CLAIMS

1. A method for knowledge extraction based on TextCNN, comprising:

S10, collecting first training data, and constructing a character vector dictionary and a word vector dictionary;

S20, constructing a first convolutional neural network, and training the first convolutional neural network based on a first optimization algorithm, wherein the first convolutional neural network comprises a first embedding layer, a first multilayer convolution, and a first softmax function, wherein the first embedding layer, the first multilayer convolution, and the first softmax function are connected in turn;

S21, collecting second training data, wherein the second training data is pre-labeled data, comprising a named entity position label and a named entity relationship label, and inputting the second training data into the first embedding layer after the second training data is divided into multiple characters and special symbols are removed;

S22, performing character vector matching on the second training data with character-level, based on the character vector dictionary, in the first embedding layer, to convert the second training data into a matrix form;

S23, wherein, the first multilayer convolution performs a convolution operation on a matrix output from the first embedding layer, the first multilayer convolution comprises one group of first-type one-dimensional convolution layer located at a front portion, and at least one group of second-type one-dimensional convolution layers located at a rear portion, the first-type one-dimensional convolution layers comprise one-dimensional convolution kernels with different number of lengths and the same number of channels, the second-type one-dimensional convolution layers comprise one-dimensional convolution kernels with the same number of lengths and the same number of channels, data of each convolution layer comes from an output of previous convolution layer, and number of matrix rows is maintained during the convolution operation;

S24, outputting the first multi-layer convolution via the first softmax function to determine prediction probabilities of BEMO tagging with a plurality of subdivision categories of each character;

S25, training the first convolutional neural network, calculating a first cross-entropy of a loss function according to the prediction probability of BEMO tagging and a true BEMO label of the second training data, and minimizing the loss function through the first optimization algorithm to train the first convolutional neural network;

S30, constructing a second convolutional neural network, and training the second convolutional neural network based on a second optimization algorithm, wherein the second convolutional neural network comprises a second embedding layer, a second multilayer convolution, a pooling layer, two fully-connected layers and a second softmax function, wherein the second embedding layer, the second multilayer convolution, the pooling layer, two fully-connected layers and the second softmax function are connected in turn, the two fully-

connected layers are the same;

S31, segmenting the second training data into words, comprising

S311, performing a preliminary word segmentation on the second training data using a jieba library, and correcting the preliminary word segmentation based on a predicted word segmentation of the first convolutional neural network, if the preliminary word segmentation is different from the predicted word segmentation of the first convolutional neural network, taking the predicted word segmentation of the first convolutional neural network as standard;

S312, inputting the second training data into the second embedding layer after special symbols and non-Chinese characters in the preliminary word segmentation are removed;

S32, performing word vector matching on the second training data after segmenting, based on the word vector dictionary, in the second embedding layer, to convert the second training data into the matrix form;

S33, wherein, the second multilayer convolution performs the convolution operation on the matrix output from the second embedding layer, the second multilayer convolution comprises one group of first-type one-dimensional convolution layer located at a front portion, and at least one group of second-type one-dimensional convolution layers located at a rear portion, the first-type one-dimensional convolution layer comprises one-dimensional convolution kernels with different number of lengths and the same number of channels, the second-type one-dimensional convolution layers comprise one-dimensional convolution kernels with the same number of lengths and the same number of channels, data of each convolution layer comes from an output of previous convolution layer, and number of matrix rows is maintained during the convolution operation;

S34, inputting an output of the second multilayer convolution into the pooling layer for compression;

S35, inputting an output of the pooling layer into two fully-connected layers to perform information fusion of each channel;

S36, inputting an output of the fully-connected layers into the second softmax function to determine corresponding prediction probabilities of multiple entity relationship labels;

S37, training the second convolutional neural network, calculating a second cross-entropy of the loss function according to predicted probability of a relationship label output by the second convolutional neural network and a true relationship label of the second training data, and minimizing the loss function through the second optimization algorithm to train the second convolutional neural network;

S40, inputting to-be-predicted data into the first convolutional neural network after training and the second convolutional neural network after training,

extracting a knowledge graph triple of the to-be-predicted data according to an entity tagging prediction output by the first convolutional neural network after training and an entity relationship prediction output by the second convolutional neural network after training, wherein a class corresponding to the maximum probability value in the prediction probabilities of BEMO tagging is selected as the entity tagging prediction output by the first convolutional

neural network, a class corresponding to a prediction probability value which is greater than 0.5 is selected as the entity relationship prediction output by the second convolutional neural network, in order to extract the knowledge graph triplet of the to-be-predicted data.

2. The method as in Claim 1, wherein S10 comprises:

S11, inputting the first training data collected into a Word2Vec algorithm for training to obtain character vectors and construct the character vector dictionary, after the first training data collected is divided into characters and special symbols are removed and non-Chinese characters are removed;

S12, at the same time, inputting the first training data collected into the Word2Vec algorithm for training to obtain word vectors and construct the word vector dictionary, after the first training data collected is segmented into words and special symbols are removed and non-Chinese characters are removed.

3. The method as in Claim 2, wherein the Word2Vec algorithm for training is implemented by a gensim library in Python.

4. The method as in Claim 1, wherein the first convolutional neural network and the second convolutional neural network are constructed based on a tensorflow library in Python.

5. The method as in Claim 1, wherein the first multilayer convolution comprises 5 convolution layers, wherein the first-type one-dimensional convolution layer comprises one-dimensional convolution kernels corresponding to three types of length, the one-dimensional convolution kernels of each length corresponds to 128 channels, four groups of the second-type one-dimensional convolution layers comprise one-dimensional convolution kernels of a length of 3, and number of channels of the one-dimensional convolution kernels corresponding to the four groups of the second-type one-dimensional convolution layers is 384;

and/or, the second multilayer convolution comprises 3 convolution layers, wherein the first-type one-dimensional convolution layer comprises one-dimensional convolution kernels corresponding to three types of length, the one-dimensional convolution kernels of each length corresponds to 128 channels, two groups of the second-type one-dimensional convolution layers comprise one-dimensional convolution kernels of a length of 3, and number of channels of the one-dimensional convolution kernels corresponding to the two groups of the second-type one-dimensional convolution layers is 384.

6. The method as in Claim 1, wherein the first optimization algorithm and the second optimization algorithm are an ADAM algorithm or an rmsprop algorithm.

7. The method as in Claim 1, in the S40, if the entity tagging prediction and the entity relationship prediction contradict with each other, knowledge graph triplet extraction of the to-be-predicted data is abandoned.

8. An equipment for knowledge extraction based on TextCNN, comprising:

a character vector dictionary construction module, configured to construct a character vector dictionary according to a first training data collected;

a word vector dictionary construction module, configured to construct a word vector dictionary according to the first training data collected;

a first convolutional neural network construction and training module, configured to construct a first convolutional neural network and train the first convolutional neural network based on a first optimization algorithm, wherein the first convolutional neural network comprises a first embedding layer, a first multilayer convolution, and a first softmax function, wherein the first embedding layer, the first multilayer convolution, and the first softmax function are connected in turn;

wherein the first convolutional neural network construction and training module comprises:

a character vector pre-processing unit, second training data is pre-labeled data, comprising a named entity position label and a named entity relationship label, configured to input the second training data into the first embedding layer after the second training data is divided into multiple characters and special symbols are removed;

a character vector matrix unit, configured to perform character vector matching on the second training data with character-level, based on the character vector dictionary, in the first embedding layer, to convert the second training data into a matrix form;

a first multilayer convolution unit, configured to perform a convolution operation on a matrix output from the first embedding layer, the first multilayer convolution comprises one group of first-type one-dimensional convolution layer located at a front portion, and at least one group of second-type one-dimensional convolution layers located at a rear portion, the first-type one-dimensional convolution layers comprise one-dimensional convolution kernels with different number of lengths and the same number of channels, the second-type one-dimensional convolution layers comprise one-dimensional convolution kernels with the same number of lengths and the same number of channels, data of each convolution layer comes from an output of previous convolution layer, and number of matrix rows is maintained during the convolution operation;

a first softmax function output unit, configured to output the first multi-layer convolution via the first softmax function to determine prediction probabilities of BEMO tagging with a plurality of subdivision categories of each character;

a first convolutional neural network training unit, configured to calculate a first cross-entropy of a loss function according to the prediction probability of BEMO tagging and a true BEMO label of the second training data, and minimize the loss function through the first optimization algorithm to train the first convolutional neural network;

a second convolutional neural network construction and training module, configured to train a second convolutional neural network based on a second optimization algorithm, wherein

the second convolutional neural network comprises a second embedding layer, a second multilayer convolution, a pooling layer, two fully-connected layers and a second softmax function, wherein the second embedding layer, the second multilayer convolution, the pooling layer, two fully-connected layers and the second softmax function are connected in turn, the two fully-connected layers are the same;

wherein the second convolutional neural network construction and training module comprises:

a word vector pre-processing unit, configured to segment the second training data into words;

wherein the word vector pre-processing unit comprises:

a preliminary word segmentation subunit, configured to perform preliminary word segmentation on the second training data using a jieba library, and correct the preliminary word segmentation based on a predicted word segmentation of the first convolutional neural network, if the preliminary word segmentation is different from the predicted word segmentation of the first convolutional neural network, take the predicted word segmentation of the first convolutional neural network as standard;

a word segmentation pre-processing subunit, configured to input the second training data into the second embedding layer after special symbols and non-Chinese characters in the preliminary word segmentation are removed;

a word vector matrix unit, configured to perform word vector matching on the second training data after segmenting, based on the word vector dictionary, in the second embedding layer, to convert the second training data into the matrix form;

a second multilayer convolution unit, configured to perform the convolution operation on the matrix output from the second embedding layer, the second multilayer convolution unit comprises one group of first-type one-dimensional convolution layer located at a front portion, and at least one group of second-type one-dimensional convolution layers located at a rear portion, the first-type one-dimensional convolution layer comprises one-dimensional convolution kernels with different number of lengths and the same number of channels, the second-type one-dimensional convolution layers comprise one-dimensional convolution kernels with the same number of lengths and the same number of channels, data of each convolution layer comes from an output of previous convolution layer, and number of matrix rows is maintained during the convolution operation;

the pooling layer, configured to input an output of the second multilayer convolution to the pooling layer for compression;

the fully-connected layer, configured to input an output of the pooling layer into two fully-connected layers to perform information fusion of each channel;

a second softmax function output unit, configured to input an output of the fully-connected layers into the second softmax function to determine corresponding prediction probabilities of multiple entity relationship labels;

a second convolutional neural network training unit, configured to calculate a second

cross-entropy of the loss function according to predicted probability of a relationship label output by the second convolutional neural network and a true relationship label of the second training data, and minimize the loss function through the second optimization algorithm to train the second convolutional neural network;

a knowledge graph triple extraction module, configured to input to-be-predicted data into the first convolutional neural network after training and the second convolutional neural network after training,

extract a knowledge graph triple of the to-be-predicted data according to an entity tagging prediction output by the first convolutional neural network after training and an entity relationship prediction output by the second convolutional neural network after training, wherein a class corresponding to the maximum probability value in the prediction probabilities of BEMO tagging is selected as the entity tagging prediction output by the first convolutional neural network, a class corresponding to a prediction probability value which is greater than 0.5 is selected as the entity relationship prediction output by the second convolutional neural network, in order to extract the knowledge graph triplet of the to-be-predicted data.

9. The equipment as in Claim 8, wherein the character vector dictionary construction module, configured to input the first training data collected into a Word2Vec algorithm for training to obtain character vectors and construct the character vector dictionary after the first training data collected is divided into characters and special symbols are removed and non-Chinese characters are removed;

10. The equipment as in Claim 9, wherein the character vector dictionary construction module, configured to input the first training data collected into the Word2Vec algorithm for training to obtain word vectors and construct the word vector dictionary after the first training data collected is segmented into words and special symbols are removed and non-Chinese characters are removed.

11. The equipment as in Claim 8, wherein the first multilayer convolution comprises 5 convolution layers, wherein the first-type one-dimensional convolution layer comprises one-dimensional convolution kernels corresponding to three types of length, the one-dimensional convolution kernels of each length corresponds to 128 channels, four groups of the second-type one-dimensional convolution layers comprise one-dimensional convolution kernels of a length of 3, and number of channels of the one-dimensional convolution kernels corresponding to the four groups of the second-type one-dimensional convolution layers is 384;

and/or, the second multilayer convolution comprises 3 convolution layers, wherein the first-type one-dimensional convolution layer comprises one-dimensional convolution kernels corresponding to three types of length, the one-dimensional convolution kernels of each length are corresponds to 128 channels, two groups of the second-type one-dimensional convolution layers comprise one-dimensional convolution kernels of a length of 3, and number of channels of the one-dimensional convolution kernels corresponding to the two groups of the second-

type one-dimensional convolution layers is 384.

12. A computing device comprising a memory, a processor, and computer program stored in the memory and executable on the processor, wherein the processor executes the computer programs for implementing steps of a method for knowledge extraction based on TextCNN, the method comprising:

S10, collecting first training data, and constructing a character vector dictionary and a word vector dictionary;

S20, constructing a first convolutional neural network, and training the first convolutional neural network based on a first optimization algorithm, wherein the first convolutional neural network comprises a first embedding layer, a first multilayer convolution, and a first softmax function, wherein the first embedding layer, the first multilayer convolution, and the first softmax function are connected in turn;

S21, collecting second training data, wherein the second training data is pre-labeled data, comprising a named entity position label and a named entity relationship label, and inputting the second training data into the first embedding layer after the second training data is divided into multiple characters and special symbols are removed;

S22, performing character vector matching on the second training data with character-level, based on the character vector dictionary, in the first embedding layer, to convert the second training data into a matrix form;

S23, wherein, the first multilayer convolution performs a convolution operation on a matrix output from the first embedding layer, the first multilayer convolution comprises one group of first-type one-dimensional convolution layer located at a front portion, and at least one group of second-type one-dimensional convolution layers located at a rear portion, the first-type one-dimensional convolution layers comprise one-dimensional convolution kernels with different number of lengths and the same number of channels, the second-type one-dimensional convolution layers comprise one-dimensional convolution kernels with the same number of lengths and the same number of channels, data of each convolution layer comes from an output of previous convolution layer, and number of matrix rows is maintained during the convolution operation;

S24, outputting the first multi-layer convolution via the first softmax function to determine prediction probabilities of BEMO tagging with a plurality of subdivision categories of each character;

S25, training the first convolutional neural network, calculating a first cross-entropy of a loss function according to the prediction probability of BEMO tagging and a true BEMO label of the second training data, and minimizing the loss function through the first optimization algorithm to train the first convolutional neural network;

S30, constructing a second convolutional neural network, and training the second convolutional neural network based on a second optimization algorithm, wherein the second convolutional neural network comprises a second embedding layer, a second multilayer

convolution, a pooling layer, two fully-connected layers and a second softmax function, wherein the second embedding layer, the second multilayer convolution, the pooling layer, two fully-connected layers and the second softmax function are connected in turn, the two fully-connected layers are the same;

S31, segmenting the second training data into words, comprising:

S311, performing a preliminary word segmentation on the second training data using a jieba library, and correcting the preliminary word segmentation based on a predicted word segmentation of the first convolutional neural network, if the preliminary word segmentation is different from the predicted word segmentation of the first convolutional neural network, taking the predicted word segmentation of the first convolutional neural network as standard;

S312, inputting the second training data into the second embedding layer after special symbols and non-Chinese characters in the preliminary word segmentation are removed;

S32, performing word vector matching on the second training data after segmenting, based on the word vector dictionary, in the second embedding layer, to convert the second training data into the matrix form;

S33, wherein, the second multilayer convolution performs the convolution operation on the matrix output from the second embedding layer, the second multilayer convolution comprises one group of first-type one-dimensional convolution layer located at a front portion, and at least one group of second-type one-dimensional convolution layers located at a rear portion, the first-type one-dimensional convolution layer comprises one-dimensional convolution kernels with different number of lengths and the same number of channels, the second-type one-dimensional convolution layers comprise one-dimensional convolution kernels with the same number of lengths and the same number of channels, data of each convolution layer comes from an output of previous convolution layer, and number of matrix rows is maintained during the convolution operation;

S34, inputting an output of the second multilayer convolution into the pooling layer for compression;

S35, inputting an output of the pooling layer into two fully-connected layers to perform information fusion of each channel;

S36, inputting an output of the fully-connected layers into the second softmax function to determine corresponding prediction probabilities of multiple entity relationship labels;

S37, training the second convolutional neural network, calculating a second cross-entropy of the loss function according to predicted probability of a relationship label output by the second convolutional neural network and a true relationship label of the second training data, and minimizing the loss function through the second optimization algorithm to train the second convolutional neural network;

S40, inputting to-be-predicted data into the first convolutional neural network after training and the second convolutional neural network after training,

extracting a knowledge graph triple of the to-be-predicted data according to an entity tagging prediction output by the first convolutional neural network after training and an entity

relationship prediction output by the second convolutional neural network after training, wherein a class corresponding to the maximum probability value in the prediction probabilities of BEMO tagging is selected as the entity tagging prediction output by the first convolutional neural network, a class corresponding to a prediction probability value which is greater than 0.5 is selected as the entity relationship prediction output by the second convolutional neural network, in order to extract the knowledge graph triplet of the to-be-predicted data.

13. The computing device as in claim 12, wherein S10 comprises:

S11, inputting the first training data collected into a Word2Vec algorithm for training to obtain character vectors and construct the character vector dictionary, after the first training data collected is divided into characters and special symbols are removed and non-Chinese characters are removed;

S12, at the same time, inputting the first training data collected into a Word2Vec algorithm for training to obtain word vectors and construct the word vector dictionary, after the first training data collected is segmented into words and special symbols are removed and non-Chinese characters are removed.

14. The computing device as in claim 13, wherein the Word2Vec algorithm for training is implemented by a gensim library in Python.

15. The computing device as in claim 12, wherein the first convolutional neural network and the second convolutional neural network are constructed based on a tensorflow library in Python.

16. The computing device as in claim 12, wherein the first multilayer convolution comprises 5 convolution layers, wherein the first-type one-dimensional convolution layer comprises one-dimensional convolution kernels corresponding to three types of length, the one-dimensional convolution kernels of each length corresponds to 128 channels, four groups of the second-type one-dimensional convolution layers comprise one-dimensional convolution kernels of a length of 3, and number of channels of the one-dimensional convolution kernels corresponding to the four groups of the second-type one-dimensional convolution layers is 384;

and/or, the second multilayer convolution comprises 3 convolution layers, wherein the first-type one-dimensional convolution layer comprises one-dimensional convolution kernels corresponding to three types of length, the one-dimensional convolution kernels of each length corresponds to 128 channels, two groups of the second-type one-dimensional convolution layers comprise one-dimensional convolution kernels of a length of 3, and number of channels of the one-dimensional convolution kernels corresponding to the two groups of the second-type one-dimensional convolution layers is 384.

17. A computer readable storage medium, storing computer programs thereon, the

computer programs executed by a processor to implement steps of a method for knowledge extraction based on TextCNN, the method comprising:

S10, collecting a first training data, and constructing a character vector dictionary and a word vector dictionary;

S20, constructing a first convolutional neural network, and training the first convolutional neural network based on a first optimization algorithm, wherein the first convolutional neural network comprises a first embedding layer, a first multilayer convolution, and a first softmax function, wherein the first embedding layer, the first multilayer convolution, and the first softmax function are connected in turn;

S21, collecting a second training data, wherein the second training data is pre-labeled data, comprising a named entity position label and a named entity relationship label, and inputting the second training data into the first embedding layer after the second training data is divided into multiple characters and special symbols are removed;

S22, performing character vector matching on the second training data with character-level, based on the character vector dictionary, in the first embedding layer, to convert the second training data into a matrix form;

S23, wherein the first multilayer convolution performs a convolution operation on a matrix output from the first embedding layer, the first multilayer convolution comprises one group of first-type one-dimensional convolution layer located at a front portion, and at least one group of second-type one-dimensional convolution layers located at a rear portion, the first-type one-dimensional convolution layers comprise one-dimensional convolution kernels with different number of lengths and the same number of channels, the second-type one-dimensional convolution layers comprise one-dimensional convolution kernels with the same number of lengths and the same number of channels, data of each convolution layer comes from an output of previous convolution layer, and number of matrix rows is maintained during the convolution operation;

S24, outputting the first multi-layer convolution via the first softmax function to determine prediction probabilities of BEMO tagging with a plurality of subdivision categories of each character;

S25, training the first convolutional neural network, calculating a first cross-entropy of a loss function according to the prediction probability of BEMO tagging and a true BEMO label of the second training data, and minimizing the loss function through the first optimization algorithm to train the first convolutional neural network;

S30, constructing a second convolutional neural network, and training the second convolutional neural network based on a second optimization algorithm, wherein the second convolutional neural network comprises a second embedding layer, a second multilayer convolution, a pooling layer, two fully-connected layers and a second softmax function, wherein the second embedding layer, the second multilayer convolution, the pooling layer, two fully-connected layers and the second softmax function are connected in turn, the two fully-connected layers are the same;

S31, segmenting the second training data into words, comprising:

S311, performing preliminary word segmentation on the second training data using a jieba library, and correcting the preliminary word segmentation based on a predicted word segmentation of the first convolutional neural network, if the preliminary word segmentation is different from the predicted word segmentation of the first convolutional neural network, taking the predicted word segmentation of the first convolutional neural network as standard;

S312, inputting the second training data into the second embedding layer after special symbols and non-Chinese characters in the preliminary word segmentation are removed;

S32, performing word vector matching on the second training data after segmenting, based on the word vector dictionary, in the second embedding layer, to convert the second training data into the matrix form;

S33, wherein the second multilayer convolution performs the convolution operation on the matrix output from the second embedding layer, the second multilayer convolution comprises one group of first-type one-dimensional convolution layer located at a front portion, and at least one group of second-type one-dimensional convolution layers located at a rear portion, the first-type one-dimensional convolution layer comprises one-dimensional convolution kernels with different number of lengths and the same number of channels, the second-type one-dimensional convolution layers comprise one-dimensional convolution kernels with the same number of lengths and the same number of channels, data of each convolution layer comes from an output of previous convolution layer, and number of matrix rows is maintained during the convolution operation;

S34, inputting an output of the second multilayer convolution into the pooling layer for compression;

S35, inputting an output of the pooling layer into two fully-connected layers to perform information fusion of each channel;

S36, inputting an output of the fully-connected layers into the second softmax function to determine corresponding prediction probabilities of multiple entity relationship labels;

S37, training the second convolutional neural network, calculating a second cross-entropy of the loss function according to predicted probability of a relationship label output by the second convolutional neural network and a true relationship label of the second training data, and minimizing the loss function through the second optimization algorithm to train the second convolutional neural network;

S40, inputting to-be-predicted data into the first convolutional neural network after training and the second convolutional neural network after training,

extracting a knowledge graph triple of the to-be-predicted data according to an entity tagging prediction output by the first convolutional neural network after training and an entity relationship prediction output by the second convolutional neural network after training, wherein a class corresponding to the maximum probability value in the prediction probability of BEMO tagging is selected as the entity tagging prediction output by the first convolutional neural network, a class corresponding to a prediction probability value which is greater than

0.5 is selected as the entity relationship prediction output by the second convolutional neural network, in order to extract the knowledge graph triplet of the to-be-predicted data.

18. The computer readable storage medium as in Claim 17, wherein S10 comprises:

S11, inputting the first training data collected into a Word2Vec algorithm for training to obtain character vectors and construct the character vector dictionary, after the first training data collected is divided into character and special symbols are removed and non-Chinese characters are removed;

S12, at the same time, inputting the first training data collected into a Word2Vec algorithm for training to obtain word vectors and construct the word vector dictionary, after the first training data collected is segmented into words and special symbols are removed and non-Chinese characters are removed.

19. The computer readable storage medium as in Claim 17, wherein the first convolutional neural network and the second convolutional neural network are constructed based on a tensorflow library in Python.

20. The computer readable storage medium as in Claim 17, wherein the first multilayer convolution comprises 5 convolution layers, wherein the first-type one-dimensional convolution layer comprises one-dimensional convolution kernels corresponding to three types of length, the one-dimensional convolution kernels of each length corresponds to 128 channels, four groups of the second-type one-dimensional convolution layers comprise one-dimensional convolution kernels of a length of 3, and number of channels of the one-dimensional convolution kernels corresponding to the four groups of the second-type one-dimensional convolution layers is 384;

and/or, the second multilayer convolution comprises 3 convolution layers, wherein the first-type one-dimensional convolution layer comprises one-dimensional convolution kernels corresponding to three types of length, the one-dimensional convolution kernels of each length corresponds to 128 channels, two groups of the second-type one-dimensional convolution layers comprise one-dimensional convolution kernels of a length of 3, and number of channels of the one-dimensional convolution kernels corresponding to the two groups of the second-type one-dimensional convolution layers is 384.