



Dynamic Spam Filtering Using Ontology



Minoo Raki*

Bitá Shadgar

School of Computer Architecture, Faculty of Computer Science, Azad University of Dezfoul, Khuzestan, Iran.

Assist Professor in Faculty of Computer Science at Shahid Chamran University of Ahvaz, Ahvaz, Khuzestan, Iran

Introduction

The growth of using Email has resulted in dramatic increasing of the unsolicited bulk Email, known as Spam.

Several approaches have been proposed to deal with Spam, and filtering is the most famous one.

Aim

To propose a dynamic Content-Based Spam filtering using WordNet Ontology and Wikipedia Knowledge base.

To reach to high accuracy in spam detection

Methods & Materials

Codes are written in JAVA language using NetBeans IDE

RapidMiner is applied for mixing different Experts using machine learning methods.

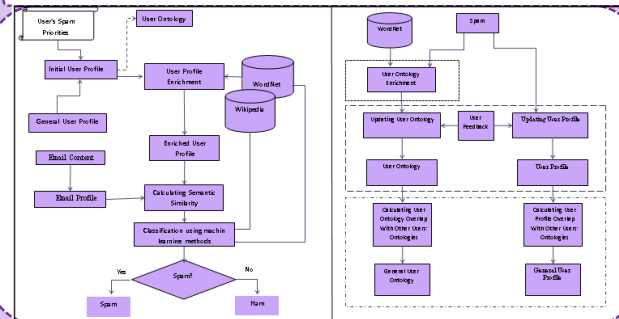
LingSpam is used as a data set

Generating a general filtering model and a base Ontology using machine learning methods

Updating User Ontology according to his feedbacks.

Updating General User Ontology periodically to obtain dynamism.

This method is done in 2 separate steps which is shown in Figure1 below.



Results & Discussion

Consideration of meaning as well as using semantic similarity methods has led to overcoming the problem of «Word Ambiguity».

Despite of machine learning based models, Proposed filter is data set independent.

the idea of combining experts causes filtering error correction.

organizing users' priorities in different user profiles provides personalization.

user priorities improve during the enrichment process.

The general user profile enables users to share their experiences of Spam.

as the most important feature the proposed filter model is dynamic.

Conclusion

The proposed Method is:

Dynamic

Data Set Independent

Personalized

Able to overcome Word Ambiguity

Able to correct errors

Using the benefits of Ontology and Knowledge base.

keeping its high 98.26 percent recall and 92.32 percent accuracy in deal with any changes in user priorities or incoming Emails.

Its recall is 2.26 percent higher than a previous filter which was evaluated using the same data set.

Acknowledgements

I would like to thank Dr. Bitá Shadgar for her expert advice and encouragement throughout this difficult project.

Tables

Comparing proposed dynamic method with Previous methods

Method	Recal l	Precisi on	Accura cy	measure F-
Our Proposed method using KNN	98.26	92.94	92.32	95.52
proposed by Laorden	96.00	99.00	-	-
proposed by Li & Huang	-	-	98.70	-

Graphs

Chart1: Knowledge Based Filtering Results(Based on Wikipedia & WordNet)

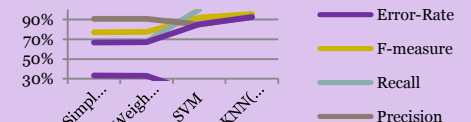


Chart2: Filtering Methods Based on Mixture of Experts

