USER INTERFACES OF SPAM FILTERS: REVIEW AND SUGGESTIONS

Completed Research Paper

Dongmin Kim
University of New Brunswick in Saint John
Faculty of Business
PO Box 5050
Saint John, NB, CANADA E2L 4L5
dongmin@unbsj.ca

Richard K. Cho
University of New Brunswick in Saint John
Faculty of Business
PO Box 5050
Saint John, NB, CANADA E2L 4L5
rcho@unbsj.ca

Ming Yang
University of New Brunswick in Saint John
Computer Science Department
doreen_yangming@hotmail.com

Abstract

Many organizations use anti-spam filters or spam filters, which are known to be among the best available measures to fight against spam. Although spam filters are useful, there is no perfect filter. Spam filters sometimes inaccurately classify legitimate e-mails into spams and spams into legitimate e-mails. This study reviews the current state of user interfaces of spam filters from the aspect of perceived restrictiveness and management of filtration levels. The main focuses of the review are: (1) whether spam filters have functions for users to report misclassifications, (2) whether spam filters provide functions to adjust filtration levels, and (3) whether spam filters produce effective management reports (e.g., frequency of misclassifications) so that managers in an organization can make a decision of an appropriate filtration level. In this study, 35 popular spam filtering software are examined and suggestions to improve user interfaces of spam filters are made.

Keywords: spam, anti-spam filters, user control, filtration level, Human computer interaction, HCI.

Introduction

Spam, unsolicited commercial e-mails, accounts for around 85 percent of total e-mails sent over networks (Goodman et al. 2007, Messagelabs 2011). Bill Gates alone receives about four million spams per year (BBC News 2004). Spam is no longer just nuisance for end users (Fallows 2003, Kraut et al. 2005). Spam has become one of the key concerns that an organization deals with. Spam's costs to an organization include not only waste of computer and network resources but also loss of employees' productivity (Denning 2006, Eighme 2006, Hann et al. 2006, Swartz 2005). According to Nucleus Research (2007), spam costs $70 billion to all US businesses annually.

Many organizations use anti-spam filters or spam filters, which are considered to be among the best available measures to fight against spam (Pavlov et al. 2005). Although spam filters are useful, there is no perfect filter. Spam filters sometimes inaccurately classify legitimate e-mails into spams and spams into legitimate e-mails.

Many studies (Androutsopoulos et al. 2000a; 2000b, Sahami et al. 1998, Zorkadis et al. 2005) on anti-spam filters focus on the filtering algorithms per se. Although these studies provide useful information regarding the adoption stage (e.g., selecting a certain anti-spam filter among many available filters available in the industry), they provide few suggestions regarding the use stage (e.g., monitoring performance of a given anti-spam filter, and taking actions to improve performance of a given anti-spam filter in an organization). In other words, there is a paucity of research on how managers should manage a given anti-spam filter in an organization. This is an important question to Chief Information Officers (CIOs) because spam control can influence the reputation of an Information Technology (IT) department. It is known that employees perceive spam control failure as a security failure (Tuesday 2003). The more spams that appear in an employee's e-mail box, the more likely employees, including top executives, think that their IT department is not capable of handling security issues.

In this paper, interfaces of 35 spam filters are reviewed. The main purposes of this review are (1) to examine how well these spam filters equip managers in an organization to manage a given spam filters and (2) to provide suggestions for future development of spam filters.

Literature Review

Several researchers (Silver 1988; Wang and Benbasat 2009) discuss the perceived restrictiveness of online recommendation agents (e.g., product recommendation systems) as an important factor that undermines intentions to make use of such decision aids. According to Wang and Benbasat (2009), online recommendation agents usually employ only pre-embedded decision rules; hence, users are unable to use variations of such decision rules, which results in their feeling restricted. We view anti-spam filters as a kind of agent in the sense that anti-spam filters act on behalf of users with pre-determined rules. Because spam filters are not perfect, as many other recommendation agents are, spam filters inevitably produce misclassification errors (i.e., sometimes classify legitimate e-mails into spams and spams into legitimate e-mails). Wang and Benbasat (2009) suggest that perceived restrictiveness negatively influences intentions to use a recommendation agent.

In the case of spam filters, we identified several potential areas in which users are likely to feel restrictiveness.

(1) When users are unable to report spam as spam to spam filters.

Spam filters often make incorrect classifications (e.g., spam as a legitimate e-mail and a legitimate e-mail as spam). From the viewpoint of users, these errors are very obvious. If users have no means of informing spam filters of these obvious misclassifications, it is likely that they will feel restrictiveness.

(2) When users are unable to change the filtration level.

Let's assume that a user notices that many spams are appearing in the regular in-box. They would then want to strengthen the filtration level so that fewer spams pass the filtration. If there are no such functions to adjust the filtration level, it is likely that users will feel that the spam filters are restrictive.

Meanwhile, it is a challenging decision for managers to determine an optimal filtration level because there is trade-off in raising or lowering a filtration level of a given filter. For example, if a filtration level is set too high, more legitimate e-mails would be quarantined as spam, while fewer spams would appear in the regular in-box. By the same token, if a filtration level is set too low, more spams would appear in the regular in-box, while fewer legitimate e-mails would be quarantined as spam.

Cho and Kim (2011) argue that managers can make an effective decision on filtration levels by considering total perceived cost in using spam filters. They propose a model of total perceived cost:

Total Perceived Cost of Using Spam Filter = Perceived Cost of FN + Perceived Cost of FP

In the model, they divide the total perceived cost into two components: one is perceived cost of deleting spams from the regular in-box, namely cost of False Negative (FN), and the other is perceived cost of recovering legitimate e-mails from spam mail box, namely cost of False Positive (FP).

Based on Cho and Kim's (2011) argument, this study examines the following aspect:

(1) Whether or not the spam filter provides users with a management report regarding the trend and frequency of FN and FP.

Method

We utilized search engines (e.g., Google and Yahoo) to identify spam filters available in the industry and added publicly available e-mails (e.g., Hotmail, Yahoo mail, and Gmail) as well as an e-mail system used in the authors' university. The thirty-five spam filters are listed below.

1. Windows Live Hotmail
2. Yahoo
3. Gmail
4. Webmail used in a university
5. Spam Eater Pro
6. CA Anti-Spam
7. ChoiceMail One
8. Spam Buster
9. Cloudmark Desktop
10. Spam Agent
11. iHateSpam
12. Mailwasher Pro
13. SpamKiller
14. Mailshell Anti-Spam Desktop (SpamCatcher)
15. SpamTitan
16. SpamDrain
17. SpamAssassin
18. Kaspersky Anti-Spam
19. Clear My Mail
20. Spamjadoo
21. AntispamServant
22. Spam Eliminator
23. Spamweed
24. Spambully
25. Zaep Antispam

26. Spamfighter
27. Allume SpamCatcher
28. AntiSpam Personal
29. Symantec Brightmail Antispam
30. Symantec Premium Antispam
31. GFI Mail Essentials
32. Spam Sleuth Enterprise
33. Xmicro AntiSpam
34. Hexamail Guard Basic
35. Sophos Security Suite SBE

Although this is not an exhaustive list of spam filters, we think that reviewing these 35 filters can provide a reasonable snapshot of present spam filters' interfaces.

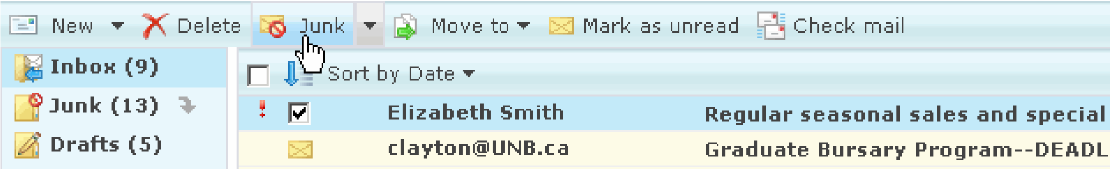
In 2009, we examined interfaces after installing a trial version of each spam filter (28 out of 35). For the other seven filters, we examined their manuals, which showed interfaces.

Results

A summary of the review results is listed in the table 1 below.

(1) Do spam filters provide functions for users to report misclassifications to spam filters?

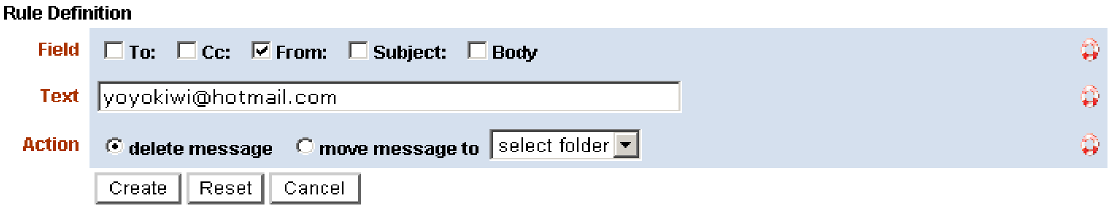
The first aspect of spam filters being examined is whether or not spam filters allow users to report classification errors (e.g., spams that are classified as legitimate e-mail and innocent e-mails that are classified as spam). If users are able to report these misclassification errors, then similar errors can be reduced in the future. For example, Hotmail (Figure 1) allows users to select one or more messages in the inbox and click the Junk button to report these spams.



	Sender	Subject
☑	Elizabeth Smith	Regular seasonal sales and special
□	clayton@UNB.ca	Graduate Bursary Program--DEADL

Figure 1. Users can report spam in Hotmail.

However, some spam filters do not allow users to do this (Figure 2). The only way for users to block a spam is to set up their own filtering rules to specify either the sender’s address or domain so that emails from the specified address or domain will be blocked in the future.



Rule Definition

Field □ To: □ Cc: ☑ From: □ Subject: □ Body

Text yoyokiwi@hotmail.com

Action ● delete message ○ move message to select folder

Figure 2 Users need to add their own filtering rules.

Among the 35 spam filters that have been examined, only 54 percent provide such a function. Nineteen of them allow users to report spams, while 16 of them do not provide this kind of function.

Table 1. Summary of Review

No.	Name of Spam Filter	End User Reporting of Misclassification	Filtration Level Adjustment	Statistics of FN and FP
1	Windows Live Hotmail	Y	Y	N
2	Yahoo	Y	N	N
3	Gmail	Y	N	N
4	A Webmail in a university	N	N	N
5	Spam Eater Pro	N	N	N
6	CA Anti-Spam	N	Y	N
7	ChoiceMail One	N	N	N
8	Spam Buster	N	N	N
9	Cloudmark Desktop	Y	N	N
10	Spam Agent	Y	N	N
11	iHateSpam	Y	Y	N
12	Mailwasher Pro	Y	Y	N
13	SpamKiller	Y	Y	N
14	Mailshell Anti-Spam Desktop (SpamCatcher)	Y	Y	N
15	SpamTitan	N	Y	N
16	SpamDrain	Y	N	N
17	SpamAssassin	Y	Y	Y
18	Kaspersky Anti-Spam	N	Y	N
19	Clear My Mail	N	N	N
20	Spamjadoo	N	N	N
21	AntispamServant	N	Y	Y
22	Spam Eliminator	Y	N	N
23	Spamweed	Y	Y	N
24	Spambully	Y	Y	Y
25	Zaep Antispam	N	N	N
26	Spamfighter	Y	Y	N
27	Allume SpamCatcher	Y	Y	N
28	AntiSpam Personal	N	N	N
29	Symantec Brightmail Antispam	N	Y	N
30	Symantec Premium Antispam	N	Y	N
31	GFI Mail Essentials	N	N	N
32	Spam Sleuth Enterprise	Y	Y	N
33	Xmicro AntiSpam	Y	N	N
34	Hexamail Guard Basic	Y	Y	N
35	Sophos PureMessage for Microsoft Exchange	N	Y	N
	Number of Filters That Support Functions in The Column Heading	19	19	3

If users have no means to inform misclassification errors to a spam filter, it is likely that they feel that the spam filter is restrictive, considering that the spam filter would make similar incorrect classifications repetitively (e.g., spam as a legitimate e-mail and a legitimate e-mail as spam).

(2) Do spam filters provide functions for users to adjust filtration level of spam filters?

The second aspect of spam filters being examined is whether or not they allow firms or users to control the spam filtering strength according to their individual preferences. Some spam filters allow users to set their own spam filtering level which make spam control more flexible. Our focus here is whether a CIO has a means to respond to the poor performance of a given spam filter. If too many spams are classified as legitimate e-mail, a CIO would want to raise the filtration level. If too many legitimate emails are quarantined as spam, a CIO would want to lower the filtration level. Fifty-four percent of the surveyed spam filters provide this function; 19 of them provide users with a function to control the spam filtering level according to their individual preferences, while 16 of them do not.

(3) Do spam filters provide users with a management report regarding the frequency of FN (false positive) and FP (false negative)?

This aspect is about management reporting of spam filters on how effectively a given anti-spam filter works. In the adoption stage of anti-spam filters, average filtration rate (i.e., spams filtered out of total spams) would be an important factor. However, in the use stage, managers need to attend to the trend of FP and FN. Only nine percent of the surveyed spam filters (three out of 35) provided users with a report regarding the statistics of FP and FN.

Discussion and Conclusion

This study reviewed interfaces of 35 filters from the aspects of restrictiveness and management reporting functions. From the viewpoint of perceived restrictiveness, only 54 percent (19 out of 35) of spam filters provide functions for users to report classification errors and to adjust filtration levels. Given that perceived restrictiveness influences intentions to use a spam filter, spam filter developers need to consider providing these functions to reduce perceived restrictiveness of a spam filter.

In terms of management reporting, only three spam filters (out of 35) provided some kinds of statistics about misclassifications (e.g., frequency of FN and FP). As Cho and Kim (2011) suggested, managers need to monitor performances of a given spam filter and to regularly make a filtration decision to serve users better. In this regard, we recommend that spam filter developers include a function which provides statistics and trends of misclassifications in a graphical format so that managers can make decisions based on the trend changes of misclassification. For example, a sudden increase in false positives or false negatives may signal that a present spam filter is performing poorly. In such a case, managers can review and adjust the current filtration level or review the need for another spam filter.

This study assists practitioners and researchers in several ways. First, we provided a snapshot of 35 spam filters available in the industry. This snapshot can be used a base for researchers to study interfaces of spam filters. Users can include the three aspects discussed in this study as a part of selection criteria of spam filters in their organization. Second, this snap shot highlighted the main areas requiring improvement. In our opinion, spam filter developers in general seem to pay little attention on how IT managers should monitor and manage spam filters. Spam filter developers can use this study to differentiate themselves from competitors by adding additional functionality (e.g., statistics of misclassification) to help IT managers to manage spam. Anecdotal evidence from an interview with a university in the western part of North America suggest that management of spam filters is viewed as technical staffs' job not as IT managers' job. As discussed in the introduction, considering that spam control failure is viewed as IT security control failure, the authors argue that IT managers need to adopt the filtration trend as their regular review item and that spam filter developers need to consider providing the functions discussed in this paper in order to help IT managers to manage spam filters in an organization.

Acknowledgements

This study was supported by a grant from the Social Sciences and Humanities Research Council of Canada (SSHRC).

References

Androutsopoulos, I., G. Paliouras, V. Karkaletsis, G. Sakkis, C. D. Spyropoulos, P. Stamatopoulos. 2000a. Learning to Filter Spam E-Mail: A Comparison of a Naive Bayesian and a Memory-Based Approach. *Proceedings of the 4th European Conference on Principles and Practice of Knowledge Discovery in Databases PKDD.*

Androutsopoulos, I., J. Koutsias, K. V. Chandrinos, C. D. Spyropoulos. 2000b. An Experimental Comparison of Naive Bayesian and Keyword-Based Anti-Spam Filtering with Personal E-Mail Messages. *Proceedings of the 23rd annual international ACM SIGIR conference*. 160-167.

BBC News. 2004. Bill Gates 'most spammed person'. BBC News. (Nov. 18). http://news.bbc.co.uk/2/.

Cho, R. K., Kim, D. 2011. New Model for Optimal Threshold Level in Receiver Operating Characteristic (ROC): The Case of Filtering Spam Mail. Working Paper, University of New Brunswick in Saint John.

Denning, P. J. 2006. Infoglut. *Communications of the ACM.* 49(7). 15-19.

Eighme, J. E. 2006. "Spam: Choosing the Right Defense." *CPA Journal* 76(7). 48-50.

Fallows, D. 2003. SPAM: How It Is Hurting Email and Degrading Life on the Internet. http://www.pewinternet.org/pdfs/PIP_Spam_Report.pdf.

Goodman, J., G. V. Cormack, D. Heckerman. 2007. Spam and the Ongoing Battle for the Inbox. *Communication of the ACM.* 50(2). 25-33.

Hann, I. H., K. L. Hui, Y. L. Lai, S. Y. T. Lee, I. Png. 2006. Who Gets Spammed? *Communication of the ACM,* 49(10). 83-87.

Kraut, R., S. Sunder, R. Telang, J. Morris. 2005. Pricing Electronic Mail to Solve the Problem of Spam. *Human-Computer Interaction.* 20(1). 195-223.

MessageLabs. 2011. 2010 Annual Security Report. (available from http://www.messagelabs.com/mlireport/MessageLabsIntelligence_2010_Annual_Report_FINAL.pdf).

Nucleus Research 2007 Nucleus Research: Spam Costing US Businesses $712 Per Employee Each Year (available from http://nucleusresearch.com/news/press-releases/nucleus-research-spam-costing-us-businesses-712-per-employee-each-year/)

Pavlov, O., N. Melville, R. Plice. 2005. Mitigating the tragedy of the digital commons: The problem of unsolicited commercial e-mail. *Communications of AIS.* 2005(16). 73-90.

Sahami, M., S. Dumais, D. Heckerman, E. Horvitz. 1998. A Bayesian Approach To Filtering Junk Email. AAAI Workshop on Learning for Text Categorization. Madison. Wisconsin. ftp://ftp.research.microsoft.com/pub/ejh/junkfilter.pdf

Silver, M.S. 1988 User Prescriptions of Decision Support System Restrictiveness: An Experiment. *Journal of Management Information Systems* (5:1), Summer88, pp 51-65.

Swartz, N. 2005. Deleting Spam Costs Businesses Billions. Information Management Journal. 39(3), 10-10

Tuesday, V. 2003. Spam Issue Viewed As IT Security Failure. *Computerworld.* 37(2). 36.

Wang, W., and Benbasat, I. 2009 Interactive Decision Aids for Consumer Decision Making in e-commerce: The Influence of Perceived Strategy Restrictiveness," *MIS Quarterly* (33:2), pp 293-320.

Zorkadis, V., D.A. Karras, M. Panayotou. 2005. Efficient Information Theoretic Strategies for Classifier Combination, Feature Extraction and Performance Evaluation in Improving False Positives and False Negatives for Spam E-Mail Filtering. *Neural Networks.* 18(5-6). 799-807.