Author's response to reviews

Title: Using n-gram analysis to cluster heartbeat signals

Authors:

Yu-Chen Huang (s978704@mail.yzu.edu.tw)
Hanjun Lin (s978702@mail.yzu.edu.tw)
Yeh-Liang Hsu (mehsu@saturn.yzu.edu.tw)
Jun-Lin Lin (jun@saturn.yzu.edu.tw)

Version: 4 Date: 22 April 2012

Author's response to reviews: see over
Dear Editor,

I truly appreciate the comments from the reviewers. I have carefully revised the paper according to the reviewers’ comments. Please see our response below. Please let me know if there are any questions. Thank you very much.

Sincerely,

Yeh-Liang Hsu, PhD

**Reviewer's report: (1)**
1. Minor Essential Revisions
2. In general, the paper presents a clear concept of using n-grams, symbolization and clustering to improve classification of heartbeat signal collections. However, further studies would be required to show robustness for heartbeat signals as opposed to being viable and realistic with other applications as discussed at the top of pg 13. This was not included in discussion and conclusion.

**Our Response:**
A sentence is added at the end of “Conclusion” section for this future study required (p.17).

3. Abstract - Not necessary for sub-sections

**Our Response:**
There are subsection titles in the abstract in the standard template of BMC papers.

4. CHF (abbreviation expanded in abstract but not in the first instance in main body of paper). Is there an expansion for RR?

**Our Response:**
CHF is expanded (p.4). RR interval refers to Inter-beat interval and is also specified (p.7).

5. Reference numbers in-text requires a thorough format checking as numbers are repeated outside the brackets. on pg 9, is it reference 16 or 17?
Our Response:
All reference numbers are checked again.

6. What is being referred to with 'Some early traditional linear methods …' on top of pg 4
Our Response:
References for the early traditional linear methods are added (p.4).

7. Reference Peng's team in-text as opposed to Peng et. al…. can be changed (to avoid personalization)
Our Response:
"developed by Peng’s team" is deleted to avoid personalization (p.6).

8. Pg 9 – Example 1 : Is the heart rate time series referring to the series of the data collection
Our Response:
This sentence is rewritten into “There are 142 samples of heart rate time series data in this example, which can be divided into 5 groups in this example, including…”, to make it clear (p.11).

9. Pg 10 – 2nd para – Cluster number 20, or is it number of clusters is 20/ k=20. Explanation of results does not include explanation of superiority in performance with CHF.
Our Response:
The cluster number 20 means k=20. This is revised to make it clear (p.12). CHF did not show significant superiority in our result.

10. Pg 11 – since there is repetition of steps in generating the symbolic sequence, can these steps be summarized once and referenced to thereafter with changes in numbers (parameters).
Our Response:
We did summarize the steps in “Method” section, then reference to it in the examples to minimize repetition.

11. Pg 13 – claim that the criterions of minimal total variance is not used in previous studies require reference to the previous studies being implied.
Our Response:
A reference is added (p.16).
Reviewer’s report

Title: Using n-gram analysis to cluster heartbeat signals
Version: 1 Date: 8 April 2012
Reviewer: Chunhua Bian

Reviewer’s report:
The manuscript presents a method named AllA to use the Simple K-Means algorithm for symbolization, which offers a new way to represent subtle variations between two interbeat intervals. Then the authors classify, i). the patients with Atrial Fibrillation (AF), Congestive Heart Failure (CHF) and healthy people, ii). The patients with apnea with AllA method. Six classifiers, including Bayesian Network, Logistic, Naïve Bayesian, Neural Network, Support Vector Matrix (SVM) and Tree-J48 are used. The present version of the manuscript still needs some revision. I would recommend publication provided the authors have considered the remarks I give below.

Major Compulsory Revisions:
1. On page 4, "At present, there are three different approaches for using non-linear symbolic sequences to represent heart beat time series." "The first approach is based on the deviation of the heart rate time series from the local mean", "The second approach is to symbolize the increase or decrease of the momentary heart rate by two different symbols." "The third approach is to divide the range between minimum and maximum heart rate into a few equidistant intervals, each interval is denoted by a unique symbol." The reference was not summarized completely. In Ref(1) and Ref(2) permutation entropy and modified PE method has been studied, which is symbolized by the order of the interval value.

Our Response:
Both references are added (p.6).

2. The databases used in the present paper are from physionet, which have been studied by many researchers. The authors should have a comparison between the results in the presented paper with those of other methods with the same database.

Our Response:
Accuracy was not presented in previous papers using the same database, and therefore cannot be compared. A sentence is added to address this (p.13).

3. What is the data length in the analysis of this paper?

Our Response:
Data lengths of all samples are specified in the tables.

4. On page 8, “This research uses 1-gram, 2-gram and 3-gram for analysis, which includes 18,278 different kinds of string combinations.” How does the number 18,278 been calculated? And does it have relevance with the cluster number k?

Our Response:
An equation is added to describe how 18,278 is calculated and its relation to $k$ (p.10).

5. In the paper the authors declares that both experiments acquired the best category results when using the Bayesian Network. While from the accuracy value in table 2 and table 5, we find no Essential difference between different classifiers.

Our Response:
Table 2 and 5 are corrected. Accuracies at the same $k$ and the best accuracy of each classifier are presented for comparison, which shows that Bayesian Network has better performance among the classifiers.

6. The English expression in the paper need to be improved.

Our Response:
We have made some corrections in English where we can.