
МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ УКРАИНЫ
НАЦИОНАЛЬНЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ УКРАИНЫ
"КИЕВСКИЙ ПОЛИТЕХНИЧЕСКИЙ ИНСТИТУТ"

Electronics and Communications

Електроніка та зв'язок

Электроника и связь

Научно-технический журнал
Издается шесть раз в год
Основан в марте 1995 года

Том 19, №6(83), 2014



Київ
НТУУ «КПІ»
2014

Биомедицинские приборы и системы

UDC 621.391

M.A. Zhukov¹, A.O. Popov¹, Ph.D., V.I. Kharytonov², I.A. Chaikovsky³

¹National Technical University of Ukraine «Kyiv Polytechnic Institute»,
off.423, Politekhnichna Str., 16, Kyiv, 03056, Ukraine.

²TMO «Psychiatry», Kyiv, Ukraine.

³Glushkov Institute of Cybernetics of NAS of Ukraine, Glushkov ave. 40, Kyiv, Ukraine.

Optimal Bin Number Selection for Mutual Information Calculation Between EEG and Cardiorhythmogram Signals

In the present work the problem of optimal bin number selection for equidistant Mutual Information (MI) estimator between electroencephalogram (EEG) and cardiorhythmogram (CRG) is addressed. In the previously developed method the bin number selected based on the finding an optimal bin number on the MI values on the range of bin numbers. With application to the real raw EEG and CRG signals it was found that for closely placed or symmetrical channels of EEG data the method can be applied, and the true value of MI value can be found with proposed method. In application to MI calculation between raw EEG and CRG signals that are not significantly coupled, true MI value cannot be estimated with proposed method for small sample size. Reference 12, figures 4.

Keywords: mutual information, bins, sample size, bin number, probability density estimation.

Abstract

Mutual information (MI) is a measure of the amount of information that one random variable contains about another random variable. It is the reduction in the uncertainty of one random variable due to the knowledge of the other [1]. In order to calculate the MI value between some datasets the knowledge of probability distributions is needed. Most frequently used estimators of probability density function are based on histograms (with fixed or adaptive bin size [6][5]), k-nearest neighbors and kernels [2][8][11][10].

Histograms are used extensively as nonparametric density estimators both to visualize data and to obtain various parameters and characteristics such as entropy, of the underlying density. To estimate the probability distribution the procedure of partitioning the variable values into

discrete bins is used to build histograms. Since histogram-based estimator is used in MI calculation for finite discrete data series, one can anticipate the dependence of calculated MI value on the binning. The general idea is to choose a number of bins sufficiently large to capture the major features in the data while ignoring fine details due to random sampling fluctuations' [4]. But the problem may arise since the accuracy and precision of the probability distribution approximation heavily influences the resulting MI value and may give spurious results.

In the paper [12] the MI between two jointly correlated gaussian datasets was estimated for a wide range of sample size/bin number pairs. In this paper the goal of applying the proposed method on real EEG and CRG signals is pursued. It should be determined how the MI value depends on bin size for real supposedly uncorrelated or somehow coupled signals.

Mutual information

Mutual information is given by the formula:

$$I(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right) \quad (1)$$

where $p(x, y)$ is the joint probability distribution function (PDF) of random variables X and Y , and $p(x)$ and $p(y)$ are the marginal probability distribution functions of X and Y respectively (Fig. 1). When X and Y are statistically independent $p(x, y) = p(x)p(y)$ and thus the value of log expression becomes zero as well as the overall MI. It can be concluded that high value of mutual information between two signals represents a large reduction of uncertainty and zero value indicates the independence of random variables.

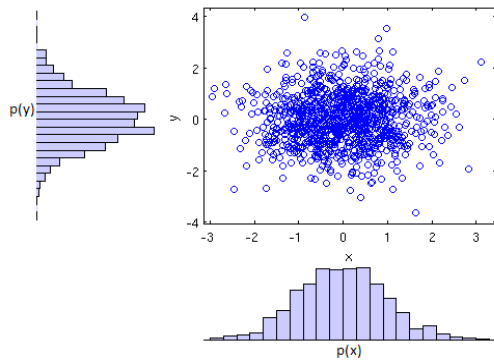


Fig. 1. Marginal density functions of Gaussian distributed random variables X and Y and observations from the joint probability distribution (top, sample size $N = 10^3$) are shown

Method for bin size selection

There are various different methods as well as rules of thumb [9][3][7] that attempt to give the optimal number of bins but usually the shape of the underlying distribution of the data should be known or there are some other limitations that restrict the use of the method.

MI value heavily depends on both the bin size and sample size [12], which could be arbitrary in case of real life applications. Thus the natural question arises about the optimal choice of one parameter given the value of another.

To find the optimal bin number for getting the most correct value of MI, we propose to use the behavior of MI dependence. The main idea of the proposed technique to select the optimal bin number for MI calculation is to choose the number of bins from the range of bin numbers, for which MI does not change much. It is suggested that selection of bin number in the range where change in bin number would not cause the significant change in MI. The algorithm of finding the optimal bin number is

- Calculate difference quotient

$$MD(bin_i) = \frac{MI_{i+1} - MI_i}{bin_{i+1} - bin_i} \text{ for } i = 2.. \frac{N}{100} .$$
- Find second order approximate of the MD using least-squares approximation

$$MD_{appr} = a \cdot bin^2 + b \cdot bin + c .$$
- Find the bin_i where MI starts to rise:

$$MD(bin_i) > 0 .$$
- Find tangent line to this point and its cross-section with bin axes where $MD(bin_i) = 0$.

- Assume that the lower boundary of bin number is represented by the found value of bin .

MI calculation for real EEG and CRG signals

In the experiment simultaneously recorded EEG and ECG signals of 16 healthy children during sleep were chosen. EEGs were recorded from 19 scalp loci (Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2) of the international 10-20 system with reference to ipsilateral ear. One ECG channel recorded from II standard Einthoven's lead. ECGs were used for derivation of RR-intervals for further analysis. All signals were recorded with 250 Hz sampling frequency.

MI values were calculated using proposed method for the range of bins from 2 to 80 and for the range of sample sizes. The behavior of MI values is illustrated for representational signals in Fig. 2-4.

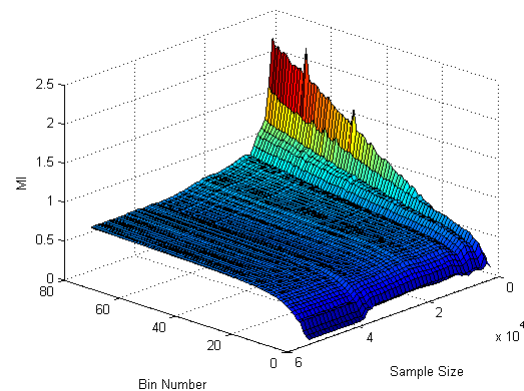


Fig. 2. MI value dependence on bin number and sample size between EEG Fp1 channel and EEG Fp2 channel signals

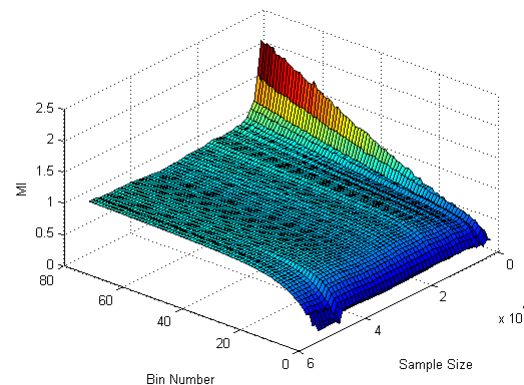


Fig. 3. MI value dependence on bin number and sample size between EEG Fz channel and EEG Fp4 channel signals

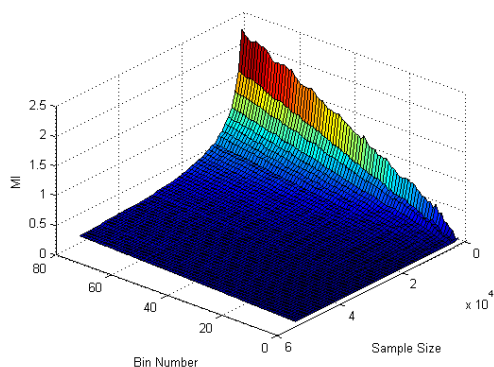


Fig. 4. MI value dependence on bin number and sample size between EEG Fz channel and CRG signals

Discussion

As can be seen from the Fig. 4 for supposedly uncoupled signals (raw CRG and Fz channel of EEG) the dependence of MI value on selected bin number differs from the coupled signals (raw signals of EEG channels Fz-Fp4, Fp1-Fp2). For the coupled signals (Fp1-Fp2 due to the symmetry of brain activity and Fz-Fp4 due to the closely situated electrodes) the true value of MI can be found for small sample size via proposed method as was shown earlier [12] for correlated Gaussian signals.

Conclusions

In the paper the method of bin number selection for mutual information calculation was tested on real EEG and CRG signals. It was shown that for coupled EEG signals the proposed method can be used for small sample size. For the case of non-correlated and supposedly uncoupled signals the dependence of MI value on bin number for small sample sizes gives no possibility for using the proposed method.

References

1. Cover, T.M., Thomas. (1991), "Elements of information theory". New York: Wiley, 1991. - ISBN 0471062596 9780471062592.
2. Doquire, G., Verleysen, M. (2012), "A Comparison of Multivariate Mutual Information Estimators for Feature Selection". SciTePress - Science and Technology Publications, Pp.176–185.
3. Freedman, D., Diaconis, P. (1981), "On the histogram as a density estimator: L 2 theory". Probab. Theory Relat. Fields". Vol. 57. No 4. Pp.453–476.
4. Knuth, K.H. (2006), "Optimal data-based binning for histograms". ArXiv Prepr. Phys.
5. Legg, P.A., Rosin, P.L., Marshall, D., Morgan, J.E. (2007) "Improving accuracy and efficiency of registration by mutual information using Sturges' histogram rule". Proc Med Image Understand Anal. Pp. 26–30.
6. Marek, T., Tichavsky, P. (2008), "On the estimation of mutual information". ROBUST'2008.
7. Shimazaki, H., Shinomoto, S. (2007), "A Method for Selecting the Bin Size of a Time Histogram". Neural Comput. Vol. 19. No 6. Pp.1503–1527.
8. Sorjamaa, A., Hao, J., Lendasse, A. (2005), "Mutual information and k-nearest neighbors approximator for time series prediction". Artificial Neural Networks: Formal Models and Their Applications–ICANN 2005.: Springer, 2005. Pp.553–558.
9. Sturges, H.A. (1926), "The Choice of a Class Interval". J. Am. Stat. Assoc. 1926. Vol. 21. No153. Pp. 65–66.
10. Suzuki, T., Sugiyama, M., Sese, J., Kanamori T. (2008), "Approximating Mutual Information by Maximum Likelihood Density Ratio Estimation". J. Mach. Learn. Res.-Proc. Track. Vol. 4. Pp. 5–20.
11. Verleysen, M. (2008), "K-Nearest Neighbours based on Mutual Information for Incomplete Data Classification". Bruges, Belgium.
12. Zhukov, M., Popov, A. (2014), "Bin Number Selection for Equidistant Mutual Information Estimator". Kyiv, Ukraine. Pp. 259–263.

Поступила в редакцію 20 сентября 2014 г.

УДК 621.391

М.А. Жуков¹, А.О. Попов¹, канд.тех.наук, В.І. Харитонов², І.А. Чайковський³

¹Національний технічний університет України «Київський політехнічний інститут», вул. Політехнічна, 16, 03056, Київ, Україна.

²ТМО "Psychiatry", Київ, Україна.

³Інститут кібернетики імені В. М. Глушкова НАН України, просп. Академіка Глушкова, 40, Київ, Україна.

Вибір оптимальної кількості бінів для розрахунку взаємної інформації між сигналами ЕЕГ та кардіоритмограми

В роботі розглянуто метод визначення взаємної інформації для оцінки взаємозв'язку між сигналами ЕЕГ та кардіоритмограми. У розробленому методі кількість бінів вибирається базуючись на значеннях взаємної інформації, які розраховані на проміжку значень бінів. При застосуванні методу до сигналів ЕЕГ та КРГ було знайдено, що метод може бути застосований при аналізі взаємозв'язку між сигналами ЕЕГ в каналах, що розташовані поряд або симетрично, згідно з системою 10-20. Для взаємної інформації між сигналами ЕЕГ та КРГ, що пов'язані у значно меншій мірі, метод не може бути застосований для малого обсягу вибірки. Бібл. 12, рис. 4.

Ключові слова: ЕКГ, варіабельність, ЕЕГ, взаємна інформація, кардіоритмограма.

УДК 621.391

М.А. Жуков¹, А.А. Попов¹, канд.тех.наук, В.И. Харитонов², И.А. Чайковский³

¹Национальный технический университет Украины «Киевский политехнический институт», ул. Политехническая, 16, 03056, Киев, Украина.

²ТМО «Psychiatry», Киев, Украина.

³Институт кибернетики імені В. М. Глушкова НАН України, просп. Академика Глушкова. 40, Киев, Украина.

Выбор оптимального количества бинов для расчета взаимной информации между сигналами ЭЭГ и кардиоритмограммы

В работе рассмотрен метод определения взаимной информации для оценки взаимосвязи между сигналами ЭЭГ и кардиоритмограммы. В разработанном методе количество бинов выбирается на промежутке значений бинов. При применении метода к сигналам ЭЭГ и КРГ было найдено, что метод может быть применим при анализе взаимосвязи между сигналами ЭЭГ в каналах, что расположены рядом или симметрично, в соответствии с системой 10-20. Для взаимной информации между сигналами ЭЭГ и КРГ, что связаны в значительно меньшей мере, метод не может применяться для малого значения выборки. Библ. 12, рис. 4.

Ключевые слова: ЭКГ, вариабельность, ЭЭГ, взаимная информация, кардиоритмограмма.

Список использованных источников

1. Cover, T.M., Thomas. Elements of information theory / T.M. Cover, Thomas - New York: Wiley, 1991. - ISBN 0471062596 9780471062592.
2. Doquire, G., Verleysen, M. A Comparison of Multivariate Mutual Information Estimators for Feature Selection / G. Doquire, M. Verleysen - : SciTePress - Science and Technology Publications, 2012. - ISBN 978-989-8425-98-0, 978-989-8425-99-7.
3. Freedman, D., Diaconis, P. On the histogram as a density estimator: L 2 theory / D. Freedman, P. Diaconis // Probability theory and related fields. - 1981. - Vol. 57. - P. 453–476.
4. Knuth, K.H. Optimal data-based binning for histograms / K.H. Knuth // arXiv preprint physics. - 2006.
5. Legg, P.A., Rosin, P.L., Marshall, D., Morgan, J.E. Improving accuracy and efficiency of registration by mutual information using Sturges' histogram rule / P.A. Legg, P.L. Rosin, D. Marshall, J.E. Morgan // Proc. Med. Image Understand. Anal. - 2007. - P. 26–30.
6. Marek, T., Tichavsky, P. On the estimation of mutual information / T. Marek, P. Tichavsky - , 2008.
7. Shimazaki, H., Shinomoto, S. A Method for Selecting the Bin Size of a Time Histogram / H. Shimazaki, S. Shinomoto // Neural Computation. - 2007. - Vol. 19. - P. 1503–1527.
8. Sorjamaa, A., Hao, J., Lendasse, A. Mutual information and k-nearest neighbors approximator for time series prediction / A. Sorjamaa, J. Hao, A. Lendasse - : Springer, 2005.
9. Sturges, H.A. The Choice of a Class Interval / H.A. Sturges // Journal of the American Statistical Association. - 1926. - Vol. 21. - P. 65–66.
10. Suzuki, T., Sugiyama, M., Sese, J., Kanamori, T. Approximating Mutual Information by Maximum Likelihood Density Ratio Estimation / T. Suzuki, M. Sugiyama, J. Sese, T. Kanamori // Journal of Machine Learning Research-Proceedings Track. - 2008. - Vol. 4. - P. 5–20.
11. Verleysen, M. K-Nearest Neighbours based on Mutual Information for Incomplete Data Classification / M. Verleysen - Bruges, Belgium: , 2008. - ISBN 2930307080 9782930307084.
12. Zhukov, M., Popov, A. Bin Number Selection for Equidistant Mutual Information Estimation / M. Zhukov, A. Popov - Kyiv, Ukraine: , 2014.