

CLAIMS

1. A computer-implemented method for verifying labels of records of a dataset, wherein the records comprise sample data and a related label out of a plurality of labels, the method comprising:
 - dividing, by one or more processors, the dataset into a training dataset comprising records relating to a selected label and an inference dataset comprising records with sample data relating to the selected label and all other labels out of the plurality of labels;
 - dividing, by one or more processors, the training dataset into a plurality of learner training datasets, wherein a learner training dataset comprises at least one sample relating to the selected label;
 - training, by one or more processors, a plurality of label-specific few-shot learners, wherein the few-shot learners are trained with one of the plurality of learner training datasets; and
 - performing, by one or more processors, inference by the plurality of trained label-specific few-shot learners on the inference dataset to generate a plurality of sets of predicted label output values.
2. The method of claim 1, further comprising:
 - aggregating, by one or more processors, the predicted label output values within each set of the sets of predicted label output values, thereby generating a respective score value for each of the plurality of the label-specific few-shot learners, the respective score values reflecting a variability of the predicted selected label; and
 - in response to determining that at least a predefined number of score values are below a predefined variability threshold value, determining, by one or more processors, the dataset as verified regarding the selected label.
3. The method according to claim 1 or 2, wherein at least one of the few-shot learners is a one-shot learner.
4. The method according to one of claims 1 to 3, wherein at least one of the few-shot learners is using a Bayesian few-shot learning algorithm.

5. The method according to one of claims 1 to 4, further comprising:
training, by one or more processors, one of the plurality of label specific few-shot learners with the records of the dataset relating the selected label.
6. The method according to one of claims 1 to 5, wherein each of the plurality of label specific few-shot learners has been pre-trained with data samples relating to the selected label.
7. The method according to one of claims 1 to 6, further comprising:
in response to determining that none of the plurality of learner training datasets result in a confidence level above a confidence threshold level, dividing, by one or more processors, the dataset into a second training dataset comprising records relating to a second selected label and a second inference dataset comprising records with sample data relating to the second selected label and all other labels out of the plurality of labels.
8. The method according to one of claims 1 to 7, further comprising:
increasing, by one or more processors, the number of the plurality of learner training datasets from one repetition cycle to a next repetition cycle until a measured improvement of the results is below a predefined threshold value.
9. The method according to one of claims 1 to 8, further comprising:
decreasing, by one or more processors, the training dataset by learner training datasets of the plurality of learner training datasets which result in high confidence levels of the related labels for a next repetition cycle.
10. The method according to one of claims 1 to 9, wherein each of the plurality of label-specific few-shot learners is adapted for a classification in two classes comprising a fit-to-label class and a not-fit-to-label class.

11. The method according to one of claims 1 to 10, wherein dividing the dataset into a training dataset comprising records and an inference dataset further comprises:

 sorting, by one or more processors, the sample data according to known confidence level of a respective related label; and

 selecting, by one or more processors, a portion of sample data for the training dataset such that the sample data of the portion have higher confidence levels of the related label than the remaining portion of the dataset.

12. The method according to one of claims 1 to 11, further comprising:

 upon determining a record having a related confidence level in the dataset, adding, by one or more processors, the respective record to the training dataset.

13. A computer program product for verifying labels of records of a dataset, wherein the records comprise sample data and a related label out of a plurality of labels, the computer program product comprising:

 one or more computer readable storage media and program instructions stored on the one or more computer readable storage media, the program instructions comprising:

 program instructions to divide the dataset into a training dataset comprising records relating to a selected label and an inference dataset comprising records with sample data relating to the selected label and all other labels out of the plurality of labels;

 program instructions to divide the training dataset into a plurality of learner training datasets, wherein a learner training dataset comprises at least one sample relating to the selected label;

 program instructions to train a plurality of label-specific few-shot learners, wherein the few-shot learners are trained with one of the plurality of learner training datasets; and

 program instructions to perform inference by the plurality of trained label-specific few-shot learners on the inference dataset to generate a plurality of sets of predicted label output values.

14. The computer program product of claim 13, further comprising program instructions, stored on the one or more computer readable storage media, to:

aggregate the predicted label output values within each set of the sets of predicted label output values, thereby generating a respective score value for each of the plurality of the label-specific few-shot learners, the respective score values reflecting a variability of the predicted selected label; and

in response to determining that at least a predefined number of score values are below a predefined variability threshold value, determine the dataset as verified regarding the selected label.

15. The computer program product of claim 13 or 14, wherein at least one of the few-shot learners is a one-shot learner.

16. The computer program product of one of claims 13 to 15, wherein at least one of the few-shot learners is using a Bayesian few-shot learning algorithm.

17. The computer program product of one of claims 13 to 16, further comprising program instructions, stored on the one or more computer readable storage media, to:

train one of the plurality of label specific few-shot learners with the records of the dataset relating the selected label.

18. A computer system for verifying labels of records of a dataset, wherein the records comprise sample data and a related label out of a plurality of labels, the computer system comprising:

one or more computer processors;

one or more computer readable storage media; and

program instructions stored on the computer readable storage media for execution by at least one of the one or more processors, the program instructions comprising:

program instructions to divide the dataset into a training dataset comprising records relating to a selected label and an inference dataset comprising records with sample data relating to the selected label and all other labels out of the plurality of labels;

program instructions to divide the training dataset into a plurality of learner

training datasets, wherein a learner training dataset comprises at least one sample relating to the selected label;

program instructions to train a plurality of label-specific few-shot learners, wherein the few-shot learners are trained with one of the plurality of learner training datasets; and

program instructions to perform inference by the plurality of trained label-specific few-shot learners on the inference dataset to generate a plurality of sets of predicted label output values.

19. The computer system of claim 18, further comprising program instructions, stored on the computer readable storage media for execution by at least one of the one or more processors, to:

aggregate the predicted label output values within each set of the sets of predicted label output values, thereby generating a respective score value for each of the plurality of the label-specific few-shot learners, the respective score values reflecting a variability of the predicted selected label; and

in response to determining that at least a predefined number of score values are below a predefined variability threshold value, determine the dataset as verified regarding the selected label.

20. The computer system of claim 18 or 19, wherein at least one of the few-shot learners is a one-shot learner.