International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

Series X

Supplement 37
(09/2022)

SERIES X: DATA NETWORKS, OPEN SYSTEM COMMUNICATIONS AND SECURITY

ITU-T X.1231 – Supplement on countering spam based on machine learning

ITU-T X-series Recommendations – Supplement 37

ITU-T X-SERIES RECOMMENDATIONS

DATA NETWORKS, OPEN SYSTEM COMMUNICATIONS AND SECURITY

PUBLIC DATA NETWORKS	X.1–X.199
OPEN SYSTEMS INTERCONNECTION	X.200–X.299
INTERWORKING BETWEEN NETWORKS	X.300–X.399
MESSAGE HANDLING SYSTEMS	X.400–X.499
DIRECTORY	X.500–X.599
OSI NETWORKING AND SYSTEM ASPECTS	X.600–X.699
OSI MANAGEMENT	X.700–X.799
SECURITY	X.800–X.849
OSI APPLICATIONS	X.850–X.899
OPEN DISTRIBUTED PROCESSING	X.900–X.999
INFORMATION AND NETWORK SECURITY	
General security aspects	X.1000–X.1029
Network security	X.1030–X.1049
Security management	X.1050–X.1069
Telebiometrics	X.1080–X.1099
SECURE APPLICATIONS AND SERVICES (1)	
Multicast security	X.1100–X.1109
Home network security	X.1110–X.1119
Mobile security	X.1120–X.1139
Web security (1)	X.1140–X.1149
Application Security (1)	X.1150–X.1159
Peer-to-peer security	X.1160–X.1169
Networked ID security	X.1170–X.1179
IPTV security	X.1180–X.1199
CYBERSPACE SECURITY	
Cybersecurity	X.1200–X.1229
Countering spam	X.1230–X.1249
Identity management	X.1250–X.1279
SECURE APPLICATIONS AND SERVICES (2)	
Emergency communications	X.1300–X.1309
Ubiquitous sensor network security	X.1310–X.1319
Smart grid security	X.1330–X.1339
Certified mail	X.1340–X.1349
Internet of things (IoT) security	X.1350–X.1369
Intelligent transportation system (ITS) security	X.1370–X.1399
Distributed ledger technology (DLT) security	X.1400–X.1429
Application Security (2)	X.1450–X.1459
Web security (2)	X.1470–X.1489
CYBERSECURITY INFORMATION EXCHANGE	
Overview of cybersecurity	X.1500–X.1519
Vulnerability/state exchange	X.1520–X.1539
Event/incident/heuristics exchange	X.1540–X.1549
Exchange of policies	X.1550–X.1559
Heuristics and information request	X.1560–X.1569
Identification and discovery	X.1570–X.1579
Assured exchange	X.1580–X.1589
Cyber Defence	X.1590–X.1599
CLOUD COMPUTING SECURITY	
Overview of cloud computing security	X.1600–X.1601
Cloud computing security design	X.1602–X.1639
Cloud computing security best practices and guidelines	X.1640–X.1659
Cloud computing security implementation	X.1660–X.1679
Other cloud computing security	X.1680–X.1699
QUANTUM COMMUNICATION	
Terminologies	X.1700–X.1701
Quantum random number generator	X.1702–X.1709
Framework of QKDN security	X.1710–X.1711
Security design for QKDN	X.1712–X.1719
Security techniques for QKDN	X.1720–X.1729
DATA SECURITY	
Big Data Security	X.1750–X.1759
Data protection	X.1770–X.1789
IMT-2020 SECURITY	X.1800–X.1819

For further details, please refer to the list of ITU-T Recommendations.

Supplement 37 to ITU-T X-series Recommendations

ITU-T X.1231 – Supplement on countering spam based on machine learning

Summary

Supplement 37 to Recommendation ITU-T X.1231 defines a technical framework for countering spam based on machine learning (ML). It may help some relevant persons and companies in spam management, reduce the benefit losses of users and providers, improve user experience and promote the healthy development of telecommunication business.

This Supplement to Recommendation ITU-T X.1231 provides some general scenarios, and characteristics of spam, and defines the general technical framework, and work flows about countering spam based on ML.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T X Suppl. 37	2022-09-02	17	11.1002/1000/15119

Keywords

Machine learning, spam, technical framework.

* To access the Recommendation, type the URL http://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID. For example, http://handle.itu.int/11.1002/1000/11830-en.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

This is an informative ITU-T publication. Mandatory provisions, such as those found in ITU-T Recommendations, are outside the scope of this publication. This publication should only be referenced bibliographically in ITU-T Recommendations.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this publication may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the publication development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents/software copyrights, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the appropriate ITU-T databases available via the ITU-T website at http://www.itu.int/ITU-T/ipr/.

Table of Contents

Introduction

With the development of the information industry, spam is becoming a widespread problem. It can pose a great threat to the safe application and dissemination of information, and cause trouble and inconvenience in people's lives. In general, spam includes e-mail spam, short message service (SMS) spam, multimedia messaging service (MMS) spam, instant messaging spam, harassment phone calls, fraudulent phone calls, etc.

Machine learning is a core field of artificial intelligence (AI) that uses statistical techniques to give computer systems the ability to learn from data, without being explicitly programmed. Compared to normal information, spam has specific characteristics, such as bulk, repetitiveness, hidden or false message origins, etc.; and according to these characteristics, machine learning can learn from this data and make predictions by building models. This Supplement to Recommendation ITU-T X.1231 focuses on countering spam using machine learning.

Supplement 37 to ITU-T X-series Recommendations

ITU-T X.1231 – Supplement on countering spam based on machine learning

1 Scope

This Supplement provides a technical framework for countering spam based on machine learning (ML) to achieve spam governance and control. This framework specifies functional components, work flows of spam recognition and management. In addition, it provides an introduction to ML and covers some general scenarios and characteristics of spam.

2 References

[ITU-T X.1231] Recommendation ITU-T X.1231 (2008), *Technical strategies for countering spam*.

[ITU-T Y.3531] Recommendation ITU-T Y.3531 (2020), *Cloud computing – Functional requirements for machine learning as a service*.

3 Definitions

3.1 Terms defined elsewhere

This Supplement uses the following terms defined elsewhere:

3.1.1 machine learning (ML) [b-ITU-T Y.3172]: Processes that enable computational systems to understand data and gain knowledge from it without necessarily being explicitly programmed.

3.1.2 multimedia message (MMS) spam [b-ITU-T X.1247]: Spam sent via MMS.

3.1.3 SMS spam [b-ITU-T X.1242]: Spam sent via SMS**.**

3.1.4 spam over instant messaging [b-ITU-T X.1244]: A spam targeting users of instant messaging service.

3.1.5 spammer [b-ITU-T X.1240]: An entity or a person creating and sending spam.

3.1.6 voice spam [b-ITU-T X.1246]: Unsolicited, automatically dialled, pre-recorded phone calls, usually with the objective of marketing commercial products or services. The content of voice spam ranges from advertisement of goods to offensive pornographic materials. Voice spam may have various kinds of harmful effects on users and operators.

3.2 Terms defined in this Supplement

None.

4 Abbreviations and acronyms

This Supplement uses the following abbreviations and acronyms:

AI Artificial Intelligence

ML Machine Learning

MMS Multimedia Messaging Service

RPC Remote Procedure Call

SMS Short Message Service

5 Conventions

This Supplement uses the following conventions:

The keywords "is recommended" indicate a requirement which is recommended but which is not absolutely required. Thus, this requirement need not be present to claim conformance.

6 Overview

With the rapid development of technology, the problem of spam is getting more and more attention. Without control measures, spam could lead to immeasurable losses to telecommunication operators, service providers and business users and negatively impact common users. Therefore, spam needs to be dealt with urgently. To counter spam, it is first needed to understand the scenarios with which spam typically appears and its common characteristics.

Spam is electronic information delivered from spammers to receivers by terminals such as computers, mobile phones and so on. It usually spreads over networks, such as communication networks and the Internet, and is sent repeatedly in a low-cost way and received in a timely manner. Receivers of spam usually do not expect to receive it and spammers do not usually intend to acquire recipients' contact information when they send e-mails, multimedia messages and short messages, etc. in bulk. Actually, some software programs can be used to gather communication addresses or telephone numbers from the web or create communication addresses automatically.

Spam includes many categories, such as email spam, short message service (SMS) spam, multimedia messaging service (MMS) spam, voice spam and instant messaging spam. Generally, no matter what type of the spam, they all contain a large number of advertisements, fraud messages, phishing messages and even viruses spread unbridled in the network.

7 Introduction of ML

Machine learning (ML) is a technique that enables machines or computers to learn how to perform tasks. It uses statistical techniques to give computer systems the ability to learn from data, without being explicitly programmed.

Compared to normal information, spam has some specific characteristics, such as bulk, repetitiveness, hidden or false message origins, etc. and ML can learn from data from these characteristics and make predictions by building models. In fact, there are already many companies that use ML to counter spam, for example, using ML to counter iMessage spam, and some companies also use ML and artificial intelligence (AI) to counter spam on their websites or applications.

Based on [ITU-T Y.3531], Figure 7-1 illustrates the generic process of ML, which includes ML data acquisition, ML data processing, ML model development and ML model deployment.

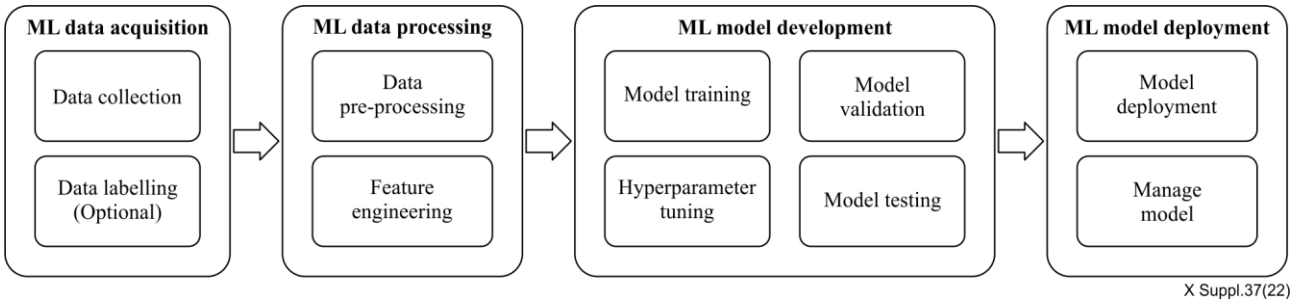

Figure 7-1 – Generic process of machine learning

– **ML data acquisition** collects data for training, which will be grouped into different datasets for ML model training, validation and testing.

– **ML data processing** handles data, including removing inaccurate or incomplete data, feature selection; scaling; and extraction, in order to improve learning performance or to create meaningful information from data.

– **ML model development** is to train and optimize the ML model and ML model deployment is to utilize the model to perform tasks.

– **ML algorithm** is one of the most important parts in ML tasks. It is mainly classified into four categories: supervised learning; unsupervised learning; semi-supervised learning; and reinforcement learning. It is recommended to choose the appropriate ML algorithm according to the practical goals and needs.

8 Technical framework for countering spam based on ML

8.1 General structure

The general structure for countering spam based on ML can be described as in Figure 8-1. It is mainly composed of three layers and five logical modules:

– **Data layer:** composed of the data preprocessing and model training modules

– **Service layer:** composed of the real-time detection and the disposition modules

– **Access layer:** composed of the system management module

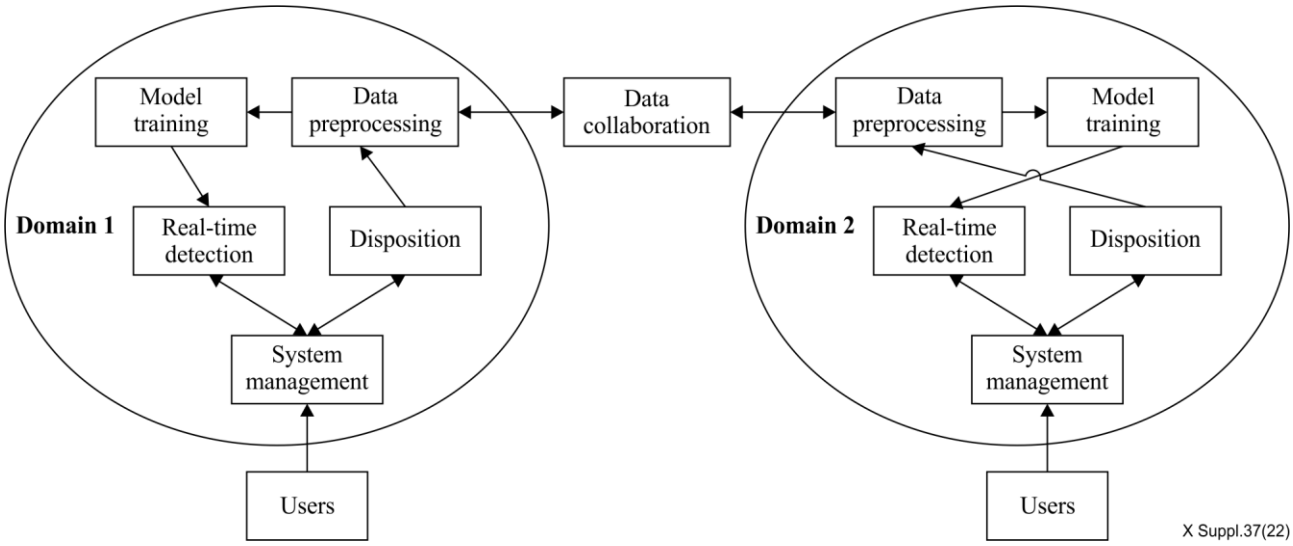

Figure 8-1 – General structure

It is recommended to share the feedback of users' data between different domains.

8.2 Reference model

The reference model for countering spam based on ML is as shown in Figure 8-2.

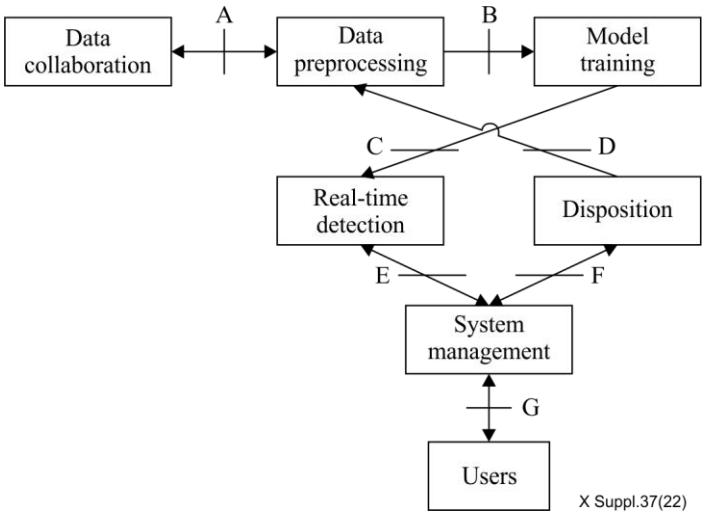

Figure 8-2 – Reference model

The interfaces of the reference model are as follows:

– **Interface A** is between the data collaboration module and the data preprocessing module. Data for different domains collaboration are transmitted through interface A.

– **Interface B** is between the data preprocessing module and the model training module. It is used to transmit the normalizing data.

– **Interface C** is between the model training module and the real-time detection module. It is used to transmit the models and the rules, which are used to analyse spam.

– **Interface D** is between the data preprocessing module and the disposition module. It is used to transmit the spam statistics in order to update the datasets, improve the performance of later recognition model.

– **Interface E** is between the system management module and the real-time detection module. It is used to transmit the management instruction, including selecting models, updating models, etc.

– **Interface F** is between the system management module and the disposition module. It is used to transmit the disposition log, disposition means, management of service states, etc.

– **Interface G** is between the system management module and the users. It is used to transmit the query of users if the users have some question of the recognition result, the query of service subscription, spam report of users, and the alert messages to spammers.

8.3 Module functions

8.3.1 Access layer

The access layer (the outer layer) is connected to the users directly and is mainly responsible for the users accessing the filtering system or sending alert message to users, who are also spammers.

8.3.1.1 Service management module

The service management module is an integrated service management platform in the structure, which includes the following functions:

– **Providing query of service subscription:** Spam filtering is a kind of optional services for users. Therefore, service providers need to provide a filtering spam function for users' subscriptions. If the user has the service subscription, the disposition module will filter the spam if it is recognized as a spam by the real-time detection module and the user will not

receive the spam again. Otherwise, all messages will be transferred using the normal (non-filtering) process, and the user will receive spam with wide ranging possibilities.

– **Receiving feedback of users:** The service management module provides a mechanism to help users send spam reports to itself, which can expand the datasets and may be of great help to improve the system performance. At the same time, if the user is recognized as a spammer, the system management module will send some alert messages to the user and if the user has questions about the result, he can send his questions to the service management module.

– **Saving the operation log and system log:** The service management module provides accounting functions and saves all the operation logs, system logs and statistics information. If the user has questions about the result, this part can provide corresponding record, which can explain why this user is recognized as a spammer.

8.3.2 Service layer

The service layer implements the core functions of the filtering system and it provides a real-time detection function and corresponding disposition function.

8.3.2.1 Real-time detection module

The functions of the real-time detection module include:

– **Model application:** The process of applying the trained model is in this module, which requires a proper coding interface (or microservice). It can provide a real-time detection function.

– **Automatic model deployment:** It is recommended to provide an online automatic model deployment and update. The model training module will adjust the model or strategies all the time according to the practical results, feedback of users and new datasets. So if there is a better model, it will be updated to the real-time detection module.

8.3.2.2 Disposition module

The functions of the disposition module include:

– **Making further judgement:** After a suspicious user is parted from the definite legitimate users, it is recommended to make further judgements to decrease the possibility of false alarms, such as manual determinations and feedback from recipients.

– **Disposing:** According to the final result and policies, the disposition module may block the spam, send alert message to spammers, stop providing service for spammers, etc.

– **Providing disposition log:** It is important to send the disposition result to the system management module, as well as the message of the spam and the information of spammer.

– **Sending the spam message to the data preprocessing module:** After further judgement, it is recommended to send the spam message to the data preprocessing module. The data preprocessing module will take it as a new data example and use it to improve the recognition models.

8.3.3 Data layer

The data layer is mainly responsible for preparing the datasets and model training. It is recommended to include multiple datasets from different legitimate sources. This data can be stored in permanent media such as disks, tapes, etc., as database-format or as text-format. It is also recommended to contain multiple model types, such as the traditional ML model, AI model and hybrid model.

8.3.3.1 Data preprocessing

The functions of the data preprocessing module include:

– **Data collection:** Data collection is one of the most important procedures for high quality models. Generally, mass data and reliable sources are two good attributes for training datasets. Data will be collected through interfaces, offline files, remote procedure calls (RPCs), etc.

– **Whitelist filtering:** Some data may be marked as spam wrongly, so it is recommended to filter or modify the trusted data through a whitelist.

– **Data normalization:** After data collection and whitelist filtering, the data may either not have a uniform format or be missing some key attributes. There may be some duplicated data as well, so data normalization is recommended to be taken into consideration according to the practical tasks and demands.

Data normalization generally includes data extraction, data cleaning and data conversion as shown in Figure 8-3.

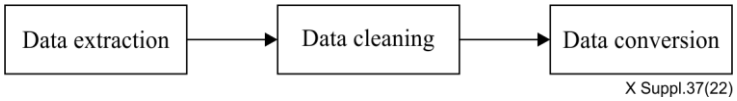

Figure 8-3 – Data normalization procedures

1) Data extraction is to extract useful data from the massive data and appropriately reduce the data dimension, which is the primary processing work after data collection and whitelist filtering.

2) Data cleaning is to ensure the accuracy of data through proper correction, completion, elimination and other technical operations.

3) Data conversion is to convert the data into the same format, which is also convenient to dispose of in the code.

8.3.3.2 Model training

The functions of the model training module include:

– **Dataset generation:** The process of dataset generation is very important, because the selection of the dimension, quantity and demarcation of the final dataset will have a big influence on the accuracy of model selection and model learning.

Dataset generation mainly includes dimension reduction and dataset partition:

- **Dimension reduction** is used to try to reduce the number of data features, and it is usually performed in two ways: feature selection and feature extraction.

  1) **Feature extraction** may combine some different features to get new features according to the relationship between features. Feature extraction will thus change the original feature space and have some new features.

  2) **Feature selection** selects a subset from the original feature space. It is a process of choosing which features are more important according to the purpose and practical demands.

– **Dataset partition** divides the data into three parts: train dataset, validation dataset and test dataset.

1) Train dataset is used to estimate the model.

2) Validation set is used to adjust the model parameters to obtain the optimal model.

3) Test set is used to test the performance of the optimal model.

These three parts are randomly selected from the sample. One of the most typical dataset partitions is that the training dataset accounts for 50% of the total sample, while the others account for 25% each.

– **Model generation:** After data preparation, model generation will use different models to test datasets and find the best model by comparing the results of those models. Generally, model generation includes algorithm selection, model training and model testing as shown in Figure 8-4.

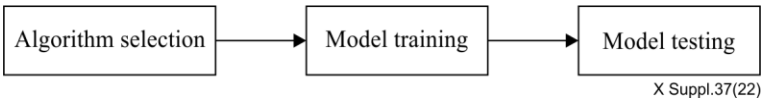

Figure 8-4 – Model generation procedures

1) **Algorithm selection:** There are four main types of algorithms: supervised learning, semi supervised learning, unsupervised learning, and reinforcement learning. Algorithm selection needs to consider the datasets and practical demands.
2) **Model training:** After algorithm selection, the model training can be carried out. Parameters selection and adjustment are recommended to be paid more attention, and the model training process needs a lot of time. Those two parts mainly depend on the engineer's experience and the difficulty of the algorithm and its parameters.
3) **Model testing** will use the training dataset to evaluate the performance. Through this part, the accuracy of the model can be further optimized.

8.3.3.3 Data collaboration

The functions of the data collaboration module include:

– **Data sharing:** It is also recommended to share some data between different domains under user permission, regulations and national laws. It helps to expand the amount of the data of different domains.

9 Work flows

In general, it can be considered that countering spam based on ML consists of eight procedures as shown in Figure 9-1.

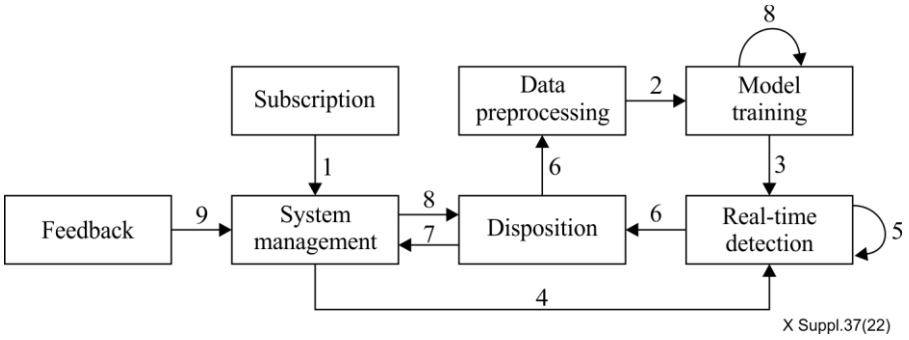

Figure 9-1 – Countering spam based on ML processing procedures

– **Procedure 1: Subscription**

Service providers provide a filtering spam service, and the users subscribe to that service by using the system management module. It is recommended that the system management module provides some simple service subscription methods, such as SMS.

– **Procedure 2: Data preprocessing**

The work of data preprocessing is mainly preparing data for model training, including data collection, filtering and data normalization. After preparing those data, the data preprocessing module will send them to the model training module.

It is recommended to get more data in order to gain better performance. The source of data mainly contains users' feedback, collaboration with other domains, etc.

– **Procedure 3: Model training**

The model training will divide those data into different datasets, and then generate some models.

– **Procedure 4: Model deployment**

The system management module will select the best model by comparing the results of model training, and send deployment instructions to the real-time detection module, which is deployed in the production environment.

– **Procedure 5: Real-time detection**

The real-time detection module will analyse whether there is a spam or not according to the model policies. Once there is a spam confirmed by the model, the real-time detection module will send the spam to the disposition module.

Since different types of spam have different characteristics, it is recommended to deploy multiple different models at the same time. It also needs to adjust or update the model according to the practical results.

– **Procedure 6: Disposition**

The disposition module will make further judgements after receiving the possible results from the real-time detection module, and dispose the spam according to rules, such as blocking it or sending alert message to the spammers.

After that, the disposition module will send the disposition result to the system management module. It is also recommended to send the spam message to the data preprocessing module as new data examples.

– **Procedure 7: Accounting**

The system management module will save all the operation log, system log and statistics information.

– **Procedure 8: Adjustment and optimization**

The system management module will adjust the disposition measures and strategies according to the user's feedback. In addition, the model training will optimize those model and related parameters based on new data samples and the practical results.

– **Procedure 9: Feedback**

The participation of the spam's receivers will be helpful for countering spam effectively and efficiently. They can feedback the spam they received or if the result of spam recognition is wrong to the system management module. All the feedbacks may be beneficial to the optimization of the system and improve the accuracy of the model.

Bibliography

[b-ITU-T X.1240] Recommendation ITU-T X.1240 (2008), *Technologies involved in countering e-mail spam.*

[b-ITU-T X.1241] Recommendation ITU-T X.1241 (2008), *Technical framework for countering email spam.*

[b-ITU-T X.1242] Recommendation ITU-T X.1242 (2009), *Short message service (SMS) spam filtering system based on user-specified rules.*

[b-ITU-T X.1244] Recommendation ITU-T X.1244 (2008), *Overall aspects of countering spam in IP-based multimedia applications*.

[b-ITU-T X.1245] Recommendation ITU-T X.1245 (2010), *Framework for countering spam in IP-based multimedia applications.*

[b-ITU-T X.1246] Recommendation ITU-T X.1246 (2015), *Technologies involved in countering voice spam in telecommunication organizations*.

[b-ITU-T X.1247] Recommendation ITU-T X.1247 (2016), *Technical framework for countering mobile messaging spam.*

[b-ITU-T Y.3172] Recommendation ITU-T Y.3172 (2019), *Architectural framework for machine learning in future networks including IMT-2020.*

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	Tariff and accounting principles and international telecommunication/ICT economic and policy issues
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Environment and ICTs, climate change, e-waste, energy efficiency; construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling, and associated measurements and tests
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	**Data networks, open system communications and security**
Series Y	Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2022