[0026] FIG. 6 schematically shows a flow chart of a method of training a speech text generation model according to embodiments of the present disclosure;

[0027] FIG. 7 schematically shows a block diagram of a speech text generation apparatus according to embodiments of the present disclosure;

[0028] FIG. 8 schematically shows a block diagram of an apparatus of training a speech text generation model according to embodiments of the present disclosure; and

[0029] FIG. 9 schematically shows a block diagram of an electronic device suitable for implementing a speech text generation method and a method of training a speech text generation model according to embodiments of the present disclosure.

DETAILED DESCRIPTION OF EMBODIMENTS

[0030] Embodiments of the present disclosure will be described below with reference to the accompanying drawings. It should be understood that these descriptions are merely exemplary and are not intended to limit the scope of the present disclosure. In the following detailed descriptions, for purposes of explanation, numerous specific details are set forth in order to provide a thorough understanding of embodiments of the present disclosure. It is clear that one or more embodiments may be implemented without these specific details. In addition, in the following descriptions, descriptions of well-known structures and technologies are omitted to avoid unnecessarily obscuring the concept of the present disclosure.

[0031] Terms used herein are for the purpose of describing embodiments only and are not intended to limit the present disclosure. Terms “comprising”, “including” and the like used herein specify a presence of the feature, step, operation and/or component, but do not preclude a presence or addition of one or more other features, steps, operations or components.

[0032] All terms (including technical and scientific terms) used herein have the same meaning as commonly understood by those skilled in the art, unless otherwise defined. It should be noted that the terms used herein should be construed to have meanings consistent with the context of the present description and should not be construed in an idealized or overly rigid manner.

[0033] Where expressions like “at least one of A, B, or C, etc.” are used, they should generally be interpreted in accordance with the meaning of the expression as commonly understood by those skilled in the art (e.g., “a system having at least one of A, B or C” should include, but is not limited to, a system having A alone, having B alone, having C alone, having A and B, having A and C, having B and C, and/or having A, B, C, etc.).

method in embodiments of the present disclosure. The description of the part of the speech text generation apparatus specifically refers to the part of the speech text generation method, which will not be repeated here.

[0130] FIG. 8 schematically shows a block diagram of an apparatus of training a speech text generation model according to embodiments of the present disclosure.

[0131] As shown in FIG. 8, the apparatus 800 of training a speech text generation model may include a sample tagging module 810, a sample first determination module 820, a sample second determination module 830, a sample mask module 840, and a training module 850.

[0132] The sample tagging module 810 is used to perform a part-of-speech tagging on a training sample standard text in a training sample set and a training sample spoken text in the training sample set, so as to obtain a first sample part-of-speech tagging result of the training sample standard text, a second sample part-of-speech tagging result of the training sample spoken text, and a sample modal particle tagging result of the training sample spoken text, where the training sample spoken text is associated with the training sample standard text.

[0133] The sample first determination module 820 is used to determine a sample target part-of-speech from the first sample part-of-speech tagging result according to a distribution feature of sample modal particles.

[0134] The sample second determination module 830 is used to determine a sample predicted insertion position according to a position of a sample content corresponding to the sample target part-of-speech in the training sample standard text.

[0135] The sample mask module 840 is used to mask the sample predicted insertion position in the training sample standard text to obtain a training sample masked standard text, where the training sample masked standard text has the first sample part-of-speech tagging result.

[0136] The training module 850 is used to train an initial speech text generation model using a target training set to obtain a trained speech text generation model, where the target training set includes the training sample masked standard text, the second sample part-of-speech tagging result of the training sample spoken text, and the sample modal particle tagging result of the training sample spoken text.

[0137] According to embodiments of the present disclosure, the apparatus of training the speech text generation model may further include: a sample updating module and a sample constructing module.

[0138] The sample updating module is used to update a first sample standard text in a first sample set and a first sample spoken text in the first sample set by using a sample confusion