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

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
11 December 2014 (11.12.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/195782 A2

(51) International Patent Classification:
G06F 17/40 (2006.01)

(21) International Application Number:
PCT/IB2014/000939

(22) International Filing Date:
3 June 2014 (03.06.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
1938/MUM/2013 3 June 2013 (03.06.2013) IN

(71) Applicant: **TATA CONSULTANCY SERVICES LIMITED** [IN/IN]; Nirmal Building, 9th Floor, Nariman Point, 400 021 Mumbai (IN).

(72) Inventors: **CHAKRAVARTY, Kingshuk**; Plot A2, M2 & N2, Sector V, Block GP, Salt Lake, Electronics Complex, 700 091 Kolkata (IN). **DAS, Diptesh**; Plot A2, M2 & N2, Sector V, Block GP, Salt Lake, Electronics Complex, 700 091 Kolkata (IN). **SINHA, Aniruddha**; Plot A2, M2 & N2, Sector V, Block GP, Salt Lake, Electronics Complex, 700 091 Kolkata (IN). **KONAR, Amit**; Flat no. 2/5 (CHHAYANAT), HIG IIA, Phase-2, Calcutta Greens, Housing Complex, 1050/2 Survey Park, (Near Ajay Nagar), 700 075 Kolkata (IN).

(74) Agents: **RAE, Konpal** et al.; Lakshmikumaran & Sridharan, B6/10, Safdarjung Enclave, 110 029 New Delhi (IN).

(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, LT, 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, 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).

Declarations under Rule 4.17:

— *of inventorship (Rule 4.17(iv))*

[Continued on next page]

(54) Title: DIFFERENTIAL EVOLUTION-BASED FEATURE SELECTION

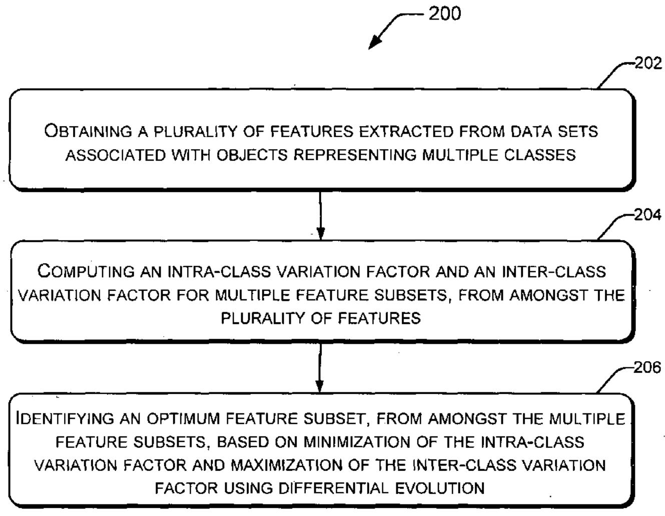
200
202
OBTAINING A PLURALITY OF FEATURES EXTRACTED FROM DATA SETS ASSOCIATED WITH OBJECTS REPRESENTING MULTIPLE CLASSES
204
COMPUTING AN INTRA-CLASS VARIATION FACTOR AND AN INTER-CLASS VARIATION FACTOR FOR MULTIPLE FEATURE SUBSETS, FROM AMONGST THE PLURALITY OF FEATURES
206
IDENTIFYING AN OPTIMUM FEATURE SUBSET, FROM AMONGST THE MULTIPLE FEATURE SUBSETS, BASED ON MINIMIZATION OF THE INTRA-CLASS VARIATION FACTOR AND MAXIMIZATION OF THE INTER-CLASS VARIATION FACTOR USING DIFFERENTIAL EVOLUTION

Figure 2

(57) Abstract: The subject matter discloses systems and methods for selection of an optimum feature subset. According to the present subject matter, the system (102) implements the described method, where the method includes obtaining a plurality of features extracted from data sets associated with objects representing multiple classes, computing an intra-class variation factor and an inter-class variation factor for multiple feature subsets, from amongst the plurality of features, and identifying an optimum feature subset, from amongst the multiple feature subsets, based on minimization of the intra-class variation factor and maximization of the inter-class variation factor using differential evolution.

WO 2014/195782 A2

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

DIFFERENTIAL EVOLUTION-BASED FEATURE SELECTION

TECHNICAL FIELD

[0001] The present subject matter relates, in general, to selection of features and particularly to selection of optimum features using differential evolution.

BACKGROUND

[0002] Objects, such as people, materials, diseases, etc., are generally identified and classified into distinct classes based on their types or characteristics. The identification and classification of the objects into the classes requires knowledge and information of object features which correlate with their types or characteristics. The known features can be used for the purposes of identification and classification of the objects. In some cases, there may be some correlated features or irrelevant features of objects. So, to identify and classify objects into their respective classes, a set of features which can distinguish the objects with respect to the classes is required. Such a set of features enables in distinct classification of different objects in distinct classes.

BRIEF DESCRIPTION OF DRAWINGS

[0003] The detailed description is described with reference to the accompanying figures. In the figures, the left-most digit(s) of a reference number identifies the figure in which the reference number first appears. The same numbers are used throughout the drawings to refer features and components.

[0004] Figure 1 illustrates a system environment implementing an optimum feature selection system, in accordance with an implementation of the present subject matter.

[0005] Figure 2 illustrates a method for selection of an optimum feature subset, in accordance with an implementation of the present subject matter.

[0006] Figure 3 illustrates a method for identification and classification of objects into classes using an optimum feature subset, in accordance with an implementation of the present subject matter.

[0007] It should be appreciated by those skilled in the art that any block diagrams herein represent conceptual views of illustrative systems embodying the principles of the present subject matter. Similarly, it will be appreciated that any flow charts, flow diagrams, state transition diagrams, pseudo code, and the like represent various processes which may be substantially represented in computer readable medium and so executed by a computing device or processor, whether or not such computing device or processor is explicitly shown.

DETAILED DESCRIPTION

[0008] The subject matter disclosed herein relates to system(s) and method(s) for selection of an optimum feature subset from a plurality of features. For the purposes of the present subject matter, feature selection refers to selection of a feature subset for identification and classification of objects into classes, and the optimum feature subset is a feature subset having a number of independent features, substantially sufficient for identification and classification of objects into classes.

[0009] The identification and classification of objects into the different classes using features that characterize the objects is known. Conventionally, for the identification and classification of objects into classes, data sets of the objects are gathered, and a plurality of features is extracted from the gathered data sets. In an example, to identify and classify people (objects) as distinct individuals (classes) based on their biometric characteristics, biometric data, such as skeleton data, is obtained for the individuals, and various gait features are extracted from the obtained biometric data. The plurality of features extracted is mapped to various possible classes, which is then used to train a supervised learning algorithm, also referred to as classifier, for subsequent identification and classification of unknown objects into the classes.

[0010] Generally, the number of features extracted from the data sets of the objects is substantially large. Some conventional classification methodologies utilize all the extracted features for the purpose of identification and classification of the objects. Such conventional methodologies thus require a large number of computational steps to identify and classify the objects, which makes them computationally expensive. Also, some of the extracted features may not be relevant or may be redundant for the classification of objects.

[0011] The extracted features which may not be relevant or may be redundant for the classification of objects, may contribute to misclassification of the objects. For this, a subset of features, from the set of extracted features, is selected. The selection of a subset of features using a classifier is also known. Conventionally, multiple random subsets of features are individually used in a classifier to identify an optimum feature subset, from amongst the subsets of features, which can identify and classify the objects. This optimum feature subset is then used to train the classifier to identify and classify the objects into the classes. Here also, only that classifier, which is trained using the optimum feature subset, can be used for the classification and identification of objects into classes. If another classifier is to be used then that classifier has to be trained using that or another optimum feature subset. Thus, the feature selection technique is classifier dependent.

[0012] The present subject matter describes system(s) and method(s) for selection of optimum feature subset from a plurality of extracted features. The selection of optimum feature subset, in accordance with the present subject matter, is classifier independent. For the selection of an optimum feature subset, a plurality of features extracted from data sets associated with objects representing multiple classes is obtained. The obtained features are analyzed and an optimum feature subset is selected based on differential evolution process.

[0013] The selection of optimum feature subset, in accordance with the present subject matter, is based on computation of an intra-class variation factor and an inter-class variation factor for a plurality of feature subsets. The intra-class variation factor refers to variations of individual or combination of features within a class. The inter-class variation factor refers to variations of individual or combination of features across multiple classes, i.e., variation of feature from one class with respect to another. In an implementation, for the selection of optimum feature subset, the intra-class variation factor is minimized and the inter-class variation factor is maximized using differential evolution process. The differential evolution process refers to an optimization search process which iteratively generates a solution (for example, a feature subset) to a problem (for example, an objective function, a fitness function, etc.,) with regard to a given condition (for example, minimization, maximization, etc.). By minimizing the intra-class variation factor for the optimum feature subset, it can be substantially ensured that the value of a particular feature, in the optimum feature subset, for a class lie in close proximity. Also, by maximizing the inter-class variation factor for the optimum feature subset, it is substantially ensured that the features, in the optimum feature subset, for each class are distinct with respect to the other classes.

[0014] The methodology of present subject matter can be implemented for selection of an optimum feature subset, to identify and classify objects into different classes using the optimum feature subset. With the optimum feature subset, the number of computations and size of storage space involved in the identification and classification stage is substantially less and the classification or recognition accuracy substantially improves. The usage of the optimum feature subset also substantially reduces the runtime complexity of identifying and classifying the objects into the classes.

[0015] In an example, the methodology of the present subject matter may be implemented for people identification, where the objects may be individuals who are to be classified as distinct individuals. In said example, the gait features, extracted from skeleton data sets at different instances for the individuals, are obtained, and an optimum gait feature subset is selected based on differential evolution process of the present subject matter. The optimum gait feature subset is then used in a classifier for the classification of the individuals.

[0016] In another example, the methodology of the present subject matter may be implemented for classification of cognitive loads on individuals, where the objects may be cognitive loads to be classified in different classes. In said example, the electroencephalography (EEG) features, extracted from EEG signals at different instances for the individuals, are obtained, and an optimum EEG feature subset is selected based on differential evolution process of the present subject matter. The optimum EEG feature subset is then used in a classifier for the classification of the cognitive loads on the individuals.

[0017] The selection of the optimum feature subset, in accordance with the present subject matter, does not involve a classifier and, thus, is independent of the classifier. This removes restrictions on the use of a particular classifier for which the optimum feature subset is obtained and which is trained to use the optimum feature subset for the identification and classification of the objects. Further, the optimum feature subset selected based on differential evolution by the minimization of intra-class variation factor and by the maximization of the inter-class variation factor is substantially accurate. With the optimum feature subset selection of the present subject matter, a substantially accurate identification and classification can be achieved.

[0018] The manner in which the system(s) and method(s) shall be implemented has been explained in details with respect to Figure 1 to Figure 3. Although the description herein is with reference to personal computer(s), the method(s) and system(s) may be implemented in other computing device(s) as well, albeit with a few variations, as will be understood by a person skilled in the art. While aspects of described methods can be implemented in any number of different computing devices, transmission environments, and/or configurations, the implementations are described in the context of the following computing device(s).

[0019] Figure 1 illustrates a system environment 100 implementing an optimum feature selection system 102, in accordance with an implementation of the present subject matter. For the purpose of description and simplicity, the optimum feature selection system 102 is hereinafter referred to as a system 102. The system 102 can be implemented as a computing device, such as a laptop computer, a desktop computer, a notebook, a workstation, a mainframe computer, and the like. The system 102 is enabled to select an optimum feature subset based on differential evolution process, in accordance with the present subject matter.

[0020] In an implementation, the system 102 includes processor(s) 104. The processor(s) 104 may be implemented as one or more microprocessors, microcomputers, microcontrollers, digital signal processors, central processing units, state machines, logic circuitries, and /or any devices that manipulate signals based on operational instructions. Among other capabilities, the processor(s) 104 is configured to fetch and execute computer-readable instructions stored in a memory.

[0021] The system 102 includes interface(s) 106. The interface(s) 106 may include a variety of machine readable instruction-based and hardware-based interfaces that allow the system 102 to communicate with other devices, including servers, data sources, and external repositories. Further, the interface(s) 106 may enable the system 102 to communicate with other communication devices, such as network entities, over a communication network.

[0022] Further, the system 102 includes a memory 108. The memory 108 may be coupled to the processor(s) 104. The memory 108 can include any computer-readable medium known in the art including, for example, volatile memory, such as static random access memory (SRAM) and dynamic random access memory (DRAM), and/or non-volatile memory, such as read only memory (ROM), erasable read only memory (ROM), erasable programmable ROM, flash memories, hard disks, optical disks, and magnetic tapes.

[0023] Further, the system 102 includes module(s) 110 and data 112. The module(s) 110 and the data 112 may be coupled to the processor(s) 104. The modules 110, amongst other things, include routines, programs, objects, components, data structures, etc., which perform particular tasks or implement particular abstract data types. The modules 110 may also be implemented as, signal processor(s), state machine(s), logic circuitries, and /or any other device or component that manipulate signals based on operational instructions. The data 112 serves, amongst other things, as a repository for storing data that may be fetched, processed, received, or generated by the module(s) 110. Although the data 112 is shown internal to the system 102, it may be understood that the data 112 can reside in an external repository (not shown in the Figure), which may be coupled to the system 102. The system 102 may communicate with the external repository through the interface(s) 106.

[0024] Further, the module(s) 110 can be implemented in hardware, as instructions executed by a processing unit, or by a combination thereof. The processing unit can comprise a computer, a processor, a state machine, a logic array or any other suitable devices capable of processing instructions. The processing unit can be a general-purpose processor which executes instructions to cause the general-purpose processor to perform the required tasks or, the processing unit can be dedicated to perform the required functions. In another aspect of the present subject matter, the module(s) 110 may be machine-readable instructions (software) which, when executed by a processor/processing unit, perform any of the desired functionalities. The machine-readable instructions may be stored on an electronic memory device, hard disk, optical disk or other machine-readable storage medium or non-transitory medium. In an implementation, the machine-readable instructions can also be downloaded to the storage medium via a network connection.

[0025] In an implementation, the module(s) 110 include a differential evolution feature selection (DEFS) module 114, and other module(s) 116. The other module(s) 116 may include programs or coded instructions that supplement applications or functions performed by the system 102. In said implementation, the data 112 includes feature data 120, fitness function data 122, optimum feature data 124, and other data 126. The other data 126 amongst other things, may serve as a repository for storing data that is processed, received, or generated as a result of the execution of one or more modules in the module(s) 110.

[0026] As shown in Figure 1, the system 102 is coupled to a data source 130 to obtain a plurality of features for the selection of an optimum feature subset. The data source 130 refers to an entity that has the data associated with the plurality of features extracted from data sets for multiple objects representing different classes. Further, as shown in Figure 1, the system 102 is coupled to a classifier 132 for classification of objects under the classes using the optimum feature subset. The classifier 132 may be trained for the optimum feature subset over different classes and, subsequently, used for the classification of unknown objects using the optimum feature subset.

[0027] For the purpose of selection of an optimum feature subset by the system 102, in an implementation, the DEFS module 114 obtains the plurality of features from the data source 130. As mentioned earlier, the features are extracted from data sets of objects representing multiple classes, taken at multiple instances of time. The data associated with plurality of features is stored in the feature data 120.

[0028] To illustrate the representation of the data associated with the plurality of features, obtained by the DEFS module 114, let us consider a case where p number of features is obtained for objects under 'c' number of classes. Let the i^{th} feature be denoted by d_i, and the i^{th} class be denoted by 'class i'. Further, as the features d_1 to d_p are obtained for each class, the data associated with the features d_1 to d_p for class 1 can be represented under the features as shown by equation (1):

$$class\ 1 = \begin{pmatrix} d_1 & d_2 & d_3 & & d_p \\ (x_1)_1^1 & (x_1)_1^2 & (x_1)_1^3 & & (x_1)_1^p \\ (x_2)_1^1 & (x_2)_1^2 & (x_2)_1^3 & \cdots & (x_2)_1^p \\ (x_3)_1^1 & (x_3)_1^2 & (x_3)_1^3 & & (x_3)_1^p \\ \vdots & \vdots & \vdots & & \vdots \\ (x_g)_1^1 & (x_g)_1^2 & (x_g)_1^3 & & (x_g)_1^p \end{pmatrix} \qquad \ldots\ldots\ldots\ldots\ldots (1)$$

where g is the size of data sets taken for class 1. Here, $(x_g)_1^p$ denotes the value of the feature d_p extracted at g^th instance for class 1.

[0029] Similarly, the data associated with the features d_l to d_p for class 2 can be represented under the features as shown by equation (2):

$$class\ 2 = \begin{pmatrix} d_1 & d_2 & d_3 & & d_p \\ (x_1)_2^1 & (x_1)_2^2 & (x_1)_1^3 & & (x_1)_2^p \\ (x_2)_2^1 & (x_2)_2^2 & (x_2)_2^3 & \cdots & (x_2)_2^p \\ (x_3)_2^1 & (x_3)_1^2 & (x_3)_1^3 & & (x_3)_2^p \\ \vdots & \vdots & \vdots & & \vdots \\ (x_w)_2^1 & (x_w)_2^2 & (x_w)_2^3 & & (x_w)_2^p \end{pmatrix} \qquad \ldots\ldots\ldots\ldots\ldots (2)$$

where w is the size of data sets taken for class 2. Here; $(x_w)_2^p$ denotes the value of the feature d_p extracted at w^th instance for class 2.

[0030] Similarly, the data associated with the features d_l to d_p for class c can be represented under the features as shown by equation (3):

$$class\ c = \begin{pmatrix} d_1 & d_2 & d_3 & & d_p \\ (x_1)_c^1 & (x_1)_c^2 & (x_1)_c^3 & & (x_1)_c^p \\ (x_2)_c^1 & (x_2)_c^2 & (x_2)_c^3 & \cdots & (x_2)_c^p \\ (x_3)_c^1 & (x_3)_c^2 & (x_3)_c^3 & & (x_3)_c^p \\ \vdots & \vdots & \vdots & & \vdots \\ (x_t)_c^1 & (x_t)_c^2 & (x_t)_c^3 & & (x_t)_c^p \end{pmatrix} \qquad \ldots\ldots\ldots\ldots\ldots (3)$$

where t is the size of data sets for class c. Here, $(x_t)_c^p$ denotes the value of the feature d_p extracted at t^th instance for class c.

[0031] In an implementation, after obtaining the plurality of features, the DEFS module 114 may normalize the values of each of the plurality of features to zero mean and unit covariance. With this, the values of the features are substantially scaled for subsequent processing.

[0032] Based on obtained features, the DEFS module 114 identifies an optimum feature subset based on a differential evolution process, such that the intra-class variation factor is minimum and the inter-class variation factor is maximum. The description below describes the procedure followed for identification of the optimum feature subset based on different evolution process.

[0033] In an implementation, for the identification of the optimum feature subset using differential evolution process, a population set comprising multiple parameter vectors for the different evolution process is formulated. Each of the parameter vectors comprises a feature subset and a Lagrange's multiplier λ. The feature subset represents and is indicative of features selected from amongst all the features obtained by the DEFS module 114. Each feature subset may have a set of features randomly selected from all of the obtained features. The Lagrange's multiplier λ is obtained from a range determined by a ratio of an inter-class variation factor and an intra-class variation factor of each of the features. The procedure of obtaining the Lagrange's multiplier λ is described later in the description.

[0034] In an implementation, each feature subset is in the form of a binary encoded decimal (BED) pattern indicative of those features which are selected to be a part of the feature subset. The BED pattern is of a size equal to the number of feature obtained by the DEFS module 114. In other words, the BED pattern is represented as a binary bit pattern, with the number of bits equal to the number of features obtained, where each bit corresponds to one feature and the value of the bits indicate the selection or the non-selection of the features in the feature subset. The 1's in the BED pattern represent the features which are selected to be the part of the feature subset and the 0's represent the features which are not selected to be the part of the feature subset. For p number of features, each feature subset is a BED pattern of p bits. In an example, for 13 number of features obtained by the DEFS module 114, the BED pattern for a feature subset may be '1011011001001'. This indicates that the features $\{d_1, d_3, d_4, d_6, d_7, d_{10}, d_{13}\}$ are selected to be the part of that feature subset.

[0035] With such representation of feature subsets, for p number of features, a total of 2^p number of BED patterns is possible. In an implementation, the population set for the differential evolution includes N number of feature subsets, where N is usually at least three times the number of the obtained features p, i.e., $N \geq 3^*p$. For example, if the total number of obtained features is 5, the total number of feature subsets N is at least equal to 15. N also denotes the number of parameter vectors in the population set.

[0036] To determine the Lagrange's multiplier λ for each parameter vector, at first a range of upper limits of the Lagrange's multiplier λ is determined. For this, the intra-class variation factor and inter-class variation factor for each of the features is computed. The intra-class variation factor of the each feature is divided by the inter-class variation factor of the same feature to obtain the upper limit of the Lagrange's multiplier λ for that feature. The upper limit of the Lagrange's multiplier λ for the j^{th} feature is given by equation (4) below:

$$\lambda_j = \frac{\mathrm{IntraVar}_j}{\mathrm{InterVar}_j} \quad \ldots\ldots\ldots\ldots\ldots (4)$$

where IntraVar$_j$ is the intra-class variation factor of the j^{th} feature and is given by the equation (5) below:

$$IntraVar_j = \sum_{i=1}^{c} \sum_{k=1}^{n} \sum_{m=k}^{n-1} \left| (x_k)_i^j - (x_{m+1})_i^j \right| \quad \ldots\ldots\ldots\ldots\ldots (5)$$

where k governs the data set at the k^{th} instance, i.e., k^{th} datapoint, i governs the class, c is total number of classes, n is size of data sets in class i, and j governs the feature for which the intra-class variation factor is to be calculated. InterVar$_j$ is the inter-class variation factor of the j^{th} feature and is given by the equation (6) below:

$$InterVar_j = \sum_{i=1}^{c} \sum_{s=1}^{c-1} \left| \frac{[m]_i^j}{[\sigma]_i^j} - \frac{[m]_{s+1}^j}{[\sigma]_{s+1}^j} \right| \quad \ldots\ldots\ldots\ldots\ldots (6)$$

where $[m]_i^j = \frac{\sum_{k=1}^{n} [x_k]_i^j}{n}$, ... (7)

$$[\sigma]_i^j = \sqrt{\frac{\sum_{k=1}^{n}\left([x_k]_i^j - [m]_i^j\right)^2}{n}} \quad .... (8)$$

where k governs the data set at the kth instance, i.e., kth datapoint, i governs the class, c is total number of classes, n is size of data sets in class i, and j governs the feature for which the inter-class variation factor is to be calculated. Similarly, the upper limits of the Lagrange's multiplier λ_1, λ_2 ... λ_p for all the p features are obtained.

[0037] In an implementation, the lower limits of the Lagrange's multiplier for the features are considered as significantly small values, let' say epsilon where epsilon is near equal to zero. After obtaining the range of upper limits and the range of lower limits for the Lagrange's multiplier, the Lagrange's multiplier for the parameter vectors are determined. In an implementation, the Lagrange's multiplier for each parameter vector is determined as a random value between the range of lower limits and the range of upper limits as obtained above.

[0038] In an implementation, the BED pattern for each of the parameter vectors is randomly generated initially. Consider an example with 4 features. The total number of feature subsets that can be represented using the BED pattern is 2^4-1, i.e., the range of the BED patterns is from '0001' to '1111'. The population set has at least 12 parameter vectors with the BED patterns randomly generated and selected from within the range of possible BED patterns.

[0039] In an implementation, the BED pattern for the each of the parameter vectors is uniformly randomly generated initially. Consider again the example with the 4 features. The 12 BED patterns for the population set are initially generated and selected randomly from within different ranges within '0001' and '1111'. In an example, the BED patterns selected uniformly may be from different ranges of '0001' to '0011', '0100' to '0111', '1000' to '1011' and '1100' to '1111'.

[0040] Further, a fitness function, denoted by J is formulated based on the intra-class variation factor, inter-class variation factor, and the Lagrange's multiplier. The fitness function J is given by equation (9):

$$J = \text{IntraVar} - \lambda^{*}\text{InterVar} \quad \ldots\ldots\ldots\ldots\ldots (9)$$

[0041] After formulating the population set with the parameter vectors having distinct feature subsets and the Lagrange's multiplier, intra-class variation factor and inter-class variation factor for the each of the feature subsets, in the population set, is computed to compute the value of fitness function. The intra-class variation factor for the each of the feature subset is computed using the values of features, as represented by equations (1) to (3), in equation (10) below:

$$IntraVar = \sum_{i=1}^{c}\sum_{j=1}^{p}\sum_{k=1}^{n}\sum_{m=k}^{n-1}\left|(x_k)_i^j - (x_{m+1})_i^j\right| \quad \ldots\ldots\ldots\ldots\ldots (10)$$

where k governs the data set at the kth instance, i.e., kth datapoint, i governs the class, c is total number of classes, n is size of the datasets in class i. Further, j governs the feature, and p is total number of features, where j belongs to those features which are selected in the feature subset for which the intra-class variation is to be computed.

[0042] The inter-class variation factor for the each of the feature subset is computed using the values of features, as represented by equations (1) to (3), in equation (11) below:

$$InterVar = \sum_{i=1}^{c}\sum_{j=1}^{p}\sum_{s=1}^{c-1}\left|\frac{[m)]_i^j}{[\sigma()]_i^j} - \frac{[m()]_{s+1}^j}{[\sigma]_{s+1}^j}\right| \quad \ldots\ldots\ldots\ldots\ldots (11)$$

$$\text{where } [m]_i^j = \frac{\sum_{k=1}^{n}(x_k)_i^j}{n}, \quad \ldots\ldots\ldots\ldots\ldots (12)$$

$$[\sigma]_i^j = \sqrt{\frac{\sum_{k=1}^{n}\left([x_k]_i^j - [m]_i^j\right)^2}{n}} \quad \ldots\ldots\ldots\ldots\ldots (13)$$

where k is governs the data set at the kth instance, i governs the class, c is total number of classes, n is total size of data sets in class i, j governs the feature, and p is total number of features. Here again j belongs to those features which are selected in the feature subset for which the inter-class variation is to be computed.

[0043] Using the Lagrange's multiplier and the above calculated intra-class variation factor and the inter-class variation factor in equation (9), the values of the fitness function are obtained for the parameter vectors in the population set.

[0044] Since each feature selected in optimum feature subset of an object in a class should have maximum amount of similarities, the intra-class variation factor of the optimum feature subset should be minimum. So the intra-class variation factor for the each of the feature subsets is to be minimized. Also, the optimum feature subset of objects in each two classes should have minimum amount of similarities, the intra-class variation factor of the optimum feature subset should be maximum. So the inter-class variation factor for the each of the feature subsets is to be maximized. Since the intra-class variation factor has to be minimized and the inter-class variation factor has to be maximized for each of the feature subsets, the fitness function J given by equation (9) has to be minimized for the feature subsets. The feature subset, from amongst the feature subsets, which has minimum value of the fitness function J is considered as the optimum feature subset. The data related to the optimum feature subset is stored in the optimum feature data 124.

[0045] The DEFS module 114 follows the differential evolution process for the minimization of the fitness function of the feature subsets and, thereby, the identification of the optimum feature subset. The differential evolution process involves four steps: initialization, mutation, recombination (also known as crossover), and selection.

[0046] In the initialization stage, a parameter vector, from amongst the parameter vectors in the population set, is selected as a target vector. Let the target vector be denoted by u_m, where m may be from 1 to the number of feature subsets N (or the number of parameter vectors). Let the BED pattern and the Lagrange's multiplier associated with the target vector u_m be denoted by BED_{um} and λ_{um}, respectively.

[0047] In the mutation stage, for the selected target vector u_m, three other parameter vectors u_p, u_q, and u_r are randomly selected from amongst the population set such that $p \neq q \neq r \neq m$. Based on the three selected parameter vectors, a donor vector v_m is generated by adding a weighted difference of any two vectors, from amongst the parameter vectors u_p, u_q, and u_r, to the remaining parameter vector as given by equation (14):

$$v_m = u_p + m_f{*}(u_q - u_r) \quad \ldots\ldots\ldots\ldots\ldots (14)$$

where m_f is mutation factor taking a value between 0 and 2. The mutation factor controls the rate of evolution of the population set. In an implementation, the mutation factor m_f is 0.9.

[0048] In the recombination or crossover stage, a trial vector t_m is generated, where each element of the trial vector t_m is selected from the elements of the target vector u_m or the donor vector v_m, depending on value of a cross-over (CR) ratio. The crossover ratio CR takes a value between 0 and 1. In an implementation, the crossover ratio CR is 0.7. The trial vector t_m is generated using equation (15) below:

$$t_{mj} = \begin{cases} v_{mj} & if\ rand\ (0,1) < CR \\ u_{mj} & otherwise \end{cases} \quad \ldots\ldots\ldots\ldots\ldots (15)$$

where t_{mj}, v_{mj}, and u_{mj} are the components (BED pattern and the Lagrange's multiplier λ) of the trial vector, donor vector, and target vector, respectively. Further, rand(0,1) is a random number generator that generates a random number between 0 and 1.

[0049] In the selection stage, the fitness function value of the trial vector t_m is calculated and compared with the fitness function value for the target vector u_m. If the fitness function value for the trial vector t_m is a lower than that for the target vector u_m, then the target vector u_m and its corresponding fitness function value are replaced by the trial vector t_m and its corresponding fitness value. Based on this revision, the target vector u_m, the Lagrange's multiplier λ_{tm} and the corresponding fitness function value are stored in the fitness function data 122.

[0050] The above procedure of mutation, recombination, and selection is iteratively repeated for all the parameter vectors as the target vectors in the population set. After all the parameter vectors of the population set are processed for the,mutation, recombination and selection, a new population set comprising the new set of target vectors as the parameter vectors and its corresponding fitness values are obtained. The differential evolution process is again performed on the new population set in a manner as described above.

[0051] The differential evolution process is continued until a stopping criterion is reached. In an implementation, the stopping criterion may be that the values of the fitness function J for the target vectors (or the parameter vectors) stops changing and has the minimum value. In an implementation, the differential evolution process may be performed for a predefined number of times.

[0052] Based on the values of fitness function corresponding to the each of the target vectors in the population set after the differential evolution process, the optimum feature subset is identified. For this, the values of fitness function for all parameter vectors of population set are compared with each other to identify that parameter vector for which the fitness function value is minimum. The BED pattern associated with that identified parameter vector is considered as the optimum feature subset. The DEFS module 114 selects the features in identified optimum feature subset as the optimum set of features. This optimum set of features is substantially sufficient for distinct identification and classification objects in different classes. The data related to the optimum feature subset is stored in the optimum feature data 124.

[0053] In an implementation, the DEFS module 114 provides the optimum feature subset to the classifier 132 for training the classifier 132 for identification and classification of the objects into the classes. In the implementation, the classifier 132 may include a supervised learning algorithm, such as a support vector machine, a naïve bayes, a decision tree, linear discriminate analysis, a neural network, and the like.

[0054] Although, as shown in Figure 1, the data source 130 that provides the plurality of features for the selection of the optimum feature subset, and the classifier 132 that receives the optimum feature subset from the system 102 for identification and classification of the objects into classes, reside outside the system 102; it may be understood by a person skilled in the art that the system 102 may obtaining the data sets for objects under the classes from devices such as a skeleton recording device, an EEG acquisition device, and the like, extract the plurality of features from the obtained data sets, identify to select the optimum feature subset, and then classify the objects into classes using a classifier. For this, in the implementation, the system 102 may have modules, such as a data acquisition module, a feature extraction module, the DEFS module 114, and a classification module, coupled to the processor(s) 104.

[0055] Figure 2 illustrates a method for selection of an optimum feature subset, in accordance with an implementation of the present subject matter. The method 200 can be implemented in the optimum feature selection system 102. The order in which the method 200 is described is not intended to be construed as a limitation, and any number of the described method blocks can be combined in any order to implement the method 200, or any alternative methods. Additionally, individual blocks may be deleted from the method 200 without departing from the spirit and scope of the subject matter described herein. Furthermore, the method 200 can be implemented in any suitable hardware.

[0056] The method 200 may be described in the general context of computer executable instructions. Generally, computer executable instructions can include routines, programs, objects, components, data structures, procedures, modules, functions, etc., that perform particular functions or implement particular abstract data types. Further, although the method 200 may be implemented in any computing device; in an example described in Figure 2, the method 200 is explained in context of the aforementioned optimum feature selection system 102, for the ease of explanation.

[0057] Referring to Figure 2, at block 202, a plurality of features extracted from data sets associated with objects representing multiple classes is obtained. The features are obtained by the system 102 from the data source 130. In an implementation, the data source 130 may obtain data sets for the objects representing multiple classes, and may extract the plurality of features from the obtained data sets.

[0058] Further, in an implementation, the values of the plurality of features are normalized to zero mean and unit covariance. After this, a population set comprising of parameter vectors is formulated for a differential evolution process. Each of the parameter vectors has a feature subset and a Lagrange's multiplier λ. The formulation of the population set is as described earlier in the description.

[0059] At block 204, an intra-class variation factor and an inter-class variation factor for multiple feature subsets, from amongst the plurality of features are computed. The multiple feature subsets are the feature subsets associated with the parameter vectors of the population set. In an implementation, the intra-class variation factor and the inter-class variation factor for the feature subsets associated with the parameter vectors in the population set are computed as described earlier in the description. Using the intra-class variation factor, the inter-class variation factor and the Lagrange's multiplier, the values of the fitness function is obtained for the parameter vectors in the population set.

[0060] Further, at block 206, the optimum feature subset is identified, from amongst the multiple feature subsets, based on minimization of the intra-class variation factor and the maximization of the inter-class variation factor using differential evolution. The multiple feature subsets are the feature subsets associated with the parameter vectors of the population set. In an implementation, the minimization of the intra-class variation factor and the maximization of the inter-class variation factor are done through the differential evolution process as described earlier in the description and the optimum feature subset is based on the feature subset having minimum value of the fitness function. The features in the identified optimum feature subset are selected as the optimum features for further processing.

[0061] In an implementation, the method 200, besides identifying the optimum feature subset, may include one of obtaining data sets for the objects representing multiple classes, extracting the plurality of features from the obtained data sets, classifying the objects into the classes based on the optimum feature subset, and a combination thereof.

[0062] Figure 3 illustrates a method 300 for identification and classification of objects into classes using an optimum feature subset, in accordance with an implementation of the present subject matter. The method 300 can be implemented in the optimum feature selection system 102. The order in which the method 300 is described is not intended to be construed as a limitation, and any number of the described method blocks can be combined in any order to implement the method 300, or any alternative methods. Additionally, individual blocks may be deleted from the method 300 without departing from the spirit and scope of the subject matter described herein. Furthermore, the method 300 can be implemented in any suitable hardware.

[0063] The method 300 may be described in the general context of computer executable instructions. Generally, computer executable instructions can include routines, programs, objects, components, data structures, procedures, modules, functions, etc., that perform particular functions or implement particular abstract data types.

[0064] Referring to Figure 3, at block 302, data sets for the objects representing multiple classes are obtained from a data acquisition device. The data acquisition device may be a skeleton recording device, an EEG acquisition device, and the like, depending on the application for which the method 300 is applied. Depending on the application, in an implementation, the data set may include skeleton points, of individuals, obtained using the skeleton recording device. In another implementation, the data sets may include EEG signals, of the individuals, obtained using the EEG acquisition device.

[0065] At block 304, a plurality of features is extracted from the data sets obtained at the block 302. In an implementation, depending on the application, the plurality of features may include area-related gait features of the object, dynamic centroid distance-related gait features of the object, angle-related gait features of the object, other static and dynamic gait features of the object and a combination thereof, or the plurality of features may include EEG features.

[0066] In an implementation, the values of the plurality of features are normalized to zero mean and unit covariance.

[0067] At block 306, an optimum feature subset is selected from amongst the plurality of features. The optimum feature subset is identified and selected based on minimization of intra-class variation factor and maximization of inter-class variation factor for multiple feature subsets through differential evolution process, described earlier in the description.

[0068] For this, a population set comprising of parameter vectors having feature subsets and Lagrange's multiplier λ is formulated for a differential evolution process, as described earlier in the description. In addition, a fitness function is formulated as described earlier in the description. After formulating the population set and the fitness function, an intra-class variation factor and an inter-class variation factor for the feature subsets associated with the parameter vectors in the population set are computed as described earlier in the description. Using the intra-class variation factor, the inter-class variation factor and the Lagrange's multiplier, the values of the fitness function is obtained for the parameter vectors in the population set. The differential evolution process is iteratively performed on the population set to minimize the intra-class variation factor and maximizing the inter-class variation factor for each of the feature sub-sets. The differential evolution process is iteratively carried out till a stopping criterion is reached as explained in the description earlier.

After this, the optimum feature subset is selected based on the feature subset having minimum value of the fitness function. The features in the identified optimum feature subset are selected as the optimum features for further processing.

[0069] At block 308, the objects are classified into classes based on the optimum feature subset. For this purpose, a classifier is used. In an implementation, the classifier may include a supervised learning algorithm, such as a support vector machine, a naïve bayes, a decision tree, linear discriminate analysis, a neural network, and the like.

[0070] Although implementations for system(s) and method(s) for optimum feature subset selection are described, it is to be understood that the present subject matter is not necessarily limited to the specific features or methods described. Rather, the specific features and methods are disclosed as implementations to select an optimum feature subset from a plurality of features.

I/We claim:

1. A computer-implemented method for differential evolution-based feature selection of an optimum feature subset from a plurality of features of objects for classification of the objects into multiple classes, the method comprising:

   obtaining the plurality of features extracted from data sets associated with the objects representing the multiple classes;

   computing, by a computing system, an intra-class variation factor and an inter-class variation factor for multiple feature subsets, from amongst the plurality of features; and

   identifying, by the computing system, the optimum feature subset, from amongst the multiple feature subsets, based on minimization of the intra-class variation factor and maximization of the inter-class variation factor using differential evolution.

2. The method as claimed in claim 1 further comprising formulating a fitness function based on the intra-class variation factor, the inter-class variation factor, and a Lagrange's multiplier.

3. The method as claimed in claim 2 further comprising formulating, by the computing system, a population set comprising parameter vectors for the differential evolution, wherein each of the parameter vectors has:

   a binary encoded decimal pattern corresponding to a feature subset, from amongst the multiple feature subsets, and

   a Lagrange's multiplier obtained from a range determined by a ratio of an inter-class variation factor and an intra-class variation factor of each of the features.

4. The method as claimed in claim 3, wherein the binary encoded decimal pattern is initially generated randomly.

5. The method as claimed in claim 3, wherein the binary encoded decimal pattern is initially generated uniformly randomly.

6. The method as claimed in claim 2, wherein the identifying of the optimum feature subset is based on the feature subset for which the corresponding fitness function has a minimum value.
7. The method as claimed in claim 1 further comprising classifying, by the computing system, the objects based on a classifier using the optimum feature subset, wherein the classifier is a learning algorithm comprising a support vector machine, a naïve bayes, a decision tree, linear discriminate analysis and a neural network.
8. The method as claimed in claim 1, wherein, for classification of individuals, the data sets are three-dimensional coordinates of skeleton points of each of the individuals, wherein the three-dimensional coordinates of skeleton points are obtained by a skeleton recording device; the plurality of features are gait features of the each of the individuals; and the each of the individuals is an object classified under a distinct class, from amongst the multiple classes.
9. The method as claimed in claim 1, wherein, for cognition load determination of individuals, the data sets are EEG signals obtained from an EEG acquisition device for each of the individuals; the plurality of features is Electroencephalography (EEG) features of the each of the individuals; and cognition load of the each of the individuals is classified under one of the multiple classes.
10. A system (102) for differential evolution-based feature selection of an optimum feature subset from a plurality of features of objects for classification of the objects into multiple classes, the system (102) comprising:

    a processor (104);

    a differential evolution feature selection (DEFS) module (114) coupled to the processor (104), to obtain the plurality of features extracted from data sets associated with the objects representing the multiple classes;

compute an intra-class variation factor and an inter-class variation factor for multiple feature subsets, from amongst the plurality of features; and

identify the optimum feature subset, from amongst the multiple feature subsets, based on minimization of the intra-class variation factor and maximization of the inter-class variation factor using differential evolution.

11. The system (102) as claimed in claim 10, wherein the DEFS module (114) formulates a fitness function based on the intra-class variation factor, the inter-class variation factor, and a Lagrange's multiplier.

12. The system (102) as claimed in claim 11, wherein the DEFS module (114) formulates a population set comprising parameter vectors for the differential evolution, wherein each of the parameter vectors has:

a binary encoded decimal pattern corresponding to a feature subset, from amongst the multiple feature subsets, and

a Lagrange's multiplier obtained from a range determined by a ratio of an inter-class variation factor and an intra-class variation factor of each of the features.

13. The system (102) as claimed in claim 12, wherein the binary encoded decimal pattern is initially generated randomly.

14. The system (102) as claimed in claim 11, wherein the DEFS module (114) minimizes the fitness function for identifying the optimum feature subset.

15. A non-transitory computer readable medium having a set of computer readable instructions that, when executed, cause a computing system to:

obtain a plurality of features extracted from data sets associated with objects representing multiple classes;

compute an intra-class variation factor and an inter-class variation factor for multiple feature subsets, from amongst the plurality of features; and

identify an optimum feature subset, from amongst the multiple feature subsets, based on minimization of the intra-class variation factor and maximization of the inter-class variation factor using differential evolution.

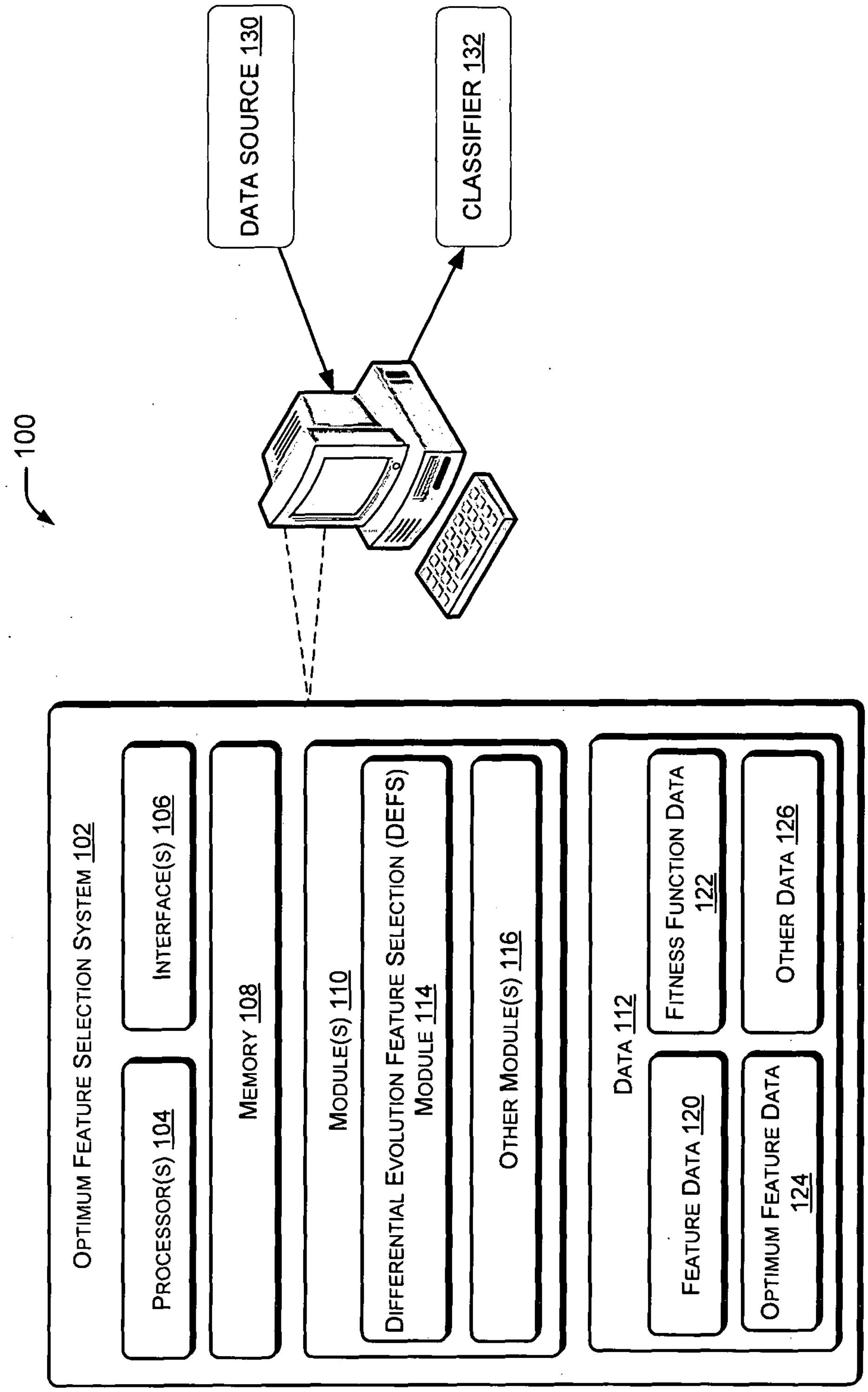
100
DATA SOURCE 130
CLASSIFIER 132
OPTIMUM FEATURE SELECTION SYSTEM 102
PROCESSOR(S) 104
INTERFACE(S) 106
MEMORY 108
MODULE(S) 110
DIFFERENTIAL EVOLUTION FEATURE SELECTION (DEFS) MODULE 114
OTHER MODULE(S) 116
DATA 112
FEATURE DATA 120
FITNESS FUNCTION DATA 122
OPTIMUM FEATURE DATA 124
OTHER DATA 126

Figure 1

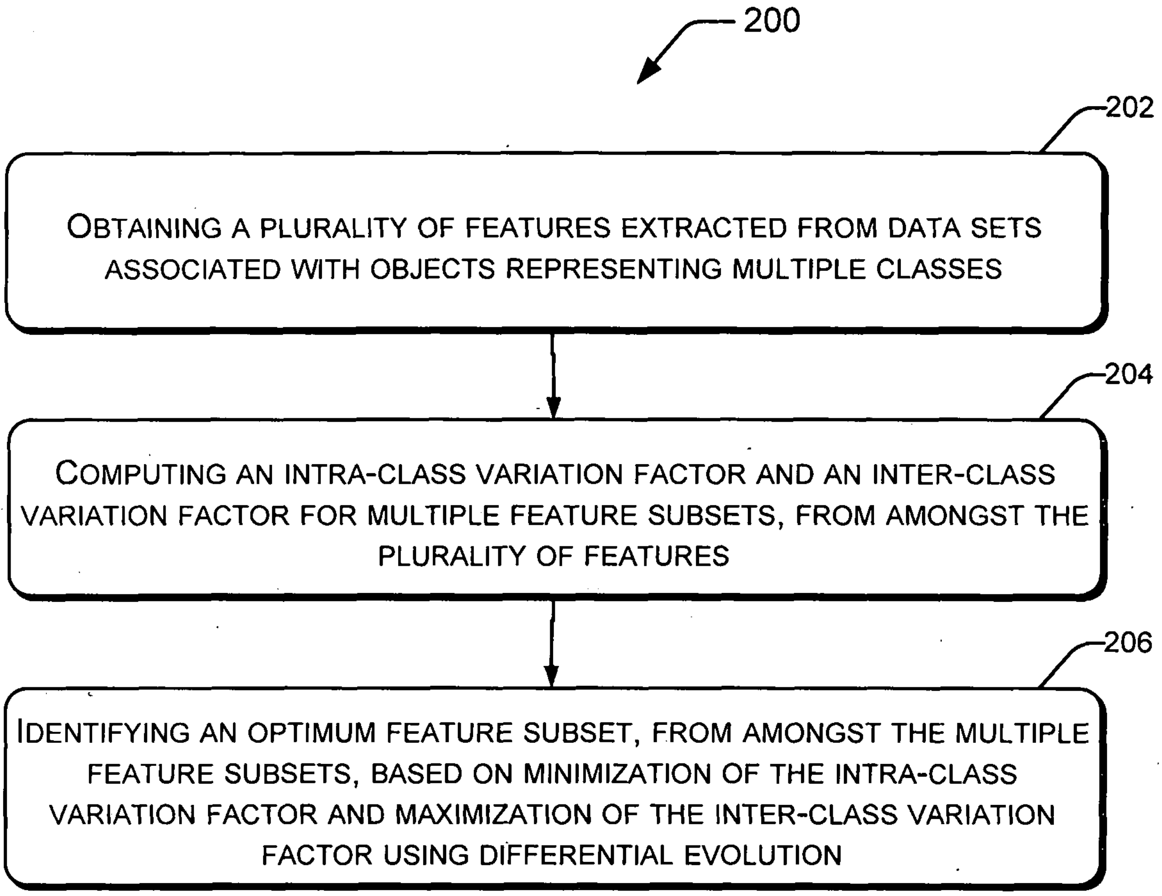
200
202
OBTAINING A PLURALITY OF FEATURES EXTRACTED FROM DATA SETS ASSOCIATED WITH OBJECTS REPRESENTING MULTIPLE CLASSES
204
COMPUTING AN INTRA-CLASS VARIATION FACTOR AND AN INTER-CLASS VARIATION FACTOR FOR MULTIPLE FEATURE SUBSETS, FROM AMONGST THE PLURALITY OF FEATURES
206
IDENTIFYING AN OPTIMUM FEATURE SUBSET, FROM AMONGST THE MULTIPLE FEATURE SUBSETS, BASED ON MINIMIZATION OF THE INTRA-CLASS VARIATION FACTOR AND MAXIMIZATION OF THE INTER-CLASS VARIATION FACTOR USING DIFFERENTIAL EVOLUTION

Figure 2

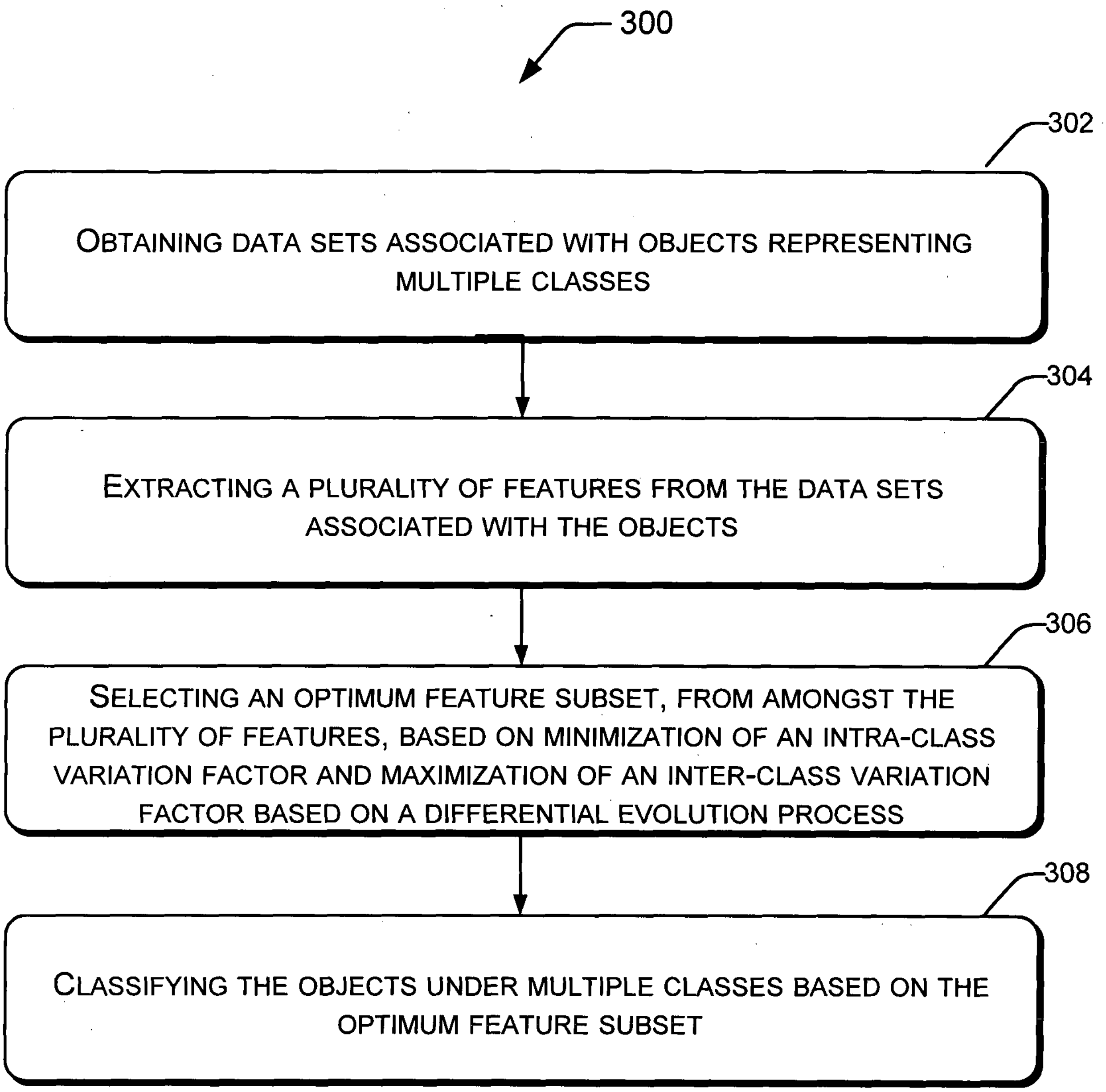
300
302
OBTAINING DATA SETS ASSOCIATED WITH OBJECTS REPRESENTING MULTIPLE CLASSES
304
EXTRACTING A PLURALITY OF FEATURES FROM THE DATA SETS ASSOCIATED WITH THE OBJECTS
306
SELECTING AN OPTIMUM FEATURE SUBSET, FROM AMONGST THE PLURALITY OF FEATURES, BASED ON MINIMIZATION OF AN INTRA-CLASS VARIATION FACTOR AND MAXIMIZATION OF AN INTER-CLASS VARIATION FACTOR BASED ON A DIFFERENTIAL EVOLUTION PROCESS
308
CLASSIFYING THE OBJECTS UNDER MULTIPLE CLASSES BASED ON THE OPTIMUM FEATURE SUBSET

Figure 3