CLAIMS

1. A system for incremental learning in machine translation, the system comprising

an input module for obtaining a source language input to be translated into a target language output;

an ensemble model module coupled to the input module, the ensemble model module comprising

a baseline model trained using a first dataset, the baseline model arranged to receive the source language input to generate a baseline model target language output with associated one or more baseline model probability scores;

at least one first adapted model trained using a second dataset, the second dataset being smaller than the first dataset and the second dataset comprises newer data obtained over time after establishment of the first dataset, the at least one first adapted model arranged to receive the source language input to generate a first adapted model target language output with associated one or more first adapted model probability scores; and

a determination module coupled to the ensemble model module for receiving the one or more baseline model probability scores and the one or more first adapted model probability scores, the determination module arranged to generate a system target language output based on the received one or more baseline model probability scores and the received one or more first adapted model probability scores.

2. The system as claimed in claim 1, further comprising the at least one first adapted model being selected from a plurality of adapted models, the plurality of adapted models comprising at least a second adapted model trained using a third dataset, the third dataset being different in size from the second dataset used to train the at least one first adapted model.

3. The system as claimed in claim 2, wherein one of the second dataset and the third dataset is a subset of the other of the second dataset and the third dataset.

4. The system as claimed in claims 2 or 3, wherein the second adapted model is an adaptation based on the baseline model and is trained using the third dataset, the

third dataset comprising the second dataset and newer data accumulated after the establishment of the second dataset.

5. The system as claimed in any one of claims 2 to 4 , wherein the second dataset and the third dataset are each a part of an ensemble dataset.

6. The system as claimed in any one of claims 1 to 5, wherein the first dataset and the second dataset each comprises language pair data of a source language and a target language.

7. The system as claimed in any one of claims 1 to 6, wherein the determination module is arranged to perform an averaging operation based on the received one or more baseline model probability scores and the received one or more first adapted model probability scores to generate the system target language output.

8. The system as claimed in any one of claims 1 to 7, wherein the baseline model and the at least one first adapted model are each based on a transformer deep learning model.

9. The system as claimed in any one of claims 1 to 8, further comprising an output module coupled to the determination module, the output module being arranged to receive the system target language output from the determination module and to provide the system target language output to a user.

10. A method of incremental learning in machine translation, the method comprising

obtaining a source language input to be translated into a target language output;

providing an ensemble model module, the ensemble model module comprising

a baseline model trained using a first dataset and at least one first adapted model trained using a second dataset, the second dataset being smaller than the first dataset and the second dataset comprises newer data obtained over time after establishment of the first dataset;

receiving, at the baseline model, the source language input;

generating, using the baseline model, a baseline model target language output with associated one or more baseline model probability scores;

receiving, at the at least one first adapted model, the source language input;

generating, using the at least one first adapted model, a first adapted model target language output with associated one or more first adapted model probability scores;

receiving at a determination module the one or more baseline model probability scores and the one or more first adapted model probability scores; and

generating, using the determination module, a system target language output based on the received one or more baseline model probability scores and the received one or more first adapted model probability scores.

11. The method as claimed in claim 10, further comprising selecting the at least one first adapted model from a plurality of adapted models, the plurality of adapted models comprising at least a second adapted model trained using a third dataset, wherein the third dataset is different in size from the second dataset used to train the at least one first adapted model.

12. The method as claimed in claim 11, wherein one of the second dataset and the third dataset is a subset of the other of the second dataset and the third dataset.

13. The method as claimed in claims 11 or 12, wherein the second adapted model is an adaptation based on the baseline model and is trained using the third dataset, the third dataset comprising the second dataset and newer data accumulated after the establishment of the second dataset

14. The method as claimed in any one of claims 11 to 13, wherein the second dataset and the third dataset are each a part of an ensemble dataset.

15. The method as claimed in any one of claims 10 to 14, wherein the first dataset and the second dataset each comprises language pair data of a source language and a target language.

16. The method as claimed in any one of claims 10 to 15, further comprising performing at the determination module an averaging operation based on the received one or more baseline model probability scores and the received one or more first adapted model probability scores to generate the system target language output.

17. The method as claimed in any one of claims 10 to 16, wherein the baseline model and the at least one first adapted model are each based on a transformer deep learning model.

18. The method as claimed in any one of claims 10 to 17, further comprising coupling an output module to the determination module;
receiving the system target language output from the determination module; and
providing the system target language output to a user using the output module.

19. A computer readable storage medium having stored thereon software instructions that, when executed by a system processing module of a system for incremental learning in machine translation, cause the system processing module to perform a method of any one of claims 10 to 18.