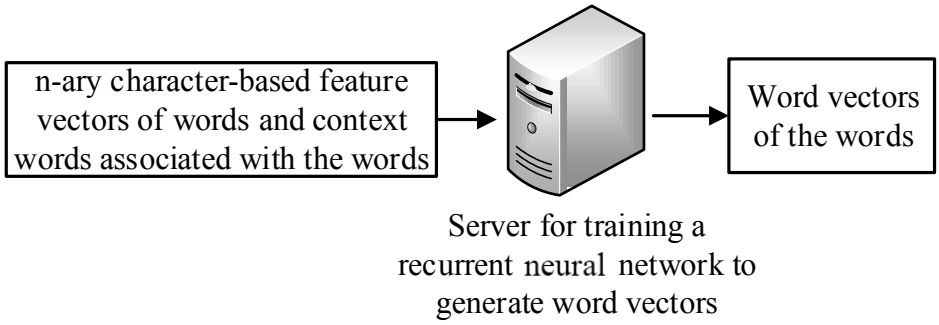

n-ary character-based feature vectors of words and context words associated with the words
Word vectors of the words
Server for training a recurrent neural network to generate word vectors

FIG. 1

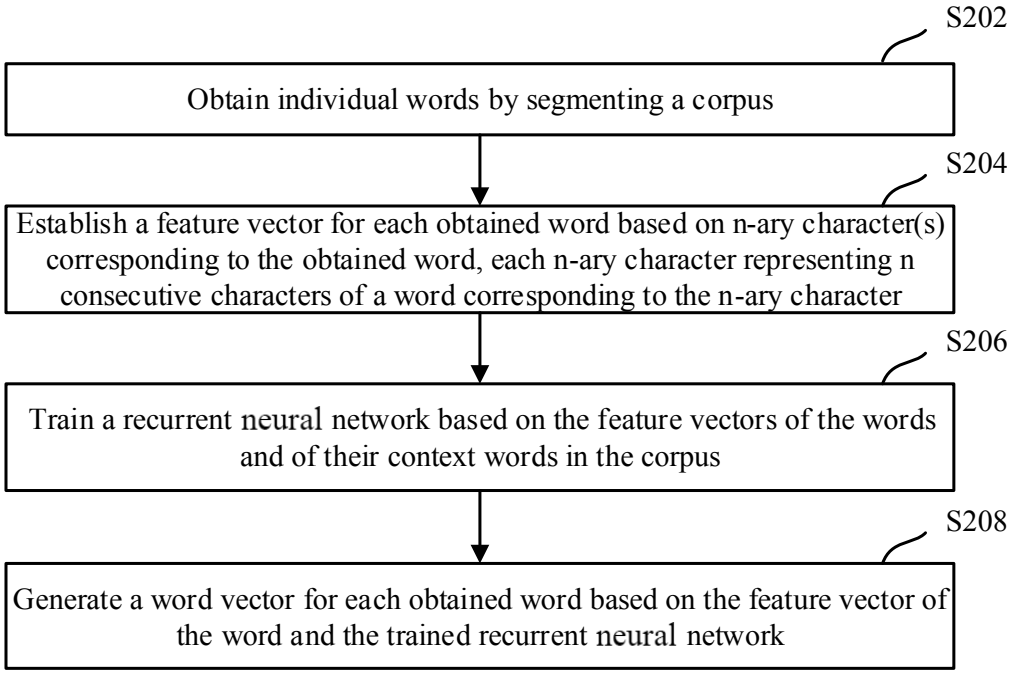

S202
Obtain individual words by segmenting a corpus
S204
Establish a feature vector for each obtained word based on n-ary character(s) corresponding to the obtained word, each n-ary character representing n consecutive characters of a word corresponding to the n-ary character
S206
Train a recurrent neural network based on the feature vectors of the words and of their context words in the corpus
S208
Generate a word vector for each obtained word based on the feature vector of the word and the trained recurrent neural network

FIG. 2

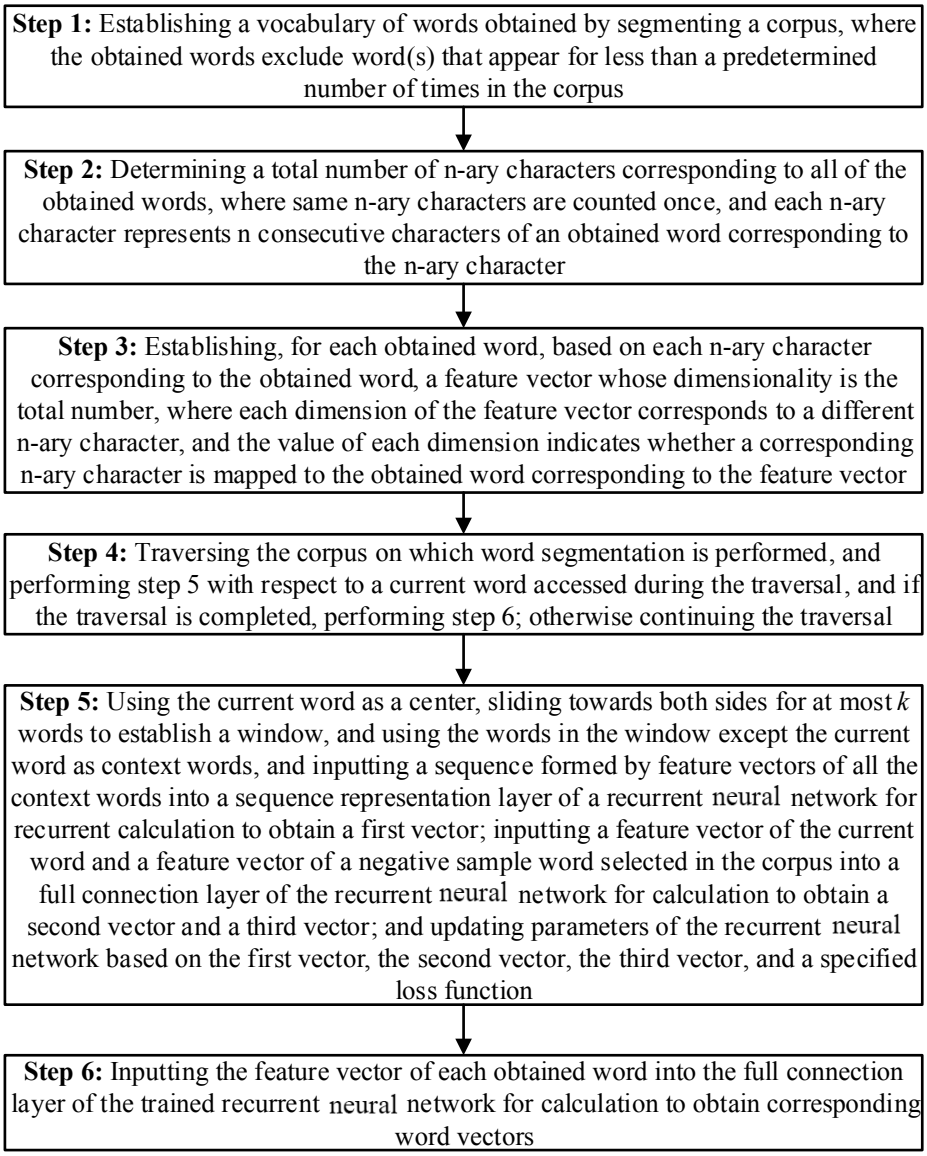

**Step 1:** Establishing a vocabulary of words obtained by segmenting a corpus, where the obtained words exclude word(s) that appear for less than a predetermined number of times in the corpus
**Step 2:** Determining a total number of n-ary characters corresponding to all of the obtained words, where same n-ary characters are counted once, and each n-ary character represents n consecutive characters of an obtained word corresponding to the n-ary character
**Step 3:** Establishing, for each obtained word, based on each n-ary character corresponding to the obtained word, a feature vector whose dimensionality is the total number, where each dimension of the feature vector corresponds to a different n-ary character, and the value of each dimension indicates whether a corresponding n-ary character is mapped to the obtained word corresponding to the feature vector
**Step 4:** Traversing the corpus on which word segmentation is performed, and performing step 5 with respect to a current word accessed during the traversal, and if the traversal is completed, performing step 6; otherwise continuing the traversal
**Step 5:** Using the current word as a center, sliding towards both sides for at most *k* words to establish a window, and using the words in the window except the current word as context words, and inputting a sequence formed by feature vectors of all the context words into a sequence representation layer of a recurrent neural network for recurrent calculation to obtain a first vector; inputting a feature vector of the current word and a feature vector of a negative sample word selected in the corpus into a full connection layer of the recurrent neural network for calculation to obtain a second vector and a third vector; and updating parameters of the recurrent neural network based on the first vector, the second vector, the third vector, and a specified loss function
**Step 6:** Inputting the feature vector of each obtained word into the full connection layer of the trained recurrent neural network for calculation to obtain corresponding word vectors

FIG. 6

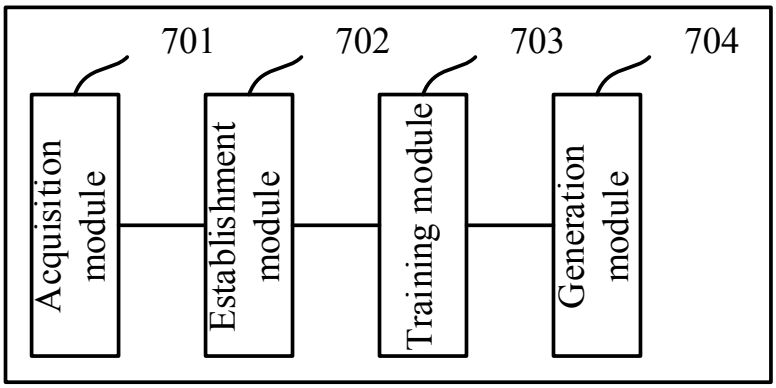

701
702
703
704
Acquisition module
Establishment module
Training module
Generation module

FIG. 7