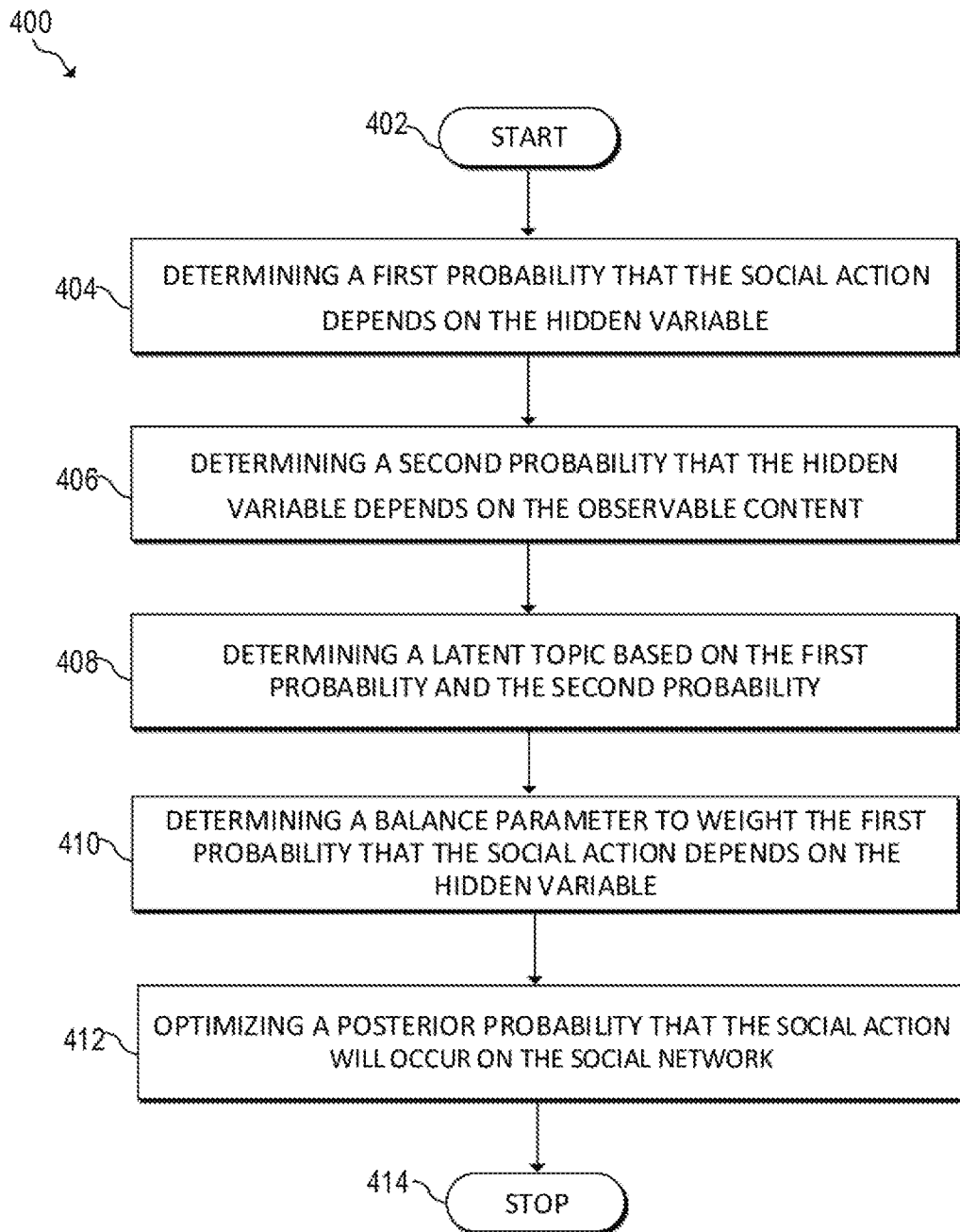
15. The non-transitory machine-readable storage medium of claim 12 further comprising:

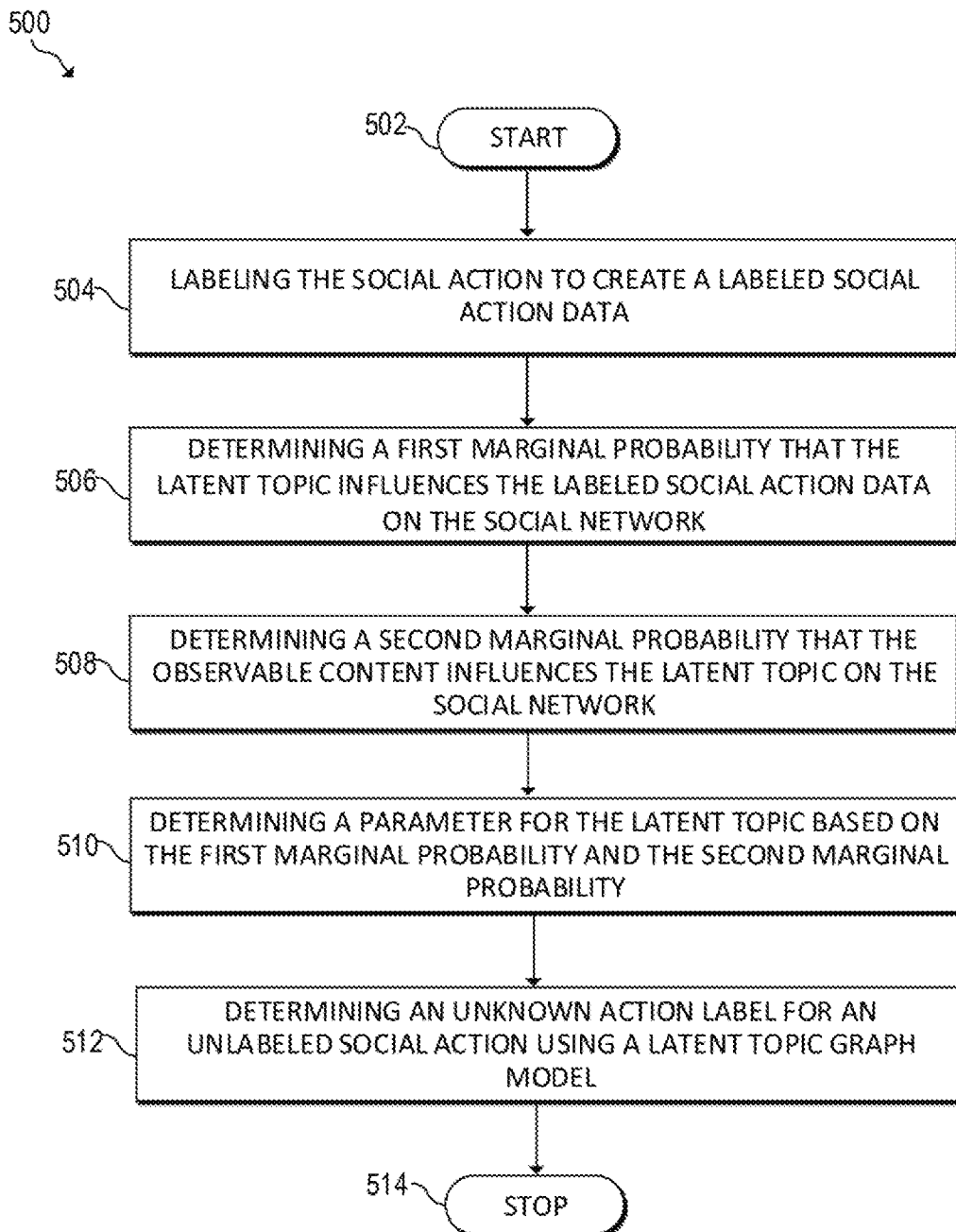
instructions to label the social action to create a labeled social action data;
instructions to determine a first marginal probability that the latent topic influences the labeled social action data on the social network;
instructions to determine a second marginal probability that the observable content action data influences the latent topic on the social network; and
instructions to determine a parameter for the latent topic based on the first marginal probability and the second marginal probability.

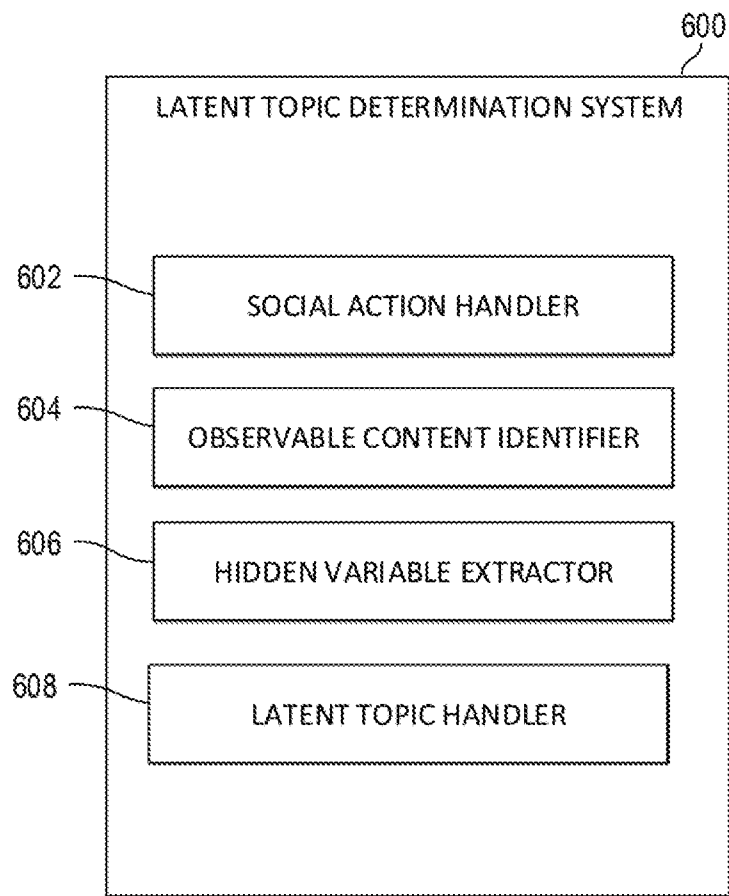
**FIG. 1**

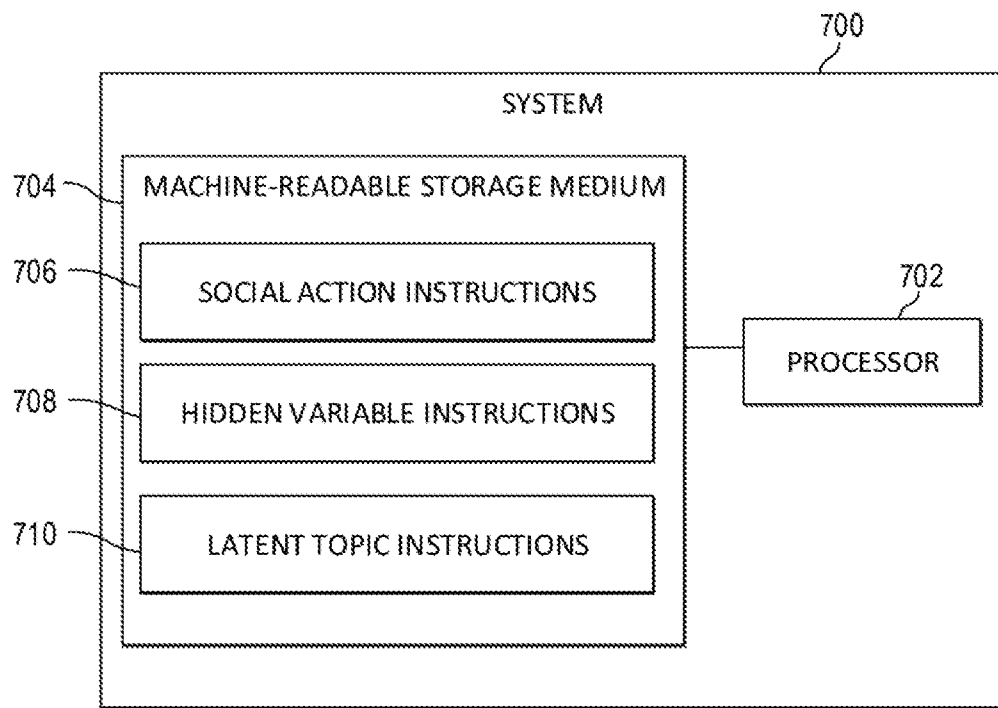
**FIG. 2**

**FIG. 3**

**FIG. 4**

**FIG. 5**

**FIG. 6**

**FIG. 7**

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2015/082512

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

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)

CNKI, CNPAT, WPI, EPODOC: latent,topics,social,action,hidden,variable,content,characteristic,probability,first,second

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2014193424 A1 (INTEL CORPET AL.) 04 December 2014 (2014-12-04) see abstract, description, page 7, line 13 to page 8, line 14, and page 18, line 6 to page 23, line 30	1-15
A	CN 103812872 A (CHINESE ACAD SCI INFORMATION ENG INST) 21 May 2014 (2014-05-21) see the whole document	1-15
A	CN 104077417 A (INST COMPUTING TECH CHINESE ACAD SCI) 01 October 2014 (2014-10-01) see the whole document	1-15
A	CN 102687166 A (NOKIA CORP) 19 September 2012 (2012-09-19) see the whole document	1-15

☐ 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

24 March 2016

Date of mailing of the international search report

31 March 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

Facsimile No. (86-10)62019451

Authorized officer

ZHANG,Guihua

Telephone No. (86-10)62089416

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2015/082512

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
WO	2014193424	A1	04 December 2014	US	2014379729	A1	25 December 2014
				CN	105247564	A	13 January 2016
				KR	20150137102	A	08 December 2015
CN	103812872	A	21 May 2014	None			
CN	104077417	A	01 October 2014	None			
CN	102687166	A	19 September 2012	WO	2011079462	A1	07 July 2011
				EP	2519920	A1	07 November 2012
				US	2012290599	A1	15 November 2012
				EP	2519920	A4	21 August 2013
				US	9031952	B2	12 May 2015