## NAME

bogotune - find optimum parameter settings for bogofilter

## SYNOPSIS

**bogotune** [-v] [-c *config*] [-C] [-d *dir*] [-D] [-r *value*] [-T *value*] -n *okfile* [[-n] *okfile* [...]]
           -s *spamfile* [[-s] *spamfile* [...]] [-M *file*]

**bogotune** [-h]

## DESCRIPTION

Bogotune tries to find optimum parameter settings for bogofilter. It needs at least one set each of spam and non-spam messages. The production wordlist is normally used, but it can be directed to read a different wordlist, or to build its own from half the supplied messages.

In order to produce useful results, bogotune has minimum message count requirements. The wordlist it uses must have at least 2,000 spam and 2,000 non-spam in it and the message files must contain at least 500 spam and 500 non-spam messages. Also, the ratio of spam to non-spam should be in the range 0.2 to 5. If you direct bogotune to build its own wordlist, it will use the half the input or 2000 messages (whichever is larger) for the wordlist.

Message files may be in mbox, maildir, or MH folder or any combination. Msg-count files can also be used, but not mixed with other formats.

## OPTIONS

The **-h** option prints the help message and exits.

The **-v** option increases the verbosity level. Level 1 displays the scan output in detail instead of using a progress meter.

The **-c** *filename* option tells bogofilter to read the config file named.

The **-C** option prevents bogotune from reading a configuration file.

The **-d** *dir* option specifies the directory for the database. See the ENVIRONMENT section for other directory setting options.

The **-D** option tells bogotune to build a wordlist in memory using the input messages. The messages are read and divided into two groups. The first group is used to build a wordlist (in ram) and the second is used for tuning. To meet the minimum requirements of 2000 messages in the wordlist and 500 messages for testing, when **-D** is used, there must be 2500 non-spam and 2500 spam in the input files.

If there are enough messages (more than 4000), they will be split evenly between wordlist and testing. Otherwise, they will be split proportionately.

The **-n** option tells bogotune that the following argument is a file (or folder) containing non-spam. Since version 1.0.3, multiple arguments to the **-n** option can be given. All non-option arguments until the next **-s** option will be treated as though they had been preceded by **-n**

The **-s** option tells bogotune that the following argument is a file (or folder) containing spam. It can be repeated as often as necessary. Since version 1.0.3, multiple arguments to the **-s** can be given. All non-option arguments until the next **-n** option will be treated as though they had been preceded by **-s**.

The **-r** *value* option tells bogotune to use the following parameter as the robx value.

The **-T** *value* option tells bogotune to use the following parameter as fp target value.

The **-M** *file* option tells bogotune to convert the file to message count format. This format provides a sorted list of each message's unique tokens, along with their ham and spam counts. Sorting hides the sense of the messages quite effectively, thus protecting privacy. The message-count format allows bogotune and bogofilter to score messages quickly without needing the original token database.

## ENVIRONMENT

Bogofilter uses a database directory, which can be set in the config file. If not set there, bogofilter will use the value of **BOGOFILTER_DIR**. Both can be overridden by the **-d***dir* option. If none of that is available, bogofilter will use directory $HOME/.bogofilter.

## BUGS

Bogotune is not particularly robust when presented with garbage input.

## AUTHOR

The bogofilter developer team.

For updates, see **the bogofilter project page**[1].

## SEE ALSO

bogofilter(1), bogolexer(1), bogoupgrade(1), bogoutil(1)

## NOTES

1.   the bogofilter project page
     http://bogofilter.sourceforge.net/