



- (51) **International Patent Classification:**
G06F 17/21 (2006.01)
- (21) **International Application Number:**
PCT/CN2015/076307
- (22) **International Filing Date:**
10 April 2015 (10.04.2015)
- (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, United States of America (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).

Published:

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

(54) **Title:** HIDDEN DYNAMIC SYSTEMS

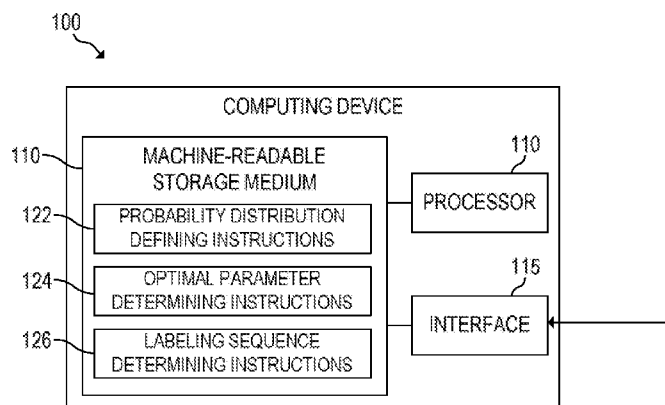


FIG. 1

(57) **Abstract:** Disclosed is a hidden dynamic system. In the hidden dynamic system, a conditional probability distribution for labeling data record segments is defined, where the conditional probability distribution models dependencies between class labels and internal substructures of the data record segments. At this stage, optimal parameter values are determined for the conditional probability distribution by applying a quasi-Newton gradient ascent method to training data, where the conditional probability distribution is restricted to a disjoint set of hidden states for each of the class labels. The conditional probability distribution and the optimal parameter values are used to determine a most probable labeling sequence for the data record segments.

HIDDEN DYNAMIC SYSTEMS

BACKGROUND

[0001] Annotating or labeling observation sequences arises in many applications across a variety of scientific disciplines, most prominently in natural language processing, information extraction, speech recognition, and bio-informatics. Recently, the predominant formalism for modeling and predicting label sequences has been based on discriminative models and variants. Conditional Random Fields (CRFs) are perhaps the most commonly used technique for probabilistic sequence modeling.

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 device for analyzing data using hidden dynamic systems;

[0004] FIG. 2 is a block diagram of an example computing device in communication with server devices for analyzing data using hidden dynamic systems;

[0005] FIG. 3 is a flowchart of an example method for execution by a computing device for analyzing data using hidden dynamic systems; and

[0006] FIG. 4 is a graph of example hidden dynamic conditional random fields (HDCRFs).

DETAILED DESCRIPTION

[0007] As detailed above, CRFs are commonly used for probabilistic sequence modeling. Structured data are widely prevalent in the real world, and

observation sequences tend to have distinct internal sub-structure and indicate predictable relationships between individual class labels, especially for natural language. For example in the task of noun phrase chunking, a noun phrase begins with a noun or a pronoun and may be accompanied by a set of modifiers. In this example, a noun phrase may contain one or more base noun phrases. In the named entity recognition task, named entities have particular characteristics in their composition. A location name can end with a location salient word but cannot end with any organization salient word. A complex, nested organization name may be composed of a person name, a location name, or even another organization name. Such complex and expressive structures can largely influence predictions. The efficiency of the CRF approach heavily depends on its first order Markov property - given the observation, the label of a token is assumed to depend only on the labels of its adjacent tokens. Further, the CRF approach models the transitions between class labels to enjoy advantages of both generative and discriminative methods capture external dynamics without consideration for internal sub-structure.

[0008] In examples described herein, the internal sub-structure in sequence data is directly modeled by incorporating a set of observed variables with additional latent, or hidden state variables to model relevant sub-structure in a given sequence, resulting in a new discriminative framework, Hidden Dynamic Conditional Random Fields (HDCRFs). The model learns the external dependencies by modeling a continuous stream of class labels and learns internal sub-structure by utilizing intermediate hidden states. HDCRFs define a conditional distribution over the class labels and hidden state labels conditioned on the observations, where dependencies between the hidden variables can be expressed by an undirected graph. Such modeling is able to deal with features that can be arbitrary functions of the observations. Efficient parameter estimation and inference can be carried out using standard graphical model algorithms such as belief propagation.

[0009] For example in web data extraction from encyclopedic pages such as WIKIPEDIA®, each encyclopedic page has a major topic or concept represented by a principal data record such as “Beijing”. A goal of HDCRFs is to extract all the interested data records such as “Beijing municipality”, “October 28”, “1420”, and “Qing Dynasty”, and assign class labels to these data records. In this example, the class labels can include pre-defined labels such as “person”, “date”, “year”, “organization”

labels assigned to each data record and hidden state variables to identify substructures like the relationship between “Beijing” and “municipality” or “Qing” and “Dynasty.” If the substructure between “Beijing” and “municipality” is identified, “Beijing municipality” can be properly labeled as an “organization.” WIKIPEDIA® is a registered trademark of the Wikimedia Foundation, Inc., which is headquartered in San Francisco, CA.

[0010] In some examples, a conditional probability distribution for labeling data record segments is defined, where the conditional probability distribution models dependencies between class labels and internal substructures of the data record segments. Data record segments may be observed data such as content from web pages, text from books, documents, etc. At this stage, optimal parameter values are determined for the conditional probability distribution by applying a quasi-Newton gradient ascent method to training data, where the conditional probability distribution is restricted to a disjoint set of hidden states for each of the class labels. The conditional probability distribution and the optimal parameter values are used to determine a most probable labeling sequence for the data record segments.

[0011] Referring now to the drawings, FIG. 1 is a block diagram of an example computing device 100 for analyzing data using hidden dynamic systems. Computing device 100 may be any computing device capable of accessing server devices, such as server devices 250A, 250N of FIG. 2. In the embodiment of FIG. 1, computing device 100 includes a processor 110, an interface 115, and a machine-readable storage medium 120.

[0012] Processor 110 may be central processing unit(s) (CPUs), microprocessor(s), and/or other hardware device(s) suitable for retrieval and execution of instructions stored in machine-readable storage medium 120. Processor 110 may fetch, decode, and execute instructions 122, 124, 126 to enable analyzing data using hidden dynamic systems (e.g., hidden states). As an alternative or in addition to retrieving and executing instructions, processor 110 may include electronic circuits comprising a number of electronic components for performing the functionality of instructions 122, 124, 126.

[0013] Interface 115 may include a number of electronic components for communicating with a server device. For example, interface 115 may be an Ethernet

interface, a Universal Serial Bus (USB) interface, an IEEE 1394 (Firewire) interface, an external Serial Advanced Technology Attachment (eSATA) interface, or any other physical connection interface suitable for communication with the server device. Alternatively, interface 115 may be a wireless interface, such as a wireless local area network (WLAN) interface or a near-field communication (NFC) interface. In operation, as detailed below, interface 115 may be used to send and receive data to and from a corresponding interface of a server device.

[0014] Machine-readable storage medium 120 may be any electronic, magnetic, optical, or other physical storage device that stores executable instructions. Thus, machine-readable storage medium 120 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. As described in detail below, machine-readable storage medium 120 may be encoded with executable instructions for analyzing data using hidden dynamic systems.

[0015] Probability distribution defining instructions 122 define a probability distribution for labeling observation sequences. Suppose X is a random variable over data sequences to be labeled, and Y is a random variable over corresponding label sequences. The distribution defines mappings between an observation sequence $X = (x_1, x_2, \dots, x_T)$ and the corresponding label sequence $Y = (y_1, y_2, \dots, y_T)$. Each y_j is a member of the possible class label set. For each sequence, a vector of sub-structure variables $S = (s_1, s_2, \dots, s_T)$ are assumed, which are not observed in training examples and, thus, form a set of hidden variables. Each s_j is a member of a finite set S_{y_j} of possible hidden states for the class label y_j . Suppose S is the set of all possible hidden states of all S_{y_j} sets. Each s_j corresponds to a labeling of x_j with some member of S , which may correspond to substructure of the sequence.

[0016] Given the above definitions, a hidden dynamic probabilistic model can be defined as follows:

$$p(Y|X) = \sum_S p(Y|S, X) \cdot p(S|X) \quad (1)$$

By definition, sequences which have any $s_j \notin S_{y_j}$ will obviously have $p(Y|S, X) = 0$, so the model above can be rewritten as:

$$p(Y|X) = \sum_{S: \forall s_j \in \mathcal{S}_{y_j}} p(S|X) \quad (2)$$

Similar to CRFs, the conditional probability distributions, $p(S|X)$, can take the form:

$$p(S|X) = \frac{1}{Z(X)} \exp \left(\sum_k \lambda_k \cdot \sum_{j=1}^T f_k(s_{j-1}, s_j, X, j) \right), \quad (3)$$

where $Z(X)$ is an instance-specific normalization function:

$$Z(X) = \sum_S \exp \left(\sum_k \lambda_k \cdot \sum_{j=1}^T f_k(s_{j-1}, s_j, X, j) \right), \quad (4)$$

and $f_k(s_{j-1}, s_j, X, j)_{k=1}^K$ is a set of real-valued feature functions.

$\Lambda = \{\lambda_k\} \in \mathbb{R}^K$ is a parameter vector that reflects the confidence of feature functions. Each feature function can be either a transition function $t_k(s_{j-1}, s_j, X, j)$ over the entire observation sequence and the hidden variables at positions i and $i-1$, or a state function $s_k(s_j, X, j)$ depends on a single hidden variable at position i . Note that the model is different from hidden conditional random fields (HCRFs), which model the conditional probability of one class label y given the observation sequence X through:

$$p(y|X) = \frac{1}{Z'(X)} \sum_{S \in y} \exp \left(\lambda \cdot f(y, S, X) \right), \quad (5)$$

where the partition function $Z'(X)$ is given by:

$$Z'(X) = \sum_{y, S \in y} \exp \left(\lambda \cdot f(y, S, X) \right). \quad (6)$$

[0017] HDCRFs combine the strengths of CRFs and HCRFs by modeling both external dependencies between class labels and internal substructure. Specifically, the weights λ_k associated with the transition function $t_k(s_{j-1}, s_j, X, j)$ model both the internal sub-structure and external dependencies between different class labels. Weights associated with a transition function for hidden states that are in the same subset $\mathcal{S}_{y_j}$

model the substructure patterns while weights associated with the transition functions for hidden states from different subsets will model the external dependencies between labels.

[0018] Optimal parameter determining instructions 124 determine optimal parameters for probability distribution. Given some training data consist of n labeled sequences $D = (X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$, the parameters $\Lambda = \{\lambda_k\}$ are set to maximize the conditional log-likelihood. Following previous work on CRFs, the following objective function can be used to estimate the parameters:

$$L(\Lambda) = \sum_{i=1}^n \log P_{\Lambda}(Y_i | X_i) \quad (7)$$

To avoid over-fitting, log-likelihood can be penalized by a prior distribution over parameters that provide smoothing to help with sparsity in the training data. A commonly used prior is a zero-mean (with variance σ^2) Gaussian. With a Gaussian prior, log-likelihood is penalized as follows:

$$L(\Lambda) = \sum_{i=1}^n \log P_{\Lambda}(Y_i | X_i) - \sum_{k=1}^K \frac{\lambda_k^2}{2\sigma^2} \quad (8)$$

[0019] Structural constraints can be encoded with an undirected graph structure, where the hidden variables $\{s_1, s_2, \dots, s_T\}$ correspond to vertices in the graph. To ensure the training and inference remains tractable, the model can be restricted to have disjoint sets (i.e., a set that contains no elements in common) of hidden states associated with each class label. A quasi-Newton gradient ascent method can be used to search for the optimal parameter values, $\Lambda^* = \arg \max_{\Lambda} L(\Lambda)$, under this criterion.

$$\begin{aligned} \forall Y \in \mathcal{Y}, j \in 1 \dots T, a \in \mathcal{S}, \\ P(s_j = a | Y, X) = \sum_{S: s_j = a} P(S | Y, X) \end{aligned} \quad (9)$$

$$\begin{aligned} \forall Y \in \mathcal{Y}, j, k \in 1 \dots T, a, b \in \mathcal{S}, \\ P(s_j = a, s_k = b | Y, X) = \sum_{S: s_j = a, s_k = b} P(S | Y, X) \end{aligned} \quad (10)$$

where $P(s_j = a|Y, X)$ and $P(s_j = a, s_k = b|Y, X)$ are marginal distributions over individual variables s_j or pairs of variables $\{s_j, s_k\}$ corresponding to edges in the graph. The gradient of $L(\Lambda)$ can be defined in terms of these marginal distributions and can therefore be calculated efficiently.

[0020] We first consider derivatives with respect to the parameters λ_k associated with a state function s_k . Taking derivatives results in:

$$\begin{aligned} \frac{\partial L(\Lambda)}{\partial \lambda_k} &= \sum_S P(S|Y, X) \sum_{j=1}^T s_k(s_j, X, j) - \sum_{Y', S} P(Y', S|X) \sum_{j=1}^T s_k(s_j, X, j) \\ &= \sum_{j,a} P(s_j = a|Y, X) s_k(j, a, X) - \sum_{Y', j, a} P(s_j = a, Y'|X) s_k(j, a, X) \end{aligned} \quad (11)$$

It shows that $\frac{\partial L(\Lambda)}{\partial \lambda_k}$ can be expressed in terms of components $P(s_j = a|X)$ and $P(Y|X)$, which can be computed using belief propagation.

[0021] For derivatives with respect to the parameters λ_l corresponding to a transition function t_l , a similar calculation provides:

$$\begin{aligned} \frac{\partial L(\Lambda)}{\partial \lambda_l} &= \sum_{j,k,a,b} P(s_j = a, s_k = b|Y, X) t_l(j, k, a, b, X) \\ &\quad - \sum_{Y', j,k,a,b} P(s_j = a, s_k = b, Y'|X) t_l(j, k, a, b, X) \end{aligned} \quad (12)$$

hence $\frac{\partial L(\Lambda)}{\partial \lambda_k}$ can also be expressed in terms of expressions (e.g., the marginal probabilities $P(s_j = a, s_k = b|Y, X)$) that can be computed efficiently using belief propagation. Gradient ascent can be performed with the limited-memory quasi-Newton BFGS optimization technique.

[0022] Labeling sequence determining instructions 126 determine a labeling sequence for observation data (e.g., data record segments). Given a new test sequence X , the most probable labeling sequence Y^* can be estimated that maximizes the conditional model:

$$Y^* = \arg \max_Y P(Y|X, \Lambda^*) \quad (13)$$

where the parameters Λ^* are learned via a training process. Assuming each class label is associated with a disjoint set of hidden states, equation (13) can be rewritten as:

$$Y^* = \arg \max_Y \sum_{S: \forall s_j \in S_{y_j}} P(Y|X, \Lambda^*) \quad (14)$$

The marginal probabilities $P(s_j = a|X, \Lambda^*)$ can be computed for all possible hidden states $a \in S$ to estimate the label y_j^* . These marginal probabilities may then be summed according to the disjoint sets of hidden states S_{y_j} , and the label associated with the optimal set can be selected. As discussed in the previous subsection, these marginal probabilities can also be computed efficiently using belief propagation. For example, the above maximal marginal probabilities approach can be used to estimate the sequence of labels because it minimizes error.

[0023] FIG. 2 is a block diagram of an example computing device 200 for analyzing data using hidden dynamic systems. Computing device 200 may be, for example, a computing device, a desktop computer, a rack-mount server, or any other computing device suitable for execution of the functionality described below. Computing device 200 is in communication with server devices 250A, 250N via a network 245.

[0024] In the embodiment of FIG. 2, computing device 200 includes interface module 210, modeling module 220, training module 226, and inference module 230. While computing device 200 may include a number of modules 210-234. Each of the modules may include a series of instructions encoded on a machine-readable storage medium and executable by a processor of computing device 200. In addition or as an alternative, each module may include one or more hardware devices including electronic circuitry for implementing the functionality described below.

[0025] Interface module 210 may manage communications with the server devices 250A, 250N. Specifically, the interface module 210 may initiate connections with the server devices 250A, 250N and then send or receive observation data (e.g., data record segments) to/from the server devices 250A, 250N.

[0026] Modeling module 220 generates hidden dynamic probabilistic models for analyzing data. Specifically, modeling module 220 may generate a probabilistic

model as described above with respect to FIG. 1. Hidden states module 222 of modeling module 220 can manage a set of hidden states to be used in probabilistic functions. The hidden states can be used to model the internal substructure of an observation sequence. External dependencies module 224 of modeling module 220 models external dependencies between class labels and the internal substructure. Weights associated with a transition function for hidden states that are in the same subset model the sub-structure patterns, while weights associated with the transition functions for hidden states from different subsets will model the external dependencies between labels.

[0027] Training module 226 is to estimate parameters of the probabilistic model. Specifically, training module 226 uses training data to maximize the conditional log-likelihood function.

[0028] Analysis module 230 is to determine the most probably labeling sequence for observation data (e.g., data record segments). Specifically, labeling sequence module 234 of analysis module 230 computes marginal probabilities for all possible hidden states to estimate a label. Then these marginal probabilities are summed according to the disjoint sets of hidden states and the label associated with the optimal set is chosen.

[0029] Server devices 250A, 250N may be any servers accessible to computing device 200 over a network 245 that is suitable for executing the functionality described below. As detailed below, each server device 250A, 250N may include a series of modules 260-264 for providing web content.

[0030] API module 260 is configured to provide access to observation data of server device A 250A. Content module 262 of API module 260 is configured to provide the observation data as content over the network 245. For example, the content can be provided as HTML pages that are configured to be displayed in web browsers. In this example, server computer device 200 obtains the HTML pages from the content module 262 for processing as observation data as described above.

[0031] Metadata module 264 of API module 260 manages metadata related to the content. The metadata describes the content and can be included in, for example, web pages provided by the content module 262. In this example, keywords describing various page elements can be embedded as metadata in the web pages.

[0032] FIG. 3 is a flowchart of an example method 300 for execution by a computing device 100 for analyzing data using hidden dynamic systems. Although execution of method 300 is described below with reference to computing device 100 of FIG. 1, other suitable devices for execution of method 300 may be used, such as computing device 200 of FIG. 2. Method 300 may be implemented in the form of executable instructions stored on a machine-readable storage medium, such as storage medium 120, and/or in the form of electronic circuitry.

[0033] Method 300 may start in block 305 and continue to block 310, where computing device 100 generates a hidden dynamic probabilistic model for analyzing data using hidden dynamic systems. The probabilistic model can include hidden states for modeling the internal substructure of an observation sequence. Further, weights associated with a transition function for hidden states that are in the same subset model the sub-structure patterns, while weights associated with the transition functions for hidden states from different subsets will model the external dependencies between labels.

[0034] In block 315, computing device 100 determines optimal parameters of the probabilistic model by applying an ascent method. In block 320, computing device 100 uses the probabilistic model and the optimal parameters to determine the most probably labeling sequence for observation data. Method 300 may then continue to block 325, where method 300 may stop.

[0035] FIG. 4 is a graph 400 of example hidden dynamic conditional random fields (HDCRFs). The graph 400 shows an observation sequence 406A-406N with potential labels 402A-402N. As shown, hidden variables 404A-404N model the internal substructure of observation sequence 406A-406N. In this example, only links with the current observation set is shown, but long range dependencies can also be possible.

[0036] In the example graph 400, each transition function defines an edge feature while each state function defines a node feature as described above with respect to FIG. 1. All the observed data 406A-406N respect the structure of graph in that no observed data depends on more than two of the hidden variables 404A-404N, and if a feature does depend on two hidden variables, there should be a

corresponding edge in the graph. Example graph 400 can be encoded arbitrarily to capture domain specific knowledge such as the internal sub-structure.

[0037] The foregoing disclosure describes a number of examples for analyzing data using hidden dynamic systems. In this manner, the example disclosed herein improves labeling of observation data by modeling both external dependencies between the class labels and internal substructure of the observation data.

CLAIMS

1. A server computing device for analyzing data using hidden dynamic systems, the computing device comprising:

a processor to:

define a conditional probability distribution for labeling a plurality of data record segments, wherein the conditional probability distribution models dependencies between class labels and internal substructures of the plurality of data record segments;

determine optimal parameter values for the conditional probability distribution by applying a quasi-Newton gradient ascent method to training data, wherein the conditional probability distribution is restricted to a disjoint set of hidden states for each of the class labels; and

use the conditional probability distribution and the optimal parameter values to determine a most probable labeling sequence for the plurality of data record segments.

2. The server computing device of claim 1, wherein the conditional probability distribution is defined as

$$p(S|X) = \frac{1}{Z(X)} \exp \left(\sum_k \lambda_k \cdot \sum_{j=1}^T f_k(s_{j-1}, s_j, X, j) \right),$$

and wherein X is an observation sequence, Y is a potential labeling sequence, S is a vector of sub-structure variables, and λ_k is a confidence parameter.

3. The server computing device of claim 2, wherein the quasi-Newton gradient ascent method is performed using a Gaussian prior defined as

$$L(\Lambda) = \sum_{i=1}^n \log P_{\Lambda}(Y_i|X_i) - \sum_{k=1}^K \frac{\lambda_k^2}{2\sigma^2},$$

and wherein Λ is a set of parameters that includes the confidence parameter.

4. The server computing device of claim 2, wherein the plurality of data segments are applied to the conditional probability distribution to determine a plurality of marginal probabilities for each of the class labels.

5. The server computing device of claim 4, wherein the plurality of marginal probabilities are summed according to the disjoint sets of hidden states to determine the most probably labeling sequence.

6. The server computing device of claim 2, wherein the confidence parameters and a transition function f_k model dependencies between the class labels and the internal substructures.

7. A method for analyzing data using hidden dynamic systems, comprising:
defining a conditional probability distribution for labeling a plurality of data record segments, wherein the conditional probability distribution models dependencies between class labels and internal substructures of the plurality of data record segments;

determining optimal parameter values for the conditional probability distribution by applying a quasi-Newton gradient ascent method to training data, wherein the conditional probability distribution is restricted to a disjoint set of hidden states for each of the class labels; and

using the conditional probability distribution and the optimal parameter values to determine a most probable labeling sequence for the plurality of data record segments, wherein the plurality of data segments are applied to the conditional probability distribution to determine a plurality of marginal probabilities for each of the class labels.

8. The method of claim 7, wherein the conditional probability distribution is defined as

$$p(S|X) = \frac{1}{Z(X)} \exp \left(\sum_k \lambda_k \cdot \sum_{j=1}^T f_k(s_{j-1}, s_j, X, j) \right),$$

and wherein X is an observation sequence, Y is a potential labeling sequence, S is a vector of sub-structure variables, and λ_k is a confidence parameter.

9. The method of claim 8, wherein the quasi-Newton gradient ascent method is performed using a Gaussian prior defined as

$$L(\Lambda) = \sum_{i=1}^n \log P_{\Lambda}(Y_i|X_i) - \sum_{k=1}^K \frac{\lambda_k^2}{2\sigma^2},$$

and wherein Λ is a set of parameters that includes the confidence parameter.

10. The method of claim 9, wherein the plurality of marginal probabilities are summed according to the disjoint sets of hidden states to determine the most probably labeling sequence.

11. The method of claim 8, wherein the confidence parameters and a transition function f_k model dependencies between the class labels and the internal substructures.

12. A non-transitory machine-readable storage medium encoded with instructions executable by a processor for analyzing data using hidden dynamic systems, the machine-readable storage medium comprising instructions to:

define a conditional probability distribution for labeling a plurality of data record segments, wherein the conditional probability distribution models dependencies between class labels and internal substructures of the plurality of data record segments;

determine optimal parameter values for the conditional probability distribution by applying a quasi-Newton gradient ascent method to training data, wherein the conditional probability distribution is restricted to a disjoint set of hidden states for each of the class labels; and

use the conditional probability distribution and the optimal parameter values to determine a most probable labeling sequence for the plurality of data record segments, wherein the plurality of data segments are applied to the conditional probability distribution to determine a plurality of marginal probabilities for each of the class labels.

13. The non-transitory machine-readable storage medium of claim 12, wherein the conditional probability distribution is defined as

$$p(S|X) = \frac{1}{Z(X)} \exp \left(\sum_k \lambda_k \cdot \sum_{j=1}^T f_k(s_{j-1}, s_j, X, j) \right),$$

and wherein X is an observation sequence, Y is a potential labeling sequence, S is a vector of sub-structure variables, and λ_k is a confidence parameter.

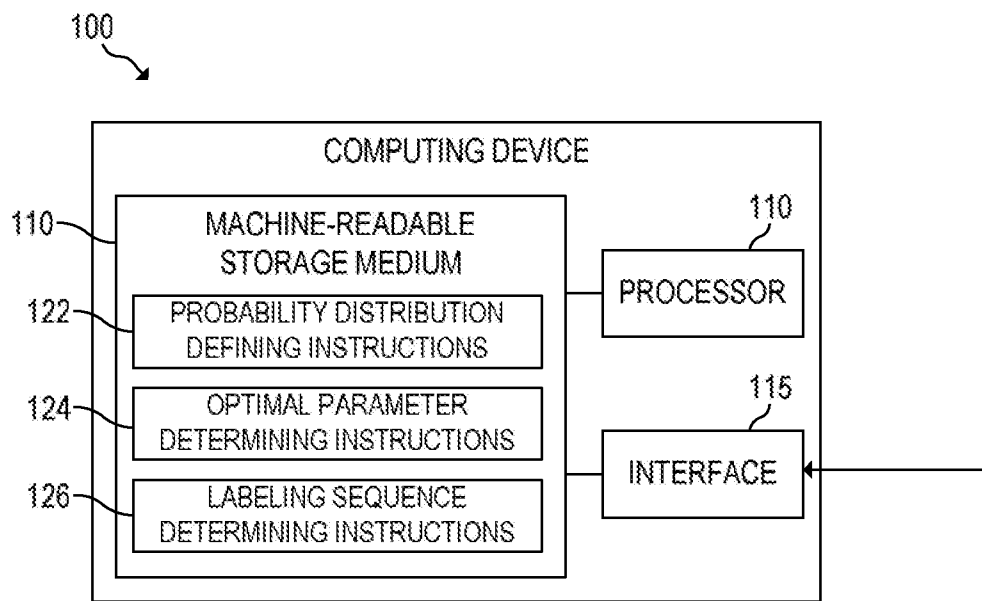
14. The non-transitory machine-readable storage medium of claim 13, wherein the quasi-Newton gradient ascent method is performed using a Gaussian prior defined as

$$L(\Lambda) = \sum_{i=1}^n \log P_{\Lambda}(Y_i|X_i) - \sum_{k=1}^K \frac{\lambda_k^2}{2\sigma^2},$$

and wherein Λ is a set of parameters that includes the confidence parameter.

15. The non-transitory machine-readable storage medium of claim 14, wherein the plurality of marginal probabilities are summed according to the disjoint sets of hidden states to determine the most probably labeling sequence.

1/4

**FIG. 1**

2/4

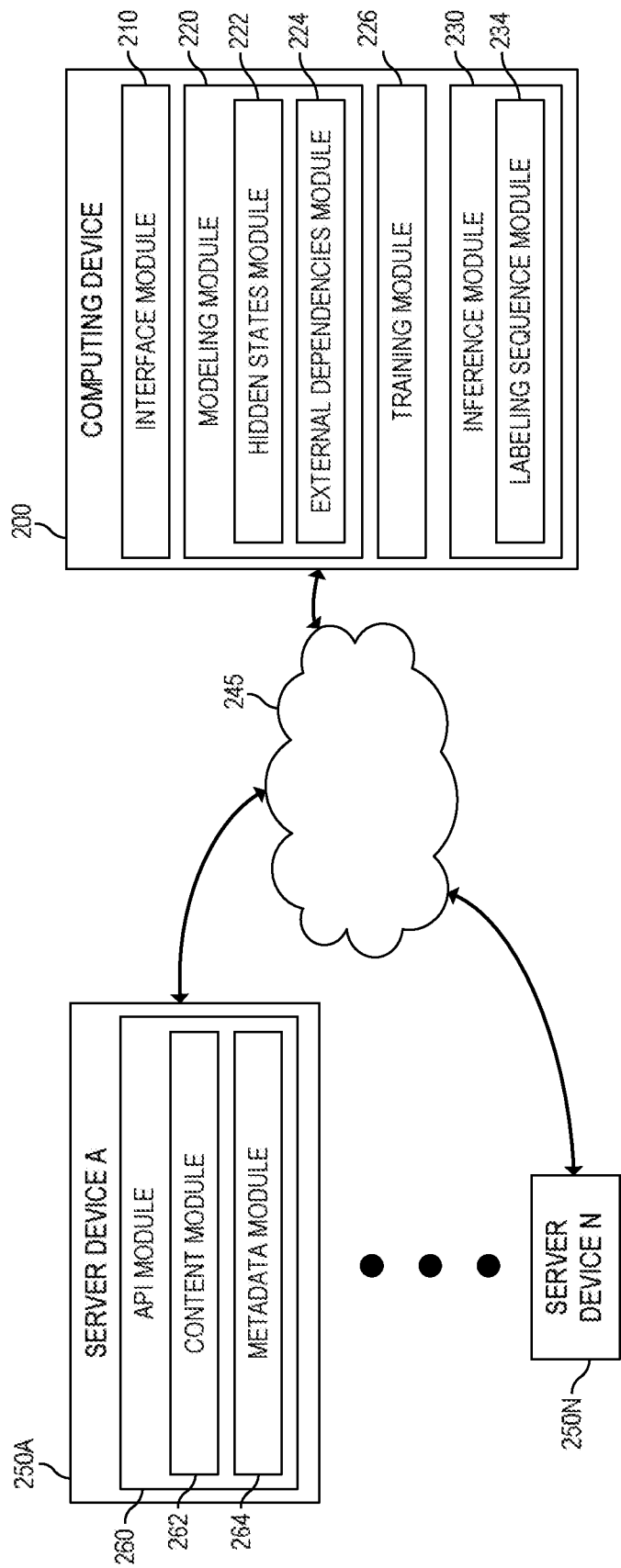
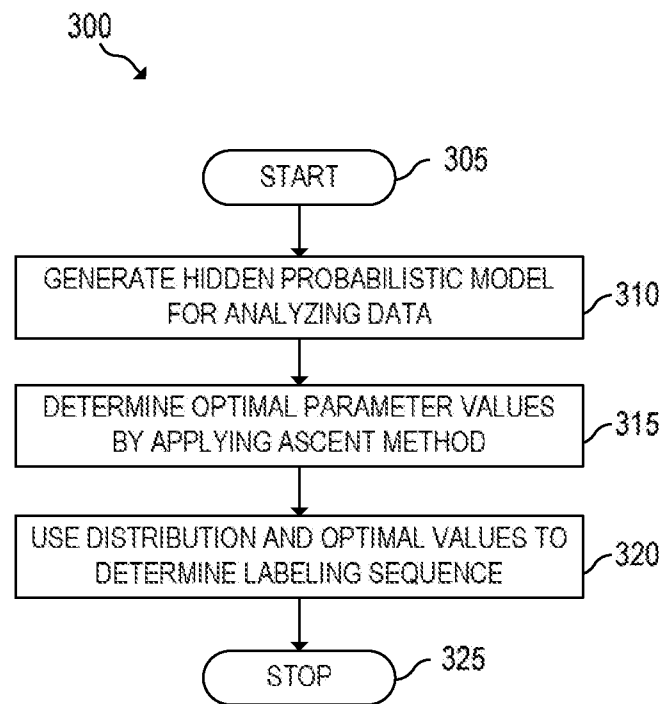


FIG. 2

3/4

**FIG. 3**

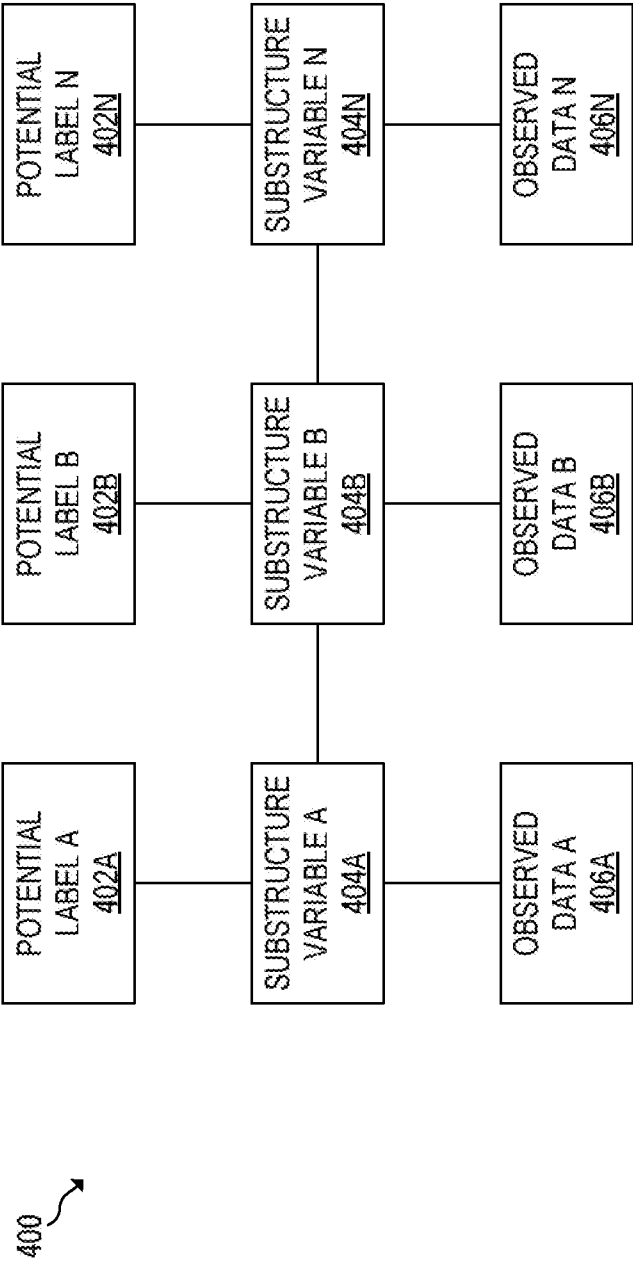


FIG. 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2015/076307

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/21(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,G06K,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)

CNABS;WPI;EPODOC;GOOGLE SCHOLAR:sequence?,distribution,conditional,hidden,record,analyz+,HDCRF?,nalis+,data,segment,probability,label+,class,dynamic

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 102681981 A (FUJITSU LTD.) 19 September 2012 (2012-09-19) paragraphs 0028-0036, 0077-0079, figures 1, 6	1-15
A	CN 102473174 A (ELEFONAKTIEBOLAGET LM ERICSSO) 23 May 2012 (2012-05-23) the whole document	1-15
A	US 2011043626 A1 (COBB, WESLEY KENNETH ET AL.) 24 February 2011 (2011-02-24) the whole document	1-15
A	CN 103198057 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 10 July 2013 (2013-07-10) 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

18 December 2015

Date of mailing of the international search report

15 January 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

YU,Liwen

Telephone No. (86-10)61648090

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2015/076307

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	102681981	A	19 September 2012	None			
CN	102473174	A	23 May 2012	WO	2011005155	A1	13 January 2011
				JP	2012533109	A	20 December 2012
				EP	2452273	A1	16 May 2012
				US	2012124607	A1	17 May 2012
				INCHENP	201108461	E	17 May 2013
US	2011043626	A1	24 February 2011	WO	2011022276	A2	24 February 2011
CN	103198057	A	10 July 2013	WO	2013102396	A1	11 July 2013
				EP	2801917	A1	12 November 2014
				JP	2015506515	A	02 March 2015
				KR	20140093762	A	28 July 2014
				US	2015019951	A1	15 January 2015