

WHAT IS CLAIMED IS:

1. A data source-based service customizing device, comprising a memory, a processor and a data source-based service customizing system, which is stored on the memory and is operated on the processor, characterized in that the data source-based service customizing system is executed by the processor to implement the following steps:

S1, acquiring user generated contents in various predetermined data sources;

S2, recognizing the user generated contents by using a user group label recognition model generated by pre-training to recognize user group labels corresponding to the various data sources;

S3, determining group services corresponding to the various data sources according to a predetermined mapping relation between the user group labels and the group services, and sending the various data sources and the corresponding group services to a predetermined terminal to perform group service customization on the various data sources;

which is characterized in that the user group label recognition model is built on the basis of the maximum entropy criterion, a model function of the user group label recognition model comprises a user behavior attribute distribution function, and the user behavior attribute distribution function is as follows:

$$p\Lambda(y | x) = \frac{1}{Z(x)} \exp\left(\sum_{j=1}^F f_j(x, y) \lambda_{j,y}\right) ; x \text{ is a user, } y \text{ is a user label, } f_j(x, y) \text{ is the}$$

j^{th} behavior attribute of the user x , $\lambda_{j,y} = \{\lambda_{1,y}, \lambda_{2,y}, \lambda_{3,y}, \dots, \lambda_{F,y}\}$ is the probability distribution of the behavior attributes of the user, and $Z(x)$ is a normalization factor;

which is characterized in that before the data source-based service customizing system is executed by the processor to execute the step S1, the following steps are also implemented:

S01, acquiring user generated contents in a preset number of data sources, taking the acquired user generated content in each data source as a group data sample, and labeling each group data sample with a corresponding user group label;

S02, classifying the group data samples into a training set at a preset first proportion and a verification set at a preset second proportion, wherein the sum of the first proportion and the second proportion is less than or equal to 1;

S03, training the predetermined user group label recognition model by using the group data samples in the training set, and after the training, verifying the accuracy of the trained user group label recognition model by using the verification set;

S04, if the accuracy is greater than a preset threshold value, ending the model training, and taking the trained user group label recognition model as the user group label recognition model in the above-mentioned step S2, or if the accuracy is less than or equal to the preset threshold value, increasing the number of data sources, and retraining the user group label recognition model on the basis of the increased number of data sources.

2. The data source-based service customizing device according to claim 1, which is characterized in that the model function of the user group label recognition model also comprises an optimization function; the optimization function is as follows:

$$\arg \max(\sum_{i=1}^N p(x_i, y_i) \log p\Lambda(y_i | x_i) - \beta \| \Lambda \|_2^2)$$
; $p(x_i, y_i)$ is a probability that a label y_i of a user x_i appears in the user group labels, $\sum_{i=1}^N p(x_i, y_i) \log p\Lambda(y_i | x_i)$ is a likelihood function of the behavior attributes of a user, $\beta \| \Lambda \|_2^2$ is a normalization factor, and β is a parameter for controlling the normalization degree of the optimization function.

3. The data source-based service customizing device according to claim 1 or 2, which is characterized in that a recognition process of the user group label recognition model comprises solving iteration for the probability distribution of the behavior attributes of the users, and each iteration process also comprises differential mode iteration.

4. A data source-based service customizing method, comprising:

S1, acquiring user generated contents in various predetermined data sources;

S2, recognizing the user generated contents by using a pre-trained user group label recognition model to recognize user group labels corresponding to the various data sources;

S3, determining group services corresponding to the various data sources according to a predetermined mapping relation between the user group labels and the group services, and sending the various data sources and the corresponding group services to a predetermined terminal to perform group service customization on the various data sources;

which is characterized in that the user group label recognition model is built on the basis of the maximum entropy criterion, a model function of the user group label

recognition model comprises a user behavior attribute distribution function, and the user behavior attribute distribution function is as follows:

$$p\Lambda(y | x) = \frac{1}{Z(x)} \exp\left(\sum_{j=1}^F f_j(x, y) \lambda_{j,y}\right) ; x \text{ is a user, } y \text{ is a user label, } f_j(x, y) \text{ is the}$$

j^{th} behavior attribute of the user x , $\lambda_{j,y} = \{\lambda_{1,y}, \lambda_{2,y}, \lambda_{3,y}, \dots, \lambda_{F,y}\}$ is the probability distribution of the behavior attributes of the user, and $Z(x)$ is a normalization factor;

before the step S1, the data source-based service customizing method further comprising:

S01, acquiring user generated contents in a preset number of data sources, taking the acquired user generated content in each data source as a group data sample, and labeling each group data sample with a corresponding user group label;

S02, classifying the group data samples into a training set at a preset first proportion and a verification set at a preset second proportion, wherein the sum of the first proportion and the second proportion is less than or equal to 1;

S03, training the predetermined user group label recognition model by using the group data samples in the training set, and after the training, verifying the accuracy of the trained user group label recognition model by using the verification set;

S04, if the accuracy is greater than a preset threshold value, ending the model training, and taking the trained user group label recognition model as the user group label recognition model in the above-mentioned step S2, or if the accuracy is less than or equal to the preset threshold value, increasing the number of data sources, and retraining the user group label recognition model on the basis of the increased number of data sources.

5. The data source-based service customizing method according to claim 4, which is characterized in that the model function of the user group label recognition model also comprises an optimization function; the optimization function is as follows:

$$\arg \max \left(\sum_{i=1}^N p(x_i, y_i) \log p\Lambda(y_i | x_i) - \beta \| \Lambda \|_2^2 \right);$$
 $p(x_i, y_i)$ is a probability that a label y_i of a user x_i appears in the user group labels, $\sum_{i=1}^N p(x_i, y_i) \log p\Lambda(y_i | x_i)$ is a likelihood function of the behavior attributes of a user, $\beta \| \Lambda \|_2^2$ is a normalization factor, and β is a parameter for controlling the normalization degree of the optimization function.

6. The data source-based service customizing method according to claim 4 or 5, which is characterized in that a recognition process of the user group label recognition model comprises solving iteration for the probability distribution of the behavior attributes of the users, and each iteration process also comprises differential mode iteration.

7. A data source-based service customizing system, comprising:

an acquisition module, which is used for acquiring user generated contents in various predetermined data sources;

a recognition module, which is used for recognizing the user generated contents by using a pre-trained user group label recognition model to recognize user group labels corresponding to the various data sources;

a determination module, which is used for determining group services corresponding to the various data sources according to a predetermined mapping

relation between the user group labels and the group services, and sending the various data sources and the corresponding group services to a predetermined terminal to perform group service customization on the various data sources;

which is characterized in that the user group label recognition model is built on the basis of the maximum entropy criterion, a model function of the user group label recognition model comprises a user behavior attribute distribution function, and the user behavior attribute distribution function is as follows:

$$p\Lambda(y | x) = \frac{1}{Z(x)} \exp\left(\sum_{j=1}^F f_j(x, y) \lambda_{j,y}\right)$$

; x is a user, y is a user label, $f_j(x, y)$ is the j^{th} behavior attribute of the user x , $\lambda_{j,y} = \{\lambda_{1,y}, \lambda_{2,y}, \lambda_{3,y}, \dots, \lambda_{F,y}\}$ is the probability distribution of the behavior attributes of the user, and $Z(x)$ is a normalization factor;

which is characterized in that the data source-based service customizing system further comprises:

a training module, which is used for acquiring user generated contents in a preset number of data sources, taking the acquired user generated content in each data source as a group data sample, and labeling each group data sample with a corresponding user group label; classifying the group data samples into a training set at a preset first proportion and a verification set at a preset second proportion, wherein the sum of the first proportion and the second proportion is less than or equal to 1; training the predetermined user group label recognition model by using the group data samples in the training set, and after the training, verifying the accuracy of the trained user group label recognition model by using the verification set; and if the accuracy is greater than a preset threshold value, ending the model training, and taking the trained user group label recognition model as the user group label recognition model, or if the

accuracy is less than or equal to the preset threshold value, increasing the number of data sources, and retraining the user group label recognition model on the basis of the increased number of data sources.

8. The data source-based service customizing system according to claim 7, which is characterized in that the model function of the user group label recognition model also comprises an optimization function; the optimization function is as follows:

$$\arg \max \left(\sum_{i=1}^N p(x_i, y_i) \log p\Lambda(y_i | x_i) - \beta \| \Lambda \|_2^2 \right);$$
 $p(x_i, y_i)$ is a probability that a label y_i of a user x_i appears in the user group labels, $\sum_{i=1}^N p(x_i, y_i) \log p\Lambda(y_i | x_i)$ is a likelihood function of the behavior attributes of a user, $\beta \| \Lambda \|_2^2$ is a normalization factor, and β is a parameter for controlling the normalization degree of the optimization function.

9. The data source-based service customizing system according to claim 7 or 8, which is characterized in that a recognition process of the user group label recognition model comprises solving iteration for the probability distribution of the behavior attributes of the users, and each iteration process also comprises differential mode iteration.

10. A computer readable storage medium, which is characterized in that the computer readable storage medium stores a data source-based service customizing system, and the data source-based service customizing system is executed by a processor to implement the following steps:

S1, acquiring user generated contents in various predetermined data sources;

S2, recognizing the user generated contents by using a user group label recognition model generated by pre-training to recognize user group labels corresponding to the various data sources;

S3, determining group services corresponding to the various data sources according to a predetermined mapping relation between the user group labels and the group services, and sending the various data sources and the corresponding group services to a predetermined terminal to perform group service customization on the various data sources;

which is characterized in that the user group label recognition model is built on the basis of the maximum entropy criterion, a model function of the user group label recognition model comprises a user behavior attribute distribution function, and the user behavior attribute distribution function is as follows:

$$p\Lambda(y | x) = \frac{1}{Z(x)} \exp\left(\sum_{j=1}^F f_j(x, y) \lambda_{j,y}\right)$$

; x is a user, y is a user label, $f_j(x, y)$ is the j^{th} behavior attribute of the user x, $\lambda_{j,y} = \{\lambda_{1,y}, \lambda_{2,y}, \lambda_{3,y}, \dots, \lambda_{F,y}\}$ is the probability distribution of the behavior attributes of the user, and Z (x) is a normalization factor;

which is characterized in that before the data source-based service customizing system is executed by the processor to execute the step S1, the following steps are also implemented:

S01, acquiring user generated contents in a preset number of data sources, taking the acquired user generated content in each data source as a group data sample, and labeling each group data sample with a corresponding user group label;

S02, classifying the group data samples into a training set at a preset first proportion and a verification set at a preset second proportion, wherein the sum of the first proportion and the second proportion is less than or equal to 1;

S03, training the predetermined user group label recognition model by using the group data samples in the training set, and after the training, verifying the accuracy of the trained user group label recognition model by using the verification set;

S04, if the accuracy is greater than a preset threshold value, ending the model training, and taking the trained user group label recognition model as the user group label recognition model in the above-mentioned step S2, or if the accuracy is less than or equal to the preset threshold value, increasing the number of data sources, and retraining the user group label recognition model on the basis of the increased number of data sources.

11. The computer readable storage medium according to claim 10, which is characterized in that the model function of the user group label recognition model also comprises an optimization function; the optimization function is as follows:

$$\arg \max \left(\sum_{i=1}^N p(x_i, y_i) \log p \Lambda(y_i | x_i) - \beta \| \Lambda \|_2^2 \right);$$
 $p(x_i, y_i)$ is a probability that a label y_i of a user x_i appears in the user group labels, $\sum_{i=1}^N p(x_i, y_i) \log p \Lambda(y_i | x_i)$ is a likelihood function of the behavior attributes of a user, $\beta \| \Lambda \|_2^2$ is a normalization factor, and β is a parameter for controlling the normalization degree of the optimization function.

12. The computer readable storage medium according to claim 10 or 11, which is characterized in that a recognition process of the user group label recognition model

comprises solving iteration for the probability distribution of the behavior attributes of the users, and each iteration process also comprises differential mode iteration.