CLAIMS

1. A text inputting method, comprising:

obtaining a user identifier, finding a user language module corresponding to the user identifier according to the user identifier; obtaining a user input, generating a candidate sentence list according to the user input; obtaining a universal language model, calculating an on-screen probability of each candidate sentence in the candidate sentence list according to the user language model and the universal language model; or

respectively calculating a standard conditional probability of each word in a user input according to a universal language model; respectively calculating a cache conditional probability of each word in the user input according to a preconfigured cache-based language modeling policy, the user input and a pre-cached user input; calculating a mixed conditional probability of each word according to the standard conditional probability and the cache conditional probability, and obtaining an on-screen probability of each candidate sentence according to the mixed conditional probability;

sorting candidate sentences in the candidate sentence list according to their on-screen probabilities; and

outputting the sorted candidate sentence list.

2. The text inputting method of claim 1, further comprising:

creating the user language model corresponding to the user identifier, and each time a user inputs a word, updating the user language model according to information of the word inputted by the user.

3. The text inputting method of claim 2, wherein the updating the user language model comprises:

recording word information and word frequency information of the word inputted by the user;

obtaining the word information and the word frequency information, dividing the word, performing a word frequency processing to the divided word according to the word frequency information; and

updating the user language model according to the divided word and the processed

word frequency.

4. The text inputting method of any one of claims 1 to 3, wherein the calculating the on-screen probability of each candidate sentence in the candidate sentence list according to the user language model and the universal language model comprises:

performing a linear interpolation to the user language model and the universal language model to generate a mixed model, calculating the on-screen probability of each candidate sentence in the candidate sentence list according to the mixed model.

5. The text inputting method of any one of claims 1 to 3, wherein the user identifier is an account registered by a user in an inputting software, or an identification number assigned for the user, or an IP address or MAC address associated with a device used by the user.

6. The text inputting method of claim 1, wherein a sentence with a largest on-screen probability is selected and outputted, and the sentence with the largest on-screen probability is cached;

the process of respectively calculating the standard conditional probability of each word in the user input according to the universal language model comprises: respectively calculating the standard conditional probability of each word in the user input according to a pre-created standard Ngram language model, comprising:

calculating the cache conditional probability of the ith word in the user input comprises:

obtaining a number of times k_i that a word sequence which includes the ith word and a predefined constant number of words emerges before the ith word in cached training material;

obtaining a number of times k_{i-1} that a word sequence which includes a predefined constant number of words emerges before the ith word in cached training material;

obtaining a time function value of the ith word; and

calculation a ratio of k_i to k_{i-1}, multiplying the ratio with the time function value of the ith word to obtain the cache conditional probability of the ith word in the user input.

7. The text inputting method of claim 6, wherein the time function value is a result obtained by dividing a preconfigured constant by a time interval between the ith word enters the cache area and the word is inputted by the user.

8. The text inputting method of claim 7, wherein the calculating the standard conditional probability of the ith word in the user input comprises:

obtaining a number of times k'_i that a word sequence which includes the ith word and a preconfigured constant number of words before the ith word emerges in training material of the standard Ngram language model;

obtaining a number of times k'_{i-1} that a word sequence which includes the preconfigured constant number of words before the ith word emerges in training material of the standard Ngram language model;

calculating a ratio of k'_i to k'_{i-1}, and taking the ratio as the standard conditional probability of the ith word of the user input.

9. The text inputting method of claim 8, wherein the calculating the mixed conditional probability of the ith word comprises:

A1, determining an interpolation parameter which is between 0 and 1;

A2, determining a product of the interpolation of the standard conditional probability of the ith word;

A3, calculating a product of a difference between 1 and the interpolation parameter and the cache conditional probability of the ith word; and

A4, calculating a sum of the products obtained in A2 and A3 and taking the sum as the mixed conditional probability of the ith word.

10. The method of claim 9, wherein the calculating the on-screen probability of the sentence comprises:

respectively obtaining the mixed conditional probability of each word included in the sentence; and

multiplying the mixed conditional probability of each word to obtain the on-screen probability of the sentence.

11. The text inputting method of claim 10, wherein the cached user input adopts a queue data structure, the value of the time interval of the ith word is a position of the ith word in the cache queue.

12. The text inputting method of any one of claims 6 to 11, further comprising:

after selecting and outputting the sentence with the maximum on-screen probability and before caching the sentence, adjusting the output sentence.

13. The text inputting method of claim 12, wherein the user input comprises: content inputted via an input method, handwritten input or voice recognizing input.

14. The text inputting method of claim 7, wherein if the pre-cached user input is null, the cache conditional probability of each word in the user input equals to the standard conditional probability of the word.

15. A text inputting method, comprising:

obtaining, by a client, a user identifier, finding a user language model corresponding to the user identifier according to the user identifier;

obtaining, by the client, a user input, uploading the user input to a server which generates a candidate sentence list according to the user input;

obtaining, by the server, a universal language model, calculating an on-screen probability of each candidate sentence in the candidate sentence list according to the user language model and the universal language model;

sorting, by the server, the candidate sentences in the candidate sentence list according to their on-screen probabilities and issuing the sorted candidate sentence list to the client; and

receiving, by the client, the sorted candidate sentence list and outputting the sorted candidate sentence list.

16. The text inputting method of claim 15, further comprising:

creating, on the server, the user language model corresponding to the user identifier, and each time a user inputs a word, updating the user language model according to information of the word inputted by the user.

17. The text inputting method of claim 16, wherein the updating the user language model comprises:

recording word information and word frequency information of the word inputted by the user;

obtaining the word information and the word frequency information, dividing the word, performing a word frequency processing to the divided word according to the word frequency information; and

updating the user language model according to the divided word and the processed word frequency.

18. The text inputting method of any one of claims 15 to 17, wherein the calculating the on-screen probability of each candidate sentence in the candidate sentence list according to the user language model and the universal language model comprises:

performing a linear interpolation to the user language model and the universal language model to generate a mixed model, calculating the on-screen probability of each candidate sentence in the candidate sentence list according to the mixed model.

19. A text inputting method, comprising:

obtaining, by a client, a user identifier, finding a user language model corresponding to the user identifier according to the user identifier;

obtaining, by the client, a user input, and generating a candidate sentence list according to the user input;

obtaining, by the client, a universal language model, calculating an on-screen probability of each candidate sentence in the candidate sentence list according to the user language model and the universal language model; and

sorting, by the client, the candidate sentences in the candidate sentence list according to their on-screen probabilities, and outputting the sorted candidate sentence list.

20. The text inputting method of claim 19, further comprising:

creating, on the client, the user language model corresponding to the user identifier, and each time a user inputs a word, updating the user language model according to information of the word inputted by the user.

21. The text inputting method of claim 20, wherein the updating the user language model comprises:

recording word information and word frequency information of the word inputted by the user;

obtaining the word information and the word frequency information, dividing the word, performing a word frequency processing to the divided word according to the word frequency information; and

updating the user language model according to the divided word and the processed word frequency.

22. The text inputting method of claim 19, wherein the calculating the on-screen probability of each candidate sentence in the candidate sentence list according to the user language model and the universal language model comprises:

performing a linear interpolation to the user language model and the universal language model to generate a mixed model, calculating the on-screen probability of each candidate sentence in the candidate sentence list according to the mixed model.

23. A text inputting system, comprising:

a searching module, to obtain a user identifier and find a user language model corresponding to the user identifier according to the user identifier;

a candidate sentence list generating module, to obtain a user input and generate a candidate sentence list according to the user input;

a probability calculating module, to calculate an on-screen probability of each candidate sentence in the candidate sentence list according to the user language model and a universal language model;

a sorting module, to sort the candidate sentences in the candidate sentence list according to their on-screen probabilities; and

an outputting module, to output the sorted candidate sentence list.

24. The text inputting system of claim 23, further comprising:

a user language model creating module, to create the user language model corresponding to the user identifier;

a user language module updating module, to update, each time a user inputs a word,

the user language model according to information of the word inputted by the user.

25. The text inputting system of claim 24, wherein the user language model updating module is further to record word information and word frequency information of the word inputted by the user, obtain the word information and the word frequency information, dividing the word, perform a frequency processing to the divided word according to the word frequency information, and update the user language model according to the divided word and the processed word frequency information.

26. The text inputting system of any one of claims 23 to 25, wherein the on-screen probability generating module is further to perform a linear interpolation to the user language model and the universal language model to generate a mixed model, calculate the on-screen probability of each candidate sentence in the candidate sentence list according to the mixed model.

27. A text processing system, comprising: a client and a server; wherein

the client is to obtain a user identifier and find a user language model corresponding to the user identifier according to the user identifier, obtain a user input, upload the user input to the server, receive a sorted candidate sentence list from the server and output the sorted candidate sentence list;

the server is to generate a candidate sentence list according to the user input, obtain a universal language model, calculate an on-screen probability of each candidate sentence in the candidate sentence list according to the user language model and the universal language model, sort the candidate sentences in the candidate sentence list according to their on-screen probabilities, and issue the sorted candidate sentence list to the client.

28. The text processing system of claim 27, wherein the server is further to perform a linear interpolation to the user language model and the universal language model to generate a mixed model, and calculate the on-screen probability of each candidate sentence in the candidate sentence list according to the mixed model.

29. The text processing system of claim 27, further comprising:

a character processing unit; wherein

the client is further to select a candidate sentence from the sorted candidate sentence list, and output the selected sentence to the character processing unit; and

the character processing unit is to perform a character processing to the candidate sentence.

30. The text processing system of claim 29, wherein the character processing unit comprises: a text file processing unit, a notebook processing unit, an instant messaging, or a presentation processing unit.

31. A text processing apparatus, comprising: a universal language model module, a cache module, a cache-based language modeling module and a mixed model module; wherein

the universal language model module is to receive a user input, calculate a standard conditional probability of each word in the user input respectively and output the standard conditional probability to the mixed model module;

the cache module is to cache a sentence outputted by the mixed model module;

the cache-based language modeling module is to respectively calculate a cache conditional probability of each word in the user input according to a preconfigured cache-based language modeling policy, the user input and the sentences cached in the cache module, and output the cache conditional probability to the mixed model module;

the mixed model module is to calculate a mixed conditional probability of each word according to the standard conditional probability and the cache conditional probability, calculate a sentence probability of each output sentence according to the mixed conditional probability, select an output sentence with the largest sentence probability and output the sentence with the largest sentence probability.

32. The text processing apparatus of claim 31, wherein the universal language model is a standard Ngram language model module, the standard Ngram language model module comprising: a first word sequence frequency counting unit, a second word sequence frequency counting unit and a standard conditional probability calculating unit;

the first word sequence frequency counting unit is to obtain a number of times k'_i that a word sequence which includes the i th word and a predefined constant number of

words before the ith word emerges in the training material of the standard Ngram language model, and output k'_i to the standard conditional probability calculating unit;

the second word sequence frequency counting unit is to obtain a number of times k'_{i-1} that a word sequence which includes the predefined constant number of words before the ith word emerges in the training material of the standard Ngram language model, and output k'_{i-1} to the standard conditional probability calculating unit; and

the standard probability calculating unit is to calculate a ratio of k'_i to k'_{i-1}, and take the ratio as the standard conditional probability of the ith word in the user input.

33. The text processing apparatus of claim 32, the cache-based language modeling module comprises: a third word sequence frequency counting unit, a fourth word sequence frequency counting unit, a time function value obtaining unit and a cache conditional probability calculating unit;

the third word sequence frequency counting unit is to obtain a number of times k_i that a word sequence which includes the ith word and the predefined constant number of words before the ith word emerges in the cached training material, and output k_i to the standard conditional probability calculating unit;

the fourth word sequence frequency counting unit is to obtain a number of times k_{i-1} that a word sequence which includes the predefined constant number of words before the ith word emerges in the cached training material, and output k_{i-1} to the standard conditional probability calculating unit;

the time function value calculating unit is to obtain a time function value of the ith word and output the value to the cache conditional probability calculating unit; and

the cache conditional probability calculating unit is to calculate a ratio of k_i to k_{i-1}, multiply the ratio with the time function value of the ith word to obtain the cache conditional probability of the ith word in the user input.

34. The text processing apparatus of claim 33, wherein the mixed model module includes: an interpolation parameter storage unit, a first multiplying unit, a second multiplying unit, a mixed conditional probability calculating unit, a sentence probability calculating unit and an output sentence selecting unit;

the interpolation parameter storage unit is to store an interpolation parameter

whose value is configured between 0 and 1 in advance;

the first multiplying unit is to calculate a product of the interpolation parameter stored in the interpolation parameter storage unit and the standard conditional probability of the ith word and output the product to the mixed conditional probability calculating unit;

the second multiplying unit is to calculate a difference between 1 and the interpolation parameter and calculate a product of the different and the cache conditional probability of the ith word, and output the product to the mixed conditional probability calculating unit;

the mixed conditional probability calculating unit is to add the products related to the ith word and take the sum as the mixed conditional probability of the ith word;

the sentence probability calculating unit is to calculate a product of the mixed conditional probability of each word to obtain the sentence probability of an output sentence; and

the output sentence selecting unit is to select a maximum sentence probability calculated by the sentence probability calculating unit and output the output sentence corresponding to the maximum sentence probability.

CLAIMS

1. A text inputting method, comprising:

obtaining a user input;

generating a candidate sentence list according to the user input;

for each candidate sentence in the candidate sentence list, respectively calculating a standard conditional probability of each word in the candidate sentence according to a universal language model;

respectively calculating a cache conditional probability of each word in the candidate sentence according to a preconfigured modeling policy, the user input and a precached user input;

calculating a mixed conditional probability of each word according to the standard conditional probability and the cache conditional probability;

obtaining an on-screen probability of the candidate sentence according to the mixed conditional probability;

sorting candidate sentences in the candidate sentence list according to their on-screen probabilities; and outputting the sorted candidate sentence list;

wherein:

the cache conditional probability of each word in the candidate sentence is positively related to a time function value of the word, and the time function value is related to a time when the word enters a cache area used for the pre-cached user input;

wherein the time function value is a result obtained by dividing a preconfigured constant by a time interval between an ith word enters the cache area and a word currently inputted by the user.

2. The text inputting method of claim 1, wherein:

the pre-cached user input corresponds to a user identifier; and the user identifier is an account registered by a user in an inputting software, or an identification number assigned for the user, or an IP address or MAC address associated with a device used by the user.

3. The text inputting method of claim 1, wherein the pre-cached user input adopts a queue data structure, the value of the time interval of the ith word is a position of the ith word in the cache queue.

4. The text inputting method of claim 1, wherein:

in the sorted candidate sentence list, a sentence with a largest on-screen probability is selected and outputted, and the sentence with the largest on-screen probability is cached in the pre-cached user input.

5. The text inputting method of claim 1, wherein the process of respectively calculating the standard conditional probability of each word in the candidate sentence according to the universal language model comprises: respectively calculating the standard conditional probability of each the candidate sentence according to a pre-created standard Ngram language model, including:

obtaining a number of times k'i that a word sequence which includes the ith word and a preconfigured constant number of words before the ith word emerges in training material of the standard N gram language model;

obtaining a number of times k'i-1 that a word sequence which includes the preconfigured constant number of words before the ith word emerges in the training material of the standard Ngram language model;

calculating a ratio of k'i to k'i-1, and taking the ratio as the standard conditional probability of the ith word of the candidate sentence.

6. The text inputting method of claim 1, wherein calculating a cache conditional probability of an ith word in the candidate sentence comprises:

obtaining a number of times ki; that a word sequence which includes the ith word and a predefined number of consecutive words before the ith word emerges in the pre-cached user input;

obtaining a number of times ki-1 that a word sequence which includes the predefined number of words before the ith word emerges in the pre-cached user input;

obtaining the time function value of the ith word; and

calculating a ratio of ki to ki-1, multiplying the ratio with the time function value of the ith word to obtain the cache-based conditional probability of the ith word in the candidate sentence.

7. The text inputting method of claim 1, wherein calculating the mixed conditional probability of an ith word in the candidate sentence comprises:

A1, determining an interpolation parameter which is between 0 and 1;

A2, determining a product of the interpolation parameter and the standard conditional probability of the ith word;

A3, calculating a product of a difference between 1 and the interpolation parameter and the cache conditional probability of the ith word; and

A4, calculating a sum of the products obtained in A2 and A3 and taking the sum as the mixed conditional probability of the ith word.

8. A text inputting method comprising:

obtaining a user input;

generating a candidate sentence list according to the user input;

for each candidate sentence in the candidate sentence list, respectively calculating a standard conditional probability of each word in the candidate sentence according to a universal language model;

respectively calculating a cache conditional probability of each word in the candidate sentence according to a preconfigured modeling policy, the user input and a pre-cached user input;

calculating a mixed conditional probability of each word according to the standard conditional probability and the cache conditional probability;

obtaining an on-screen probability of the candidate sentence according to the mixed conditional probability;

sorting candidate sentences in the candidate sentence list according to their on-screen probabilities;

and outputting the sorted candidate sentence list;

wherein:

in the sorted candidate sentence list, a sentence with a largest on-screen probability is selected and outputted, and the sentence with the largest on-screen probability is cached;

the process of respectively calculating a standard conditional probability of each word in the candidate sentence according to the universal language model comprises:

respectively calculating the standard conditional probability of each word in the candidate sentence according to a pre-created standard Ngram language model, and

calculating the cache conditional probability of an ith word in the candidate sentence comprises:

obtaining a number of times ki that a word sequence which includes the ith word and a predefined constant number of words emerges before the ith word in cached training material;

obtaining a number of times ki-1 that a word sequence which includes a predefined constant number of words emerges before the ith word in cached training material;

obtaining a time function value of the ith word; and

calculation a ratio of ki to ki-1, multiplying the ratio with the time function value of the ith word to obtain the cache conditional probability of the ith word in the user input.

9. The text inputting method of claim 8, wherein the time function value is a result obtained by dividing a preconfigured constant by a time interval between the ith word enters the cache area and a word currently inputted by the user.

10. The text inputting method of claim 8, wherein calculating a standard conditional probability of an ith word in the candidate sentence comprises:

obtaining a number of times k'; that a word sequence which includes the ith word and a preconfigured constant number of words before the ith word emerges in training material of the standard N gram language model;

obtaining a number of times k'i-1 that a word sequence which includes the preconfigured constant number of words before the ith word emerges in the training material of the standard N gram language model;

calculating a ratio of k'i to k'i-1, and taking the ratio as the standard conditional probability of the ith word of the candidate sentence.

11. The text inputting method of claim 8, wherein calculating the mixed conditional probability of an ith word in the candidate sentence comprises:

A1, determining an interpolation parameter which is between 0 and 1;

A2, determining a product of the interpolation parameter and the standard conditional probability of the ith word;

A3, calculating a product of a difference between 1 and the interpolation parameter and the cache conditional probability of the ith word; and

A4, calculating a sum of the products obtained in A2 and A3 and taking the sum as the mixed conditional probability of the ith word.

12. The text inputting method of claim 9, wherein the pre-cached user input adopts a queue data structure, the value of the time interval of the ith word is a position of the ith word in the cache queue.

13. The text inputting method of claim 8, wherein if the precached user input is null, the cache conditional probability of each word in the candidate sentence equals to the standard conditional probability of the word.

14. A text processing apparatus, comprising:

one or more processors;

memory; and

one or more program modules stored in the memory and to be executed by the one or more processors, the one or more program modules including:

a universal language model module, a cache module, a cache-based language modeling module and a mixed model module; wherein the universal language model module is to receive a user input, generate a candidate sentence list according to the user input, calculate a standard conditional probability of each word in the each candidate sentence respectively and output the standard conditional probability to the mixed model module;

the cache module is to cache a sentence outputted by the mixed model module;

the cache-based language modeling module is to respectively calculate a cache conditional probability of each word in each candidate sentence according to a preconfigured cache-based language modeling policy, the user input and the sentences cached in the cache module, and output the cache conditional probability to the mixed model module;

the mixed model module is to calculate a mixed conditional probability of each word according to the standard conditional probability and the cache conditional probability, calculate a sentence probability of each candidate sentence according to the mixed conditional probability, and select and output the sentence with the largest sentence probability; and

the cache-based language modeling module is further configured to calculate a time function value of each word, wherein the cache-based conditional probability of each word is

positively related to the time function value of the word, and the time function value is related to a time when the word enters the cache module;

wherein the time function value is a result obtained by dividing a preconfigured constant by a time interval between an ith word enters the cache area and a word currently inputted by the user.

15. The text processing apparatus of claim 14, wherein the universal language model is a standard Ngram language model module, the standard Ngram language model module comprising: a first word sequence frequency counting unit, a second word sequence frequency counting unit and a standard conditional probability calculating unit;

the first word sequence frequency counting unit is to obtain a number of times k'i that a word sequence which includes an ith word and a predefined constant number of words before the ith word emerges in the training material of the standard Ngram language model, and output k'i to the standard conditional probability calculating unit;

the second word sequence frequency counting unit is to obtain a number of times k'i-1 that a word sequence which includes the predefined constant number of words before the ith word emerges in the training material of the standard Ngram language model, and output k'i-1 to the standard conditional probability calculating unit; and

the standard probability calculating unit is to calculate a ratio of k'i to k'i-1, and take the ratio as the standard conditional probability of the ith word in the candidate sentence.

16. The text processing apparatus of claim 15, the cache-based language modeling module comprises: a third word sequence frequency counting unit, a fourth word sequence frequency counting unit, a time function value obtaining unit and a cache conditional probability calculating unit;

the third word sequence frequency counting unit is to obtain a number of times ki that a word sequence which includes the ith word and the predefined constant number of words before the ith word emerges in cached training material from the cache module, and output ki to the cache conditional probability calculating unit;

the fourth word sequence frequency counting unit is to obtain a number of times ki-1 that a word sequence which includes the predefined constant number of words before the ith word emerges in the cached training material, and output ki-1 to the cache conditional probability calculating unit;

the time function value obtaining unit is to obtain a time function value of the ith word and output the value to the cache conditional probability calculating unit; and

the cache conditional probability calculating unit is to calculate a ratio of ki to ki-1, multiply the ratio with the time function value of the ith word to obtain the cache conditional probability of the ith word in the candidate sentence.

17. The text processing apparatus of claim 16, wherein the mixed model module includes: an interpolation parameter storage unit, a first multiplying unit, a second multiplying unit, a mixed conditional probability calculating unit, a sentence probability calculating unit and an output sentence selecting unit;

the interpolation parameter storage unit is to store an interpolation parameter whose value is configured between 0 and 1 in advance;

the first multiplying unit is to calculate a product of the interpolation parameter stored in the interpolation parameter storage unit and the standard conditional probability of the ith word in the candidate sentence and output the product to the mixed conditional probability calculating unit;

the second multiplying unit is to calculate a difference between 1 and the interpolation parameter and calculate a product of the difference and the cache conditional probability of the ith word, and output the product to the mixed conditional probability calculating unit;

the mixed conditional probability calculating unit is to add the products related to the ith word and take the sum as the mixed conditional probability of the ith word;

the sentence probability calculating unit is to calculate a product of the mixed conditional probability of each word to obtain the sentence probability of the candidate sentence;

the output sentence selecting unit is to select and output the sentence with the largest sentence probability calculated by the sentence probability calculating unit.

18. The text processing apparatus of claim 14, wherein the cache module adopts a queue data structure, the value of the time interval of the ith word is a position of the ith word in the cache queue.