Review

of the paper entitled

A Novel Change Detection Approach on Content Delivery Network (CDN) using Data Streaming Analytics

submitted by

Indra Awal Priyanto, Wolosz Andras, Kadar Zoltan, Baba Adam

Reviewer: *Fruzsina Habzda*

May 6, 2023

The article subject to this review task is a writing practice with imaginary results, but actual literatury research for the Research methodology class at ELTE Faculty of Informatics.

Recommendation

strong accept / **accept** / weak accept / borderline / weak reject / reject / strong reject

Summary of the Content, Evaluation

The researchers addressed the sudden change detection in the big data paradigm using a scalable machine learning algorithms and efficient solution methodology in both static and streaming contexts. To create scalable ML models, they used Spark MLlib. Firstly, in the static context, they developed machine learning models on the historic CDN data collected from open source. (They employed several machine learning techniques: NB, LR, SVM, DT, RF, GBT, and MLP.) Secondly, in the streaming context, they used DStreams to capture data at regular intervals and employed a sliding window based change detection approach. They performed change detection using binary classification tasks.

Technical Content and Accuracy *(4 points)*

Unfortunately the novel approach is not validated in a proper way. In the results, we can not see the results of the researcher's method, but they compared some already existed methods. They showed that 'RF' is the best of them all. I also miss some

the full name of these methods, more information and references. There were some references in the text to not listed references.
More precisely:

- In the introduction some technical terms are not explained e.g. weighting. (Where do we have to weight and why?). And there are listed many machine learning techniques (NB, LR,SVM, DT, RF, GBT, and MLP) some other techniques (SMOTE-ENN, SMOTE-Tomek, ADASYN), which are also not discussed here, and there's no reference for them.
- There's also no reference, or explanation for 10-Fold CV (Cross Validation).
- "AUC is proven to be a robust measure which is quite evident in previous case studies such as credit card fraud detection" - also no reference
- "Abakiram et al. [50] proposed a real time model for change-point detection based on deep learning." - there is a reference, which you can't find in the list (there aren't 50th reference)
- ect.

Significance of the Work *(5 points)*
The topic is really significant, because the technique can help users to access streamed contents, when the network is overloaded and there are blackouts. The new method is really useful in every day life.

Appropriate Title, Introduction and Conclusion *(4 points)*
The title properly refers to what this paper is about.
The introduction was a bit hard to follow for me. For example inside "Applications" they are writing about "Quality of Service" and "Data Mining". In Quality of Service" they speak about the Change Detection's importance and application for CDN providers, but inside "Data Mining" they don't speak about the application, but the tools that are used for data mining (Spark, Hadoop, Hive, ect.). I've found it misleading.
The conclusion reflects to the content wonderfully.

Overall Organization *(4 points)*
The paper is in IMRaD structure. However, the subtitles are not always referring to the content of the paragraphs (as mentioned above).

Appropriateness for the Conf/Journal *(5 points)*

This paper is about the same topic, as reader would expect. It is about change detection in CDN. The method is data analysis with the help of Spark. Exactly, as the Abstract and the Title suggests.

Style and Clarity of the Paper *(4 points)*

Good. Nicely wrapped text. Good amount of figures. Nicely labeled figures. Lots of subheadings that helps the user to navigate. However, I don't think it is a good idea to break English words, I think it's better to use justified text instead. And the Methodology chapter should have been written in present tense. (The Conclusion is in past tense, which is good as it is.) The references are not generated, it is not automatic (There can be found some numbers referring to non-existent references. e.g. [47])

Originality of the Content *(5 points)*

Change detection in CDN data is seldom a researched topic, so as a topic, it is quite original by itself.

Sum Score *(31 points)*

Novelty of Results

Beyond State of the Art (Related Work)

"Our research article appeared dealing with CDN change detection, where a specialized language is proposed for proactive sudden change-point management in CDN data streams", as the paper declares. So I'd say, it is a novel result in this field.

Appropriateness of the Methods, Validation, References

The method seems appropriate, but they are not validated in a proper way. In the results, we can not see the results of the novel method, but some already existed methods. I also miss some references for these methods and information, and there were some references in the text, which referenced to not listed references.

Comments on Errors, Typos, Grammar, Figures

There were some typos, and grammatically incorrect sentences. But what most concerns me, is that there were references for references, which were not listed

below...

- "It is make sense ..."
- "fulfil"
- "nformation"
- "spark"
- there are several references in the text, that don't appear in the list. e.g.: 31, 27, 29, 49, 48, 50, 47. There aren't that many references listed at the end of the paper. (This appears mostly at the "related work" chapter. Was that text copied from somewhere else?)
- I'm not sure, this sentence is right grammatically: "The proposed methodology is depicted in Fig. 1, and the block diagram is depicted in Fig. 2, and can be used in a static environment, where the model is trained on historical data."
- at the 4th Fig. there's a red line in the middle, which is not part of the diagram.
- There is an incorrect text breaking at "Evaluation Measure". The bold "AUC" should be at the beginning of the next line, I think.
- "... when only one single class-related data point are given." (instead of 'are' use 'is'.)
- "... but also the showed lesser variability which supports its robustness."
- "... a scalable ML models ..."
- "Further, we conducted statistical test which turned out that RF turned out to be statistically significant."

Proposal for Improvements

Please use present tense in "Methodology". Please use justified text in the paper instead of breaking words. Please generate references. Please show reference for statements that was in someone else's research, and not you found it out in this one. I think you should write about the results of your own novel method, and maybe compare them to an already existing one.

Reviewer's Confidence

expert / high / medium / **low** / none