**A Fast Abnormal Data Cleaning Algorithm for**

**Performance Evaluation of Wind Turbine**

Zhongju Wang , Long Wang , Member, IEEE, and Chao Huang , Member, IEEE

**Abstract.** In this paper a method to remove abnormal data from wind turbines is proposed. This method is divided in data cleaning and data classification. In the former, pixels corresponding to normal data are extracted based on the characteristics of normal and abnormal data. In the later step images are classified based on the existence of the corresponding pixels. The data cleaning is implemented parallelly using CUDA. The proposed method’s effectivity is validated using 37 commercial wind turbines. This paper’s results show an increase in performance cleaning abnormal data while reducing computational time.

**Keywords:** Abnormal data cleaning, data-driven approaches, wind power curve (WPC), wind turbines.

1 Introduction

Monitoring wind turbine power generation performance is conducive to arranging maintenance plans reasonably, preventing failure occurrences, and minimizing O&M cost. To measure the wind turbine power generation performance, accurate wind power curve modeling is widely applied. With the development of information technologies, current wind turbines are generally equipped with data acquisition and monitoring systems. Supervisory control and data acquisition systems that accumulate vast historical data of wind turbines are widely used for various condition monitoring tasks of wind turbines, such as fault diagnosis and failure detection. WPC modeling-based methods are another main type of method for abnormal data cleaning. Based on the spatial distribution characteristics, wind power data points distributed outside the boundary of WPC are considered as outliers. Verma utilized a five-year historical data set to fit the WPC and applied the k-means clustering algorithm to distinguish normal wind power data points and outliers. Identified abnormal wind power data points based on the probabilistic WPC and corrected the detected outliers via the spatial correlation characteristics of neighboring wind farms. Considering abnormal data affects the error distribution in WPC modeling, Wang et al. proposed two novel asymmetric spline regression models, mixture of asymmetric Gaussian and mixture of asymmetric exponential power, for solving the problem of complex and asymmetric error distribution in WPC modeling. However, these methods required a selected training data set to fit WPC models so that abnormal data cleaning performance was influenced by the selection of the training data set. In the proposed method, the principal part of the WPC image was extracted via MMO, and the mapping relationship between wind power data points and WPC image pixels was established to mark the normal and abnormal data. Here is proposed an image thresholding method for WPC abnormal data cleaning. The proposed method filtered the abnormal data by thresholding on a gray level feature image. In conclusion, existing abnormal data cleaning methods for wind power data have the following limitations: 1) high computational overhead is required; 2) cleaning performance is seriously influenced by the selection of training data; and 3) hyperparameters need to be determined for different data sets. Considering the features of WPCs, the data cleaning stage filters abnormal data based on pixel spatial distribution characteristics of abnormal and normal data in WPC images.

2 Paper Preparation

The printing area is 122 mm × 193 mm. The text should be justified to occupy the full line width, so that the right margin is not ragged, with words hyphenated as appropriate. Please fill pages so that the length of the text is no less than 180 mm, if possible.

Use 10-point type for the name(s) of the author(s) and 9-point type for the address(es) and the abstract. For the main text, please use 10-point type and single-line spacing. We recommend the use of Computer Modern Roman or Times. Italic type may be used to emphasize words in running text. Bold type and underlining should be avoided.

Papers not complying with the LNCS style will be reformatted. This can lead to an increase in the overall number of pages. We would therefore urge you not to squash your paper.

Headings. Headings should be capitalized (i.e., nouns, verbs, and all other words except articles, prepositions, and conjunctions should be set with an initial capital) and should, with the exception of the title, be aligned to the left. Words joined by a hyphen are subject to a special rule. If the first word can stand alone, the second word should be capitalized. The font sizes are given in Table 1.

Here are some examples of headings: "Criteria to Disprove Context-Freeness of Collage Languages", "On Correcting the Intrusion of Tracing Non-deterministic Programs by Software", "A User-Friendly and Extendable Data Distribution System", "Multi-flip Networks: Parallelizing GenSAT", "Self-determinations of Man".

**Table 1.** Font sizes of headings. Table captions should always be positioned *above* the tables.

|  |  |  |
| --- | --- | --- |
| Heading level | Example | Font size and style |
| Title (centered) | **Lecture Notes …** | 14 point, bold |
| 1st-level heading | **1 Introduction** | 12 point, bold |
| 2nd-level heading | **2.1 Printing Area** | 10 point, bold |
| 3rd-level heading | **Headings.** Text follows … | 10 point, bold |
| 4th-level heading | *Remark.* Text follows … | 10 point, italic |

Lemmas, Propositions, and Theorems. The numbers accorded to lemmas, propositions, and theorems, etc. should appear in consecutive order, starting with Lemma 1, and not, for example, with Lemma 11.

2.1 Figures

Please check that the lines in line drawings are not interrupted and have a constant width. Grids and details within the figures must be clearly legible and may not be written one on top of the other. Line drawings should have a resolution of at least 800 dpi (preferably 1200 dpi). The lettering in figures should have a height of 2 mm (10-point type). Figures should be numbered and should have a caption which should always be positioned *under* the figures, in contrast to the caption belonging to a table, which should always appear *above* the table. Please center the captions between the margins and set them in 9-point type (Fig. 1 shows an example). The distance between text and figure should be about 8 mm, the distance between figure and caption about 6 mm.

To ensure that the reproduction of your illustrations is of a reasonable quality, we advise against the use of shading. The contrast should be as pronounced as possible.

If screenshots are necessary, please make sure that you are happy with the print quality before you send the files.

Remark 1. In the printed volumes, illustrations are generally black and white (halftones), and only in exceptional cases, and if the author is prepared to cover the extra costs involved, are colored pictures accepted. Colored pictures are welcome in the electronic version free of charge. If you send colored figures that are to be printed in black and white, please make sure that they really are legible in black and white. Some colors show up very poorly when printed in black and white.



**Fig. 1.** One kernel at *xs* (*dotted kernel*) or two kernels at *xi* and *xj* (*left and right*) lead to the same summed estimate at *xs*. This shows a figure consisting of different types of lines. Elements of the figure described in the caption should be set in italics, in parentheses, as shown in this sample caption.

2.2 Formulas

Displayed equations or formulas are centered and set on a separate line (with an extra line or halfline space above and below). Displayed expressions should be numbered for reference. The numbers should be consecutive within each section or within the contribution, with numbers enclosed in parentheses and set on the right margin.

|  |  |
| --- | --- |
| x + y = z . | (**1**) |

Equations should be punctuated in the same way as ordinary text but with a small space before the end punctuation mark.

2.3 Footnotes

The superscript numeral used to refer to a footnote appears in the text either directly after the word to be discussed or – in relation to a phrase or a sentence – following the punctuation mark (comma, semicolon, or period). Footnotes should appear at the bottom of the normal text area, with a line of about 5cm set immediately above them[[1]](#footnote-1).

2.4 Program Code

Program listings or program commands in the text are normally set in typewriter font, e.g., CMTT10 or Courier.

Example of a Computer Program from Jensen K., Wirth N. (1991) Pascal user manual and report. Springer, New York

program Inflation (Output)  
 {Assuming annual inflation rates of 7%, 8%, and  
 10%,... years};  
 const MaxYears = 10;  
 var Year: 0..MaxYears;  
 Factor1, Factor2, Factor3: Real;  
 begin  
 Year := 0;  
 Factor1 := 1.0; Factor2 := 1.0; Factor3 := 1.0;  
 WriteLn('Year 7% 8% 10%'); WriteLn;  
 repeat  
 Year := Year + 1;  
 Factor1 := Factor1 \* 1.07;  
 Factor2 := Factor2 \* 1.08;  
 Factor3 := Factor3 \* 1.10;  
 WriteLn(Year:5,Factor1:7:3,Factor2:7:3,  
 Factor3:7:3)  
 until Year = MaxYears  
end.

2.5 Citations

For citations in the text please use square brackets and consecutive numbers: [1], [2], [3], etc.

2.6 Page Numbering and Running Heads

There is no need to include page numbers. If your paper title is too long to serve as a running head, it will be shortened. Your suggestion as to how to shorten it would be most welcome.

3 LNCS Online

The online version of the volume will be available in LNCS Online. Members of institutes subscribing to the Lecture Notes in Computer Science series have access to all the pdfs of all the online publications. Non-subscribers can only read as far as the abstracts. If they try to go beyond this point, they are automatically asked, whether they would like to order the pdf, and are given instructions as to how to do so.

Please note that, if your email address is given in your paper, it will also be included in the meta data of the online version.

4 BibTeX Entries

The correct BibTeX entries for the Lecture Notes in Computer Science volumes can be found at the following Website shortly after the publication of the book: <http://www.informatik.uni-trier.de/~ley/db/journals/lncs.html>

Acknowledgments. The heading should be treated as a 3rd level heading and should not be assigned a number.

5 The References Section

In order to permit cross referencing within LNCS-Online, and eventually between different publishers and their online databases, LNCS will, from now on, be standardizing the format of the references. This new feature will increase the visibility of publications and facilitate academic research considerably. Please base your references on the examples below. References that don’t adhere to this style will be reformatted by Springer. You should therefore check your references thoroughly when you receive the final pdf of your paper. The reference section must be complete. You may not omit references. Instructions as to where to find a fuller version of the references are not permissible.

We only accept references written using the latin alphabet. If the title of the book you are referring to is in Russian or Chinese, then please write (in Russian) or (in Chinese) at the end of the transcript or translation of the title.

The following section shows a sample reference list with entries for journal articles [1], an LNCS chapter [2], a book [3], proceedings without editors [4] and [5], as well as a URL [6]. Please note that proceedings published in LNCS are not cited with their full titles, but with their acronyms!

References

1. Smith, T.F., Waterman, M.S.: Identification of Common Molecular Subsequences. J. Mol. Biol. 147, 195--197 (1981)

2. May, P., Ehrlich, H.C., Steinke, T.: ZIB Structure Prediction Pipeline: Composing a Complex Biological Workflow through Web Services. In: Nagel, W.E., Walter, W.V., Lehner, W. (eds.) Euro-Par 2006. LNCS, vol. 4128, pp. 1148--1158. Springer, Heidelberg (2006)

3. Foster, I., Kesselman, C.: The Grid: Blueprint for a New Computing Infrastructure. Morgan Kaufmann, San Francisco (1999)

4. Czajkowski, K., Fitzgerald, S., Foster, I., Kesselman, C.: Grid Information Services for Distributed Resource Sharing. In: 10th IEEE International Symposium on High Performance Distributed Computing, pp. 181--184. IEEE Press, New York (2001)

5. Foster, I., Kesselman, C., Nick, J., Tuecke, S.: The Physiology of the Grid: an Open Grid Services Architecture for Distributed Systems Integration. Technical report, Global Grid Forum (2002)

6. National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov

Appendix: Springer-Author Discount

LNCS authors are entitled to a 33.3% discount off all Springer publications. Before placing an order, they should send an email to [orders-HD-individuals@springer.com](mailto:orders-HD-individuals@springer.com), giving full details of their Springer publication, to obtain a so-called token. This token is a number, which must be entered when placing an order via the Internet, in order to obtain the discount.

Checklist of Items to be Sent to Volume Editors

1. A final Word or RTF file
2. A final PDF file
3. A copyright form, signed by one author on behalf of all of the authors of the paper
4. A readme giving the name and email address of the corresponding author

1. The footnote numeral is set flush left and the text follows with the usual word spacing. [↑](#footnote-ref-1)