DEDICATION

To my father

Dr. Elnageeb Adam Gammereldin

and to my mother

Miss. Widad Ahmed Yahia

To my brother and my sisters

To my teachers

To my friends

Salma Elnageeb

ACKNOLEDGMENT

I would like to express my deepest gratitude to my supervisor Dr. Mohamed Elhafiz Mustafa for his continual, guidance detail comments on nearly every page on this thesis. His valuable suggestions and unlimited support through the study were the major factores in bringing this work into existence.

Special respect and thanks to my teachers and my best friends for support and help and prayers.

Last but not least , words are not enough to express my thanks to my parents, sisters and brother for their encouragement ,support, patience and endless love.

Abstract

The massive increase of spam is posing a very serious threat to email which has become an important means for communication. Not only it annoys users, but it also consumes much of the bandwidth of the Internet. Current spam filters are based on the contents of the email one way or the other. In this thesis we present a social network-based spam detection method in which the core idea is using social network measurements as feature to be used by classifier. Two separate classification models have been designed and tested. The first is k-Nearest-Neighbor Classifiers (KNN) classifier and the second is Locally weighted learning (LWL). The experimental results have shown a great favour of using KNN model for spam detection. However, it classifies many legitimate as spam which may annoy the email user. Hence we recommend this model to be applied where the acceptance of a spam message is more danger than legitimate messages rejection. While the classification result of LWL is better than KNN result. It is clear that KNN has advantage of detecting all spammer.

المستخلص

الزيادة الهائلة فى أعداد رسائل البريد الالكترونى المزعجة شكلت تهديدا خطيرا لمستخدمى البريد الإلكتروني الذي أصبح وسيلة مهمة للإتصال. والمشكلة ليست ف قط فى إزعاج المستخدمين، ولكنها تستهلك الكثير من عرض النطاق الترددي للإنترنت. تستند المرشاحات التى تستخدم حاليا للحد من الرسائل غير المرغوب فيها على تحليل المحتوى. يقترح البحث التعرف على مرسلى الرسائل أنفسهم بالتعرف على سلوكهم. لهذا الغرض استخدم البحث الشبكات الاجتماعية للكشف عن مرسلى الرسائل غير المرغوب فيها. الفكرة الأساسية هي بناء شبكة اجتماعية ومن ثم تحليلها. تم بناء نموذجين منفصلين للتصنيف. الأول هو المصنف كى الجار الا قرب (K-Nearest-Neighbor (KNN والثاني هو تعليم الوزن المحلى (Locally Weighted learning (LWL). و قد أظهرت النتائج التجريبية ان التصنيف بإستخدام نموذج كى الجار الا قرب (K-Nearest-Neighbor (KNN للكشف عن الرسائل غير المرغوب فيها نجح فى الكشف على كل مرسلى الرسائل غير المرغوب فيها . ومع ذلك ، فإنه يصنف العديد من المرسلين الشرعين على انهم مزعجين و غير شرعين. هذا يؤدى الى ازعاج مستخدمى البريد الالكترونى لما فية من رفض لرسائل شرعية. ومن هنا فإننا ننصح أن يكون تطبيق هذا النموذج حيث قبول رسالة البريد المزعجة هو أكثر خطرا من رفض الرسائل المشروعة خطأ.
في حين أن النتيجة تصنيف تعليم الوزن المحلى (Locally Weighted learning (LWL أفضل من نتيجة كى الجار الا قرب (K-Nearest-Neighbor (KNN. فمن الواضح أن كى الجار الا قرب (K-Nearest- Neighbor (KNNيتميز بانة نجح بالكشف عن كل مرسلي الرسائل المزعجة.

Table of contents

List of figures

List of tables