

METHODS AND APPARATUSES FOR PERFORMING MODEL OWNERSHIP VERIFICATION BASED ON EXOGENOUS FEATURE

ABSTRACT

5 Embodiments of this specification provide methods and apparatuses for
performing model ownership verification based on an exogenous feature. A specific
implementation of the method includes: selecting samples from an initial sample set to
form a selected sample set; processing sample data of each selected sample in the
selected sample set to obtain a transform sample set formed by a transform sample with
10 an exogenous feature, where the exogenous feature is a feature that sample data of the
initial sample do not have; training a meta-classifier based on a target model, an
auxiliary model, and the transform sample set, where the auxiliary model is a model
trained by using the initial sample set, the target model is a model trained by using the
transform sample set and a remaining sample set in the initial sample set except the
15 selected sample set, and the meta-classifier is used to identify feature knowledge of the
exogenous feature; and inputting related data of a suspicious model into the meta-
classifier and determining, based on an output result of the meta-classifier, whether the
suspicious model is a model stolen from a deployment model, where the deployment
model has feature knowledge of the exogenous feature.

20

FIG. 1