

A new feature selection algorithm based on HGSO and EO Algorithms for Twitter spam detection

Khaoula Zineb LEGOUI^{1,2}, Sofiane MAZA^{1,2},
Abdelouahab ATTIA^{1,2}

¹Department of Computer Science, University of Mohamed El Bachir El Ibrahimi, El-Anasser, Bordj Bou Arreridj, 34030, Bordj Bou Arreridj, Algeria.

²Laboratory LMSE (Laboratoire Matériaux et Système Electronique).

Contributing authors: khaoulazineb.legoui@univ-bba.dz;
sofiane.maza@univ-bba.dz; abdelouahab.attia@univ-bba.dz;

Abstract

The rapid spread and daily usage of social networks have made them vulnerable to spammers. Therefore, detecting and eliminating spam and spammers has become more than necessary to reduce the risks that it poses to users' security. In order to achieve this goal, it is crucial to determine the exact features that help identify and classify whether a user is spam or not. This paper proposes a wrapper-based method for selecting the most important features. It is based on combining two recent metaheuristic algorithms, the Henry Gas Solubility Optimisation Algorithm (HGSO) and the Equilibrium Optimizer Algorithm (EO), with the goal of choosing a small and most influential subset of features that give good performance and help in the spammer profile detection process. For the purpose of showing the ability of the proposed method to achieve the desired goals, several comparisons are conducted on a modified Social Honeypot dataset. The first comparison is made between HGSOEO and the two algorithms (HGSO and EO) that were used to build the proposed algorithm to prove the power of hybridization. The two next comparisons are performed against some classical filter- and wrapper-based feature selection methods. The last comparison is carried out against some well-known metaheuristic algorithms for feature selection. Experiments and analysis of the results show that the proposed model is more accurate than the algorithms and methods that we compared it to.

Keywords: keyword1, Keyword2, Keyword3, Keyword4

1 Introduction

Online Social Networks (OSNs) have become the most widely used tools in the daily lives of millions of users due to the various services they offer, such as communication, interaction, and even commercial and governmental ones [1][2]. Among the OSNs is Twitter, a social networking service that had 368 million monthly active users as of 2022, as reported by Statista [3]. The extensive availability and usability are the two most important characteristics of Twitter, which have made it vulnerable to spammers, making it an unsafe environment.

Spammers are a major issue on social networks since they may utilize their accounts for a variety of purposes, such as malicious intent, posting unwanted information, spreading misinformation, creating malicious links or misdirection URLs, or stealing users' information [4][5]. These crimes affect both the Twitter platform and the privacy of users, which makes the protection of these platforms essential [6]. Manual detection of them is very difficult, if not impossible. That's why studying their behavior and choosing the features that are critical in identifying spammers' profiles represents the most important step towards protecting both the platform and the users; therefore, using feature selection (FS) is the most important way to reach this purpose.

The feature selection process aims to keep only the features that contribute highly to the detection process by eliminating irrelevant and redundant features that may mislead the detection. Based on the applications of the learning algorithms, FS has been divided into two main methods: wrapper and filter approaches [7][8]. Since feature selection is an NP-hard problem [9], many researchers rely on metaheuristic algorithms, which have proven effective in solving and finding solutions to a wide range of tasks and optimization problems, including FS.

2 Results

Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text.

3 This is an example for first level head—section head

3.1 This is an example for second level head—subsection head

3.1.1 This is an example for third level head—subsubsection head

Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text.

4 Equations

Equations in L^AT_EX can either be inline or on-a-line by itself (“display equations”). For inline equations use the $\$...\$$ commands. E.g.: The equation $H\psi = E\psi$ is written via the command `\$H \psi = E \psi\$`.

For display equations (with auto generated equation numbers) one can use the `equation` or `align` environments:

$$\|\tilde{X}(k)\|^2 \leq \frac{\sum_{i=1}^p \|\tilde{Y}_i(k)\|^2 + \sum_{j=1}^q \|\tilde{Z}_j(k)\|^2}{p+q}. \quad (1)$$

where,

$$\begin{aligned} D_\mu &= \partial_\mu - ig \frac{\lambda^a}{2} A_\mu^a \\ F_{\mu\nu}^a &= \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + gf^{abc} A_\mu^b A_\nu^a \end{aligned} \quad (2)$$

Notice the use of `\nonumber` in the `align` environment at the end of each line, except the last, so as not to produce equation numbers on lines where no equation numbers are required. The `\label{}` command should only be used at the last line of an `align` environment where `\nonumber` is not used.

$$Y_\infty = \left(\frac{m}{\text{GeV}}\right)^{-3} \left[1 + \frac{3 \ln(m/\text{GeV})}{15} + \frac{\ln(c_2/5)}{15}\right] \quad (3)$$

The class file also supports the use of `\mathbb{}`, `\mathscr{}` and `\mathcal{}` commands. As such `\mathbb{R}`, `\mathscr{R}` and `\mathcal{R}` produces \mathbb{R} , \mathscr{R} and \mathcal{R} respectively (refer Subsubsection 3.1.1).

5 Tables

Tables can be inserted via the `normal table` and `tabular` environment. To put footnotes inside tables you should use `\footnotetext[...]` tag. The footnote appears just below the table itself (refer Tables 1 and 2). For the corresponding footnotemark use `\footnotemark[...]`

Table 1 Caption text

Column 1	Column 2	Column 3	Column 4
row 1	data 1	data 2	data 3
row 2	data 4	data 5 ¹	data 6
row 3	data 7	data 8	data 9 ²

Source: This is an example of table footnote. This is an example of table footnote.

¹Example for a first table footnote. This is an example of table footnote.

²Example for a second table footnote. This is an example of table footnote.

The input format for the above table is as follows:

```

\begin{table}[<placement-specifier>]
\caption{<table-caption>}\label{<table-label>}%
\begin{tabular}{@{}llll@{}}
\toprule
Column 1 & Column 2 & Column 3 & Column 4\\
\midrule
row 1 & data 1 & data 2 & data 3 \\
row 2 & data 4 & data 5\footnotemark[1] & data 6 \\
row 3 & data 7 & data 8 & data 9\footnotemark[2]\\
\botrule
\end{tabular}
\footnotetext{Source: This is an example of table footnote.
This is an example of table footnote.}
\footnotetext[1]{Example for a first table footnote.
This is an example of table footnote.}
\footnotetext[2]{Example for a second table footnote.
This is an example of table footnote.}
\end{table}

```

Table 2 Example of a lengthy table which is set to full textwidth

Project	Element 1 ¹			Element 2 ²		
	Energy	σ_{calc}	σ_{expt}	Energy	σ_{calc}	σ_{expt}
Element 3	990 A	1168	1547 ± 12	780 A	1166	1239 ± 100
Element 4	500 A	961	922 ± 10	900 A	1268	1092 ± 40

Note: This is an example of table footnote. This is an example of table footnote this is an example of table footnote this is an example of table footnote this is an example of table footnote.

¹Example for a first table footnote.

²Example for a second table footnote.

In case of double column layout, tables which do not fit in single column width should be set to full text width. For this, you need to use `\begin{table*} ... \end{table*}` instead of `\begin{table} ... \end{table}` environment. Lengthy tables which do not fit in textwidth should be set as rotated table. For this, you need to use `\begin{sidewaystable} ... \end{sidewaystable}` instead of `\begin{table*} ... \end{table*}` environment. This environment puts tables rotated to single column width. For tables rotated to double column width, use `\begin{sidewaystable*} ... \end{sidewaystable*}`.

6 Figures

As per the \LaTeX standards you need to use eps images for \LaTeX compilation and pdf/jpg/png images for PDF \LaTeX compilation. This is one of the major difference between \LaTeX and PDF \LaTeX . Each image should be from a single input .eps/vector image file. Avoid using subfigures. The command for inserting images for \LaTeX and PDF \LaTeX can be generalized. The package used to insert images in \LaTeX /PDF \LaTeX is the graphicx package. Figures can be inserted via the normal figure environment as shown in the below example:

```
\begin{figure}[<placement-specifier>]
\centering
\includegraphics{<eps-file>}
\caption{<figure-caption>}\label{<figure-label>}
\end{figure}
```



Fig. 1 This is a widefig. This is an example of long caption this is an example of long caption this is an example of long caption this is an example of long caption

In case of double column layout, the above format puts figure captions/images to single column width. To get spanned images, we need to provide `\begin{figure*} ... \end{figure*}`.

For sample purpose, we have included the width of images in the optional argument of `\includegraphics` tag. Please ignore this.

7 Algorithms, Program codes and Listings

Packages `algorithm`, `algorithmicx` and `algpseudocode` are used for setting algorithms in \LaTeX using the format:

Table 3 Tables which are too long to fit, should be written using the “sidewaystable” environment as shown here

Projectile	Element 1 ¹			Element ²		
	Energy	σ_{calc}	σ_{expt}	Energy	σ_{calc}	σ_{expt}
Element 3	990 A	1168	1547 \pm 12	780 A	1166	1239 \pm 100
Element 4	500 A	961	922 \pm 10	900 A	1268	1092 \pm 40
Element 5	990 A	1168	1547 \pm 12	780 A	1166	1239 \pm 100
Element 6	500 A	961	922 \pm 10	900 A	1268	1092 \pm 40

Note: This is an example of table footnote this is an example of table footnote this is an example of table footnote this is an example of table footnote
this is an example of table footnote.

¹This is an example of table footnote.

```

\begin{algorithm}
\caption{<alg-caption>}\label{<alg-label>}
\begin{algorithmic}[1]
. . .
\end{algorithmic}
\end{algorithm}

```

You may refer above listed package documentations for more details before setting `algorithm` environment. For program codes, the “verbatim” package is required and the command to be used is `\begin{verbatim} ... \end{verbatim}`.

Similarly, for listings, use the `listings` package. `\begin{lstlisting} ... \end{lstlisting}` is used to set environments similar to `verbatim` environment. Refer to the `lstlisting` package documentation for more details.

A fast exponentiation procedure:

```

begin
  for i:=1 to 10 step 1 do
    expt(2,i);
    newline() od
where
proc expt(x,n) ≡
  z:=1;
  do if n=0 then exit fi;
  do if odd(n) then exit fi;
    comment: This is a comment statement;
    n:=n/2; x:=x*x od;
  { n>0 };
  n:=n-1; z:=z*x od;
  print(z).
end

```

Comments will be set flush to the right margin

```

for i:=maxint to 0 do
begin
{ do nothing }
end;
Write('Case_insensitive_');
Write('Pascal_keywords. ');

```

8 Cross referencing

Environments such as `figure`, `table`, `equation` and `align` can have a label declared via the `\label{#label}` command. For figures and table environments use the `\label{}` command inside or just below the `\caption{}` command. You can then use the `\ref{#label}` command to cross-reference them. As an example, consider the label

Algorithm 1 Calculate $y = x^n$

Require: $n \geq 0 \vee x \neq 0$ **Ensure:** $y = x^n$

```
1:  $y \leftarrow 1$ 
2: if  $n < 0$  then
3:    $X \leftarrow 1/x$ 
4:    $N \leftarrow -n$ 
5: else
6:    $X \leftarrow x$ 
7:    $N \leftarrow n$ 
8: end if
9: while  $N \neq 0$  do
10:  if  $N$  is even then
11:     $X \leftarrow X \times X$ 
12:     $N \leftarrow N/2$ 
13:  else[ $N$  is odd]
14:     $y \leftarrow y \times X$ 
15:     $N \leftarrow N - 1$ 
16:  end if
17: end while
```

declared for Figure 1 which is `\label{fig1}`. To cross-reference it, use the command `Figure \ref{fig1}`, for which it comes up as “Figure 1”.

To reference line numbers in an algorithm, consider the label declared for the line number 2 of Algorithm 1 is `\label{algn2}`. To cross-reference it, use the command `\ref{algn2}` for which it comes up as line 2 of Algorithm 1.

8.1 Details on reference citations

Standard \LaTeX permits only numerical citations. To support both numerical and author-year citations this template uses `natbib` \LaTeX package. For style guidance please refer to the template user manual.

Here is an example for `\cite{...}`: [10]. Another example for `\citep{...}`: [11]. For author-year citation mode, `\cite{...}` prints Jones et al. (1990) and `\citep{...}` prints (Jones et al., 1990).

All cited bib entries are printed at the end of this article: [12], [13], [14], [15], [16], [17], [18], [19], [20], [21] and [22].

9 Examples for theorem like environments

For theorem like environments, we require `amsthm` package. There are three types of predefined theorem styles exists—`thmstyleone`, `thmstyletwo` and `thmstylethree`

<code>thmstyleone</code>	Numbered, theorem head in bold font and theorem text in italic style
<code>thmstyletwo</code>	Numbered, theorem head in roman font and theorem text in italic style
<code>thmstylethree</code>	Numbered, theorem head in bold font and theorem text in roman style

For mathematics journals, theorem styles can be included as shown in the following examples:

Theorem 1 (Theorem subhead). *Example theorem text. Example theorem text. Example theorem text. Example theorem text. Example theorem text. Example theorem text. Example theorem text. Example theorem text. Example theorem text. Example theorem text.*

Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text.

Proposition 2. *Example proposition text. Example proposition text. Example proposition text. Example proposition text. Example proposition text. Example proposition text. Example proposition text. Example proposition text.*

Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text.

Example 1. *Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem.*

Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text.

Remark 1. *Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem.*

Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text.

Definition 1 (Definition sub head). *Example definition text. Example definition text. Example definition text. Example definition text. Example definition text. Example definition text.*

Additionally a predefined “proof” environment is available: `\begin{proof}` ... `\end{proof}`. This prints a “Proof” head in italic font style and the “body text” in roman font style with an open square at the end of each proof environment.

Proof. Example for proof text. Example for proof text. Example for proof text. Example for proof text. Example for proof text. Example for proof text. Example for proof text. \square

Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text. Sample body text.

[illegible]

For a quote environment, use `\begin{quote}...\end{quote}`

Quoted text example. Aliquam porttitor quam a lacus. Praesent vel arcu ut tortor cursus volutpat. In vitae pede quis diam bibendum placerat. Fusce elementum convallis neque. Sed dolor orci, scelerisque ac, dapibus nec, ultricies ut, mi. Duis nec dui quis leo sagittis commodo.

Sample body text. Sample body text. Sample body text. Sample body text. Sample body text (refer Figure 1). Sample body text. Sample body text. Sample body text (refer Table 3).

10 Methods

Topical subheadings are allowed. Authors must ensure that their Methods section includes adequate experimental and characterization data necessary for others in the field to reproduce their work. Authors are encouraged to include RIIDs where appropriate.

Ethical approval declarations (only required where applicable) Any article reporting experiment/s carried out on (i) live vertebrate (or higher invertebrates), (ii) humans or (iii) human samples must include an unambiguous statement within the methods section that meets the following requirements:

1. Approval: a statement which confirms that all experimental protocols were approved by a named institutional and/or licensing committee. Please identify the approving body in the methods section
2. Accordance: a statement explicitly saying that the methods were carried out in accordance with the relevant guidelines and regulations
3. Informed consent (for experiments involving humans or human tissue samples): include a statement confirming that informed consent was obtained from all participants and/or their legal guardian/s

If your manuscript includes potentially identifying patient/participant information, or if it describes human transplantation research, or if it reports results of a clinical trial then additional information will be required. Please visit (<https://www.nature.com/nature-research/editorial-policies>) for Nature Portfolio journals, (<https://www.springer.com/gp/authors-editors/journal-author/journal-author-helpdesk/publishing-ethics/14214>) for Springer Nature journals, or (<https://www.biomedcentral.com/getpublished/editorial-policies#ethics+and+consent>) for BMC.

11 Discussion

Discussions should be brief and focused. In some disciplines use of Discussion or ‘Conclusion’ is interchangeable. It is not mandatory to use both. Some journals prefer a

section ‘Results and Discussion’ followed by a section ‘Conclusion’. Please refer to Journal-level guidance for any specific requirements.

12 Conclusion

Conclusions may be used to restate your hypothesis or research question, restate your major findings, explain the relevance and the added value of your work, highlight any limitations of your study, describe future directions for research and recommendations.

In some disciplines use of Discussion or ‘Conclusion’ is interchangeable. It is not mandatory to use both. Please refer to Journal-level guidance for any specific requirements.

Supplementary information. If your article has accompanying supplementary file/s please state so here.

Authors reporting data from electrophoretic gels and blots should supply the full unprocessed scans for key as part of their Supplementary information. This may be requested by the editorial team/s if it is missing.

Please refer to Journal-level guidance for any specific requirements.

Acknowledgments. Acknowledgments are not compulsory. Where included they should be brief. Grant or contribution numbers may be acknowledged.

Please refer to Journal-level guidance for any specific requirements.

Declarations

Some journals require declarations to be submitted in a standardised format. Please check the Instructions for Authors of the journal to which you are submitting to see if you need to complete this section. If yes, your manuscript must contain the following sections under the heading ‘Declarations’:

- Funding
- Conflict of interest/Competing interests (check journal-specific guidelines for which heading to use)
- Ethics approval
- Consent to participate
- Consent for publication
- Availability of data and materials
- Code availability
- Authors’ contributions

If any of the sections are not relevant to your manuscript, please include the heading and write ‘Not applicable’ for that section.

Editorial Policies for:

Springer journals and proceedings: <https://www.springer.com/gp/editorial-policies>

Nature Portfolio journals: <https://www.nature.com/nature-research/editorial-policies>

Scientific Reports: <https://www.nature.com/srep/journal-policies/editorial-policies>

BMC journals: <https://www.biomedcentral.com/getpublished/editorial-policies>

Appendix A Section title of first appendix

An appendix contains supplementary information that is not an essential part of the text itself but which may be helpful in providing a more comprehensive understanding of the research problem or it is information that is too cumbersome to be included in the body of the paper.

References

- [1] Arora, A., Bansal, S., Kandpal, C., Aswani, R., Dwivedi, Y.: Measuring social media influencer index-insights from facebook, twitter and instagram. *Journal of retailing and consumer services* **49**, 86–101 (2019)
- [2] Bhat, S.Y., Abulaish, M.: Community-based features for identifying spammers in online social networks. In: 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013), pp. 100–107 (2013). IEEE
- [3] Statista Research Department: Number of Twitter users worldwide from 2019 to 2024(in millions). <https://www.statista.com/statistics/303681/twitter-users-worldwide/>, Last accessed on 2023-01-19 (2022)
- [4] Wang, A.H.: Detecting spam bots in online social networking sites: a machine learning approach. In: IFIP Annual Conference on Data and Applications Security and Privacy, pp. 335–342 (2010). Springer
- [5] Stringhini, G., Kruegel, C., Vigna, G.: Detecting spammers on social networks. In: Proceedings of the 26th Annual Computer Security Applications Conference, pp. 1–9 (2010)
- [6] Sohrabi, M.K., Karimi, F.: A feature selection approach to detect spam in the facebook social network. *Arabian Journal for Science and Engineering* **43**(2), 949–958 (2018)
- [7] Salappa, A., Doumpos, M., Zopounidis, C.: Feature selection algorithms in classification problems: an experimental evaluation. *Optimisation Methods and Software* **22**(1), 199–212 (2007)
- [8] Maza, S., Touahria, M.: Feature selection for intrusion detection using new multi-objective estimation of distribution algorithms. *Applied Intelligence* **49**(12), 4237–4257 (2019)

- [9] Kohavi, R., John, G.H.: Wrappers for feature subset selection. *Artificial intelligence* **97**(1-2), 273–324 (1997)
- [10] Campbell, S.L., Gear, C.W.: The index of general nonlinear DAES. *Numer. Math.* **72**(2), 173–196 (1995)
- [11] Slifka, M.K., Whitton, J.L.: Clinical implications of dysregulated cytokine production. *J. Mol. Med.* **78**, 74–80 (2000) <https://doi.org/10.1007/s001090000086>
- [12] Hamburger, C.: Quasimonotonicity, regularity and duality for nonlinear systems of partial differential equations. *Ann. Mat. Pura. Appl.* **169**(2), 321–354 (1995)
- [13] Geddes, K.O., Czapor, S.R., Labahn, G.: *Algorithms for Computer Algebra*. Kluwer, Boston (1992)
- [14] Broy, M.: Software engineering—from auxiliary to key technologies. In: Broy, M., Denert, E. (eds.) *Software Pioneers*, pp. 10–13. Springer, New York (1992)
- [15] Seymour, R.S. (ed.): *Conductive Polymers*. Plenum, New York (1981)
- [16] Smith, S.E.: Neuromuscular blocking drugs in man. In: Zaimis, E. (ed.) *Neuromuscular Junction. Handbook of Experimental Pharmacology*, vol. 42, pp. 593–660. Springer, Heidelberg (1976)
- [17] Chung, S.T., Morris, R.L.: Isolation and characterization of plasmid deoxyribonucleic acid from *Streptomyces fradiae*. Paper presented at the 3rd international symposium on the genetics of industrial microorganisms, University of Wisconsin, Madison, 4–9 June 1978 (1978)
- [18] Hao, Z., AghaKouchak, A., Nakhjiri, N., Farahmand, A.: Global integrated drought monitoring and prediction system (GIDMaPS) data sets. figshare <https://doi.org/10.6084/m9.figshare.853801> (2014)
- [19] Babichev, S.A., Ries, J., Lvovsky, A.I.: Quantum scissors: teleportation of single-mode optical states by means of a nonlocal single photon. Preprint at <https://arxiv.org/abs/quant-ph/0208066v1> (2002)
- [20] Beneke, M., Buchalla, G., Dunietz, I.: Mixing induced CP asymmetries in inclusive B decays. *Phys. Lett.* **B393**, 132–142 (1997) [arXiv:0707.3168](https://arxiv.org/abs/hep-th/9707033) [gr-qc]
- [21] Stahl, B.: DeepSIP: Deep Learning of Supernova Ia Parameters, 0.42, *Astrophysics Source Code Library* (2020), [ascl:2006.023](https://ui.adsabs.org/abs/2020ascl..0000000Stal)
- [22] Abbott, T.M.C., *et al.*: Dark Energy Survey Year 1 Results: Constraints on Extended Cosmological Models from Galaxy Clustering and Weak Lensing. *Phys. Rev. D* **99**(12), 123505 (2019) <https://doi.org/10.1103/PhysRevD.99.123505> [arXiv:1810.02499](https://arxiv.org/abs/1810.02499) [astro-ph.CO]