Published in final edited form as:

Image Anal Mov Organ Breast Thorac Images (2018). 2018 September ; 11040: 326–334.

doi:10.1007/978-3-030-00946-5_33.

On the Relevance of the Loss Function in the Agatston Score Regression from Non-ECG Gated CT Scans

Carlos Cano-Espinosa¹, Germán González², George R. Washko³, Miguel Cazorla¹, Raúl San José Estépar⁴

¹Department of Computer Science and Artificial Intelligence, University of Alicante, Alicante, Spain

²Sierra Research SL, Avda. Costa Blanca 132, Alicante, Spain

³Brigham and Women's Hospital, Pulmonary and Critical Care Medicine, 75 Francis Street, Boston, MA, USA

⁴Applied Chest Imaging Laboratory, Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

Abstract

In this work, we evaluate the relevance of the choice of loss function in the regression of the Agatston score from 3D heart volumes obtained from non-contrast non-ECG gated chest computed tomography scans. The Agatston score is a well-established metric of cardiovascular disease, where an index of coronary artery disease (CAD) is computed by segmenting the calcifications of the arteries and multiplying each calcification by a factor related to their intensity and their volume, creating a final aggregated index. Recent work has automated such task with deep learning techniques, even skipping the segmentation step and performing a direct regression of the Agatston score. We study the effect of the choice of the loss function in such methodologies. We use a large database of 6983 CT scans to which the Agatston score has been manually computed. The dataset is split into a training set and a validation set of $n = 1000$. We train a deep learning regression network using such data with different loss functions while keeping the structure of the network and training parameters constant. Pearson correlation coefficient ranges from 0.902 to 0.938 depending on the loss function. Correct risk group assignment measurements range between 59.5% and 81.7%. There is a trade-off between the accuracy of the Pearson correlation coefficient and the risk group measurement, which leads to optimize for one or the other.

Keywords

Loss functions; Agatston score; Regression Convolutional

1 Introduction

The Agatston score is a well-established metric used to measure the extent of coronary artery disease (CAD) in ECG-gated CT studies [1]. This biomarker is computed by measuring the volume and maximum intensity of the coronary artery calcifications (CAC) and adding the per-lesion score to learn a global index. The Agatston score is then used to classify subjects in five different clinically relevant risk groups, defined by the following ranges: [0, 1], (1, 100], (100, 400], (400, 1000], (1000, inf], as described [3].

Several works automate the computation of the Agatston following the same general pipeline: hearts are located using anatomy-based [9], atlas-based [7] or 2.5D object detection [4, 11] strategies and a 3D Region of Interest (ROI) is extracted around the heart. Then, each CAC candidate is categorized as relevant or not using their relative position [16], texture and size features [6] or a combination of both [12, 13]. Finally, the Agatston score is computed from the CACs. The latest work of [14, 15] uses a deep-learning solution for CAC classification. This methodology uses a database of segmented CAC as the reference standard, where each voxel is labeled to indicate if it is part of a CAC or not, to train a lesion-based or a voxel-based classifier. In contrast, the work of [4] generates the inclusion and exclusion rules of the CACs by optimizing the global score directly.

The work of [5] uses a deep learning network for the regression of image-based biomarkers, and the work of [2] does so, specifically for the problem of Agatston from CT images. The latter approach minimizes the L2 cost function between the reference standard and the regressed Agatston score. While being attractive for its simplicity and achieving a similar Pearson correlation coefficient as other deep-learning based methods, it is inferior with respect to the classification of subjects to risk-groups. In this experimental work, we explore improvements to such methodology by analyzing the relevance of the cost-function of the regression network.

2 Materials and Methods

2.1 Database

COPDGene is a multi-center observational study designed to understand the evolution and genetic signature of COPD in smokers [10]. COPDGene contains a total of 10,000 pulmonary non-ECG gated CT Scans obtained with 16 detectors scanners. Subjects are both smokers and non-smokers, with ages between 45 and 80 years and 10 years of smoking history from non-hispanic white and non-hispanic African Americans ethnicities. The Agatston score was manually estimated in 6983 of such images, forming the database on which we train and evaluate the method.

We automatically select a region of interest (ROI) centered around the heart in each CT scan (Fig. 1), using the method of [4,11]. We use a prefixed ROI size to avoid the need of re-scaling the reference standard Agatston score. Each heart ROI is further normalized to a canonical size of $64 \times 64 \times 64$ voxels to enable their processing using a 3D convolutional neural network. The images are clamped to the range $[-500, 2000]$ HUs to highlight the lesions and discard lung structures. Mistakes in the automated location of the heart were

eliminated by manual inspection, resulting in 6663 images that are divided between a training set ($n = 5663$) and a testing set ($n = 1000$).

2.2 Data Augmentation

In our database, the number of cases in high-risk groups according to their Agatston score is lower than the number of cases in low-risk groups, as shown in Fig. 2. This poses challenges when training the regression network. We use a data augmentation technique to reduce such data imbalance. The technique generates an equalized number of cases per group by generating random displacements over the three axes, using a spherical probabilistic volume. Such is done to ensure that the new augmented sample is equidistant from the center of the heart in all directions. The data augmentation is done on-the-fly and to ensure reproducibility, the random seed of the data augmentation policy is fixed.

2.3 Convolutional Network

Due to its simplicity and for comparative purposes, we use the network proposed in [2] and depicted in Fig. 3. The network consists of three 3D convolutional-max-pooling blocks with rectified linear activation functions, followed by two fully connected layers that output the regression in a single neuron with linear activation. Dropout layers are present to prevent over-fitting. At test time, the negative regressions are clipped to 0, since the Agatston score is always positive. The optimizer used is the well known Adaptive Momentum optimizer [8], with an exponential decay rate.

Cost Functions: The convolutional network is optimized with respect to four different cost functions: mean square error of normalized values, absolute difference of normalized values, mean square error of the logarithmic scaled scores and absolute difference of the logarithmic scaled scores. The definition of the cost functions is shown in Table 1. We have chosen to optimize using linear and logarithmic cost functions because while the Agatston score is computed in a linear scale, its association to risk groups follows a loosely-logarithmic scale.

Categorization: We analyzed the performance of the network with respect to the direct risk group estimation for the subject in two manners: first, we assigned to each subject his risk group and regressed the group directly using the $L1$ norm as cost function. Second, we turned the regression network into a categorization one by substituting the last neuron for five neurons followed by a softmax activation function and optimized the categorical cross entropy loss function.

2.4 Comparison Metrics

To evaluate the performance of the different cost functions, we use the well defined Pearson's correlation coefficient, the Spearman correlation coefficient and the risk-group accuracy (RGAcc), defined as the percentage of cases that correctly classified in their risk group.

3 Results

Table 1 shows the results obtained for the six defined loss functions. Better Pearson correlation coefficients are found when using linear cost functions instead of logarithmic ones. However, the RGAcc metric improves when using logarithmic cost functions. The Spearman correlation coefficient is higher in the logarithmic cost functions than in the linear ones. Such is consistent with the measurements RGAcc, which are higher in logarithmic cost functions.

Figure 4 displays the correlation plots between the reference standard and the computed Agatston score, as well as the concordance matrices of the risk groups. The correlation plots are made in both linear and logarithmic scales to provide a fair comparison between the different metrics. There is a large difference in performance between linear and logarithmic cost functions in the lowest range of the Agatston score, where the logarithmic cost functions produce more accurate results. Conversely, the logarithmic cost functions are less accurate with large Agatston scores, which leads to lower Pearson correlation coefficients. Since low-risk groups are more prevalent in the database as shown the Fig. 2, the logarithmic cost functions achieve better risk accuracy percentages.

In both logarithmic and linear scales, the mean absolute difference outperforms or equals the mean squared error for Pearson and Spearman coefficients and the risk group accuracy percentage.

The cost function that achieves a good trade-off between the correlation coefficient and the risk accuracy percentage is the mean logarithmic absolute error, with a $\rho = 0.916$ and accuracy of 81.7% and the maximum Spearman coefficient of 0.949.

When treating the Agatston risk group assignment as a classification problem, we achieve a RGAcc of 75.7% when using the categorical cross entropy loss function and of 81.1% when using a risk group regression with the mean average error loss function.

4 Discussion

In this work, we have shown that keeping the deep learning network, the training parameters and the data constant, the selection of the cost function has drastic effects on the performance of the Agatston score regression. We have found that the mean absolute error of the logarithm of the score achieves a good trade-off between the Pearson correlation coefficient and the correct classification of the subjects according to their risk group while achieving the highest Spearman correlation. Such result is of little surprise, since the risk groups are defined in a pseudo-logarithmic scale, and the errors of the linearly scaled network are in the lowest range.

As an alternative to the regression of the Agatston score and then assignment of the subjects to the risk groups, we have predicted the risk group directly using a classification network. Such framework breaks the ordering of the risk groups, and shows worse performance than the regression networks. When regressing the Agatston risk group directly, we achieve comparable, but lower, RGAcc than that obtained with our best Agatston score regression

network. This comparison favors the Agatston score regression network, since it does not only provide the risk group but also quantifies the score itself.

One valid criticism of this work is the extreme simplicity of the network used. This has been chosen since (a) it achieves results comparable to the state of the art and (b) it is faster to train than deeper convolutional networks. Still, training time is of 600 s per epoch, and convergence is normally achieved around the 100th epoch, depending on the loss function. Such numbers are due to the 3D nature of the problem and the extension of the dataset.

The Pearson correlation coefficient is often referred in the literature as the figure of merit for biomarker regression methodologies. However, in our experiments, we have found that the Spearman correlation coefficient provides a more coherent view of the data as an aggregated statistic. Such is demonstrated with the data of Table 1, where the Spearman correlation relates more to the risk accuracy group than the Pearson correlation coefficient. The Pearson correlation coefficient depends on the per-case error, and high-valued cases can strongly bias the overall measurement. The Spearman correlation coefficient assesses a monotonic relationship among variables, ignoring their absolute error. For cases such as the problem presented, it is more important to produce good rankings among the cases than minimizing the error on the Agatston score, and thus justifying Spearman's correlation as the preferred figure of merit.

Acknowledgments

This study was supported by the NHLBI awards R01HL116931, R01HL116473, and R21HL140422. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

References

1. Agatston AS, Janowitz WR, Hildner FJ, Zusmer NR, Viamonte M, Detrano R: Quantification of coronary artery calcium using ultrafast computed tomography. *J. Am. Coll. Cardiol* 15(4), 827–832 (1990) [PubMed: 2407762]
2. Cano-Espinosa C, González G, Washko GR, Cazorla M, Estépar RSJ: Automated Agatston score computation in non-ECG gated CT scans using deep learning. In: *Proceedings of the SPIE: Medical Imaging 2018*, 2 2018
3. Erbel R, Möhlenkamp S, Kerkhoff G, Budde T, Schmermund A: Non-invasive screening for coronary artery disease: calcium scoring. *Heart* 93(12), 1620–1629 (2007) [PubMed: 18003695]
4. González G, Washko GR, Estépar RSJ: Automated Agatston score computation in a large dataset of non ECG-gated chest computed tomography. In: *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pp. 53–57. IEEE (2016)
5. González G, Washko GR, Estépar RSJ: Deep learning for biomarker regression. Application to osteoporosis and emphysema on chest CT scans. In: *Proceedings of the SPIE: Medical Imaging 2018*, 2 2018
6. Isgum I, Prokop M, Niemeijer M, Viergever MA, van Ginneken B: Automatic coronary calcium scoring in low-dose chest computed tomography. *IEEE Trans. Med. Imaging* 31(12), 2322–2334 (2012) [PubMed: 22961297]
7. Isgum I, Staring M, Rutten A, Prokop M, Viergever MA, Van Ginneken B: Multi-atlas-based segmentation with local decision fusion—application to cardiac and aortic segmentation in CT scans. *IEEE Trans. Med. Imaging* 28(7), 1000–1010 (2009) [PubMed: 19131298]
8. Kingma D, Ba J: Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)

9. Reeves AP, Biancardi AM, Yankelevitz DF, Cham MD, Henschke CI: Heart region segmentation from low-dose CT scans: an anatomy based approach. In: Medical Imaging: Image Processing, p. 83142A (2012)
10. Regan EA, et al.: Genetic epidemiology of COPD (COPDgene) study design. COPD J. Chronic Obstr. Pulm. Dis 7(1), 32–43 (2011)
11. Rodriguez-Lopez S, et al.: Automatic ventricle detection in computed tomography pulmonary angiography. In: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), pp. 1143–1146. IEEE (2015)
12. Shahzad R, et al.: Vessel specific coronary artery calcium scoring: an automatic system. Acad. Radiol 20(1), 1–9 (2013) [PubMed: 22981481]
13. Wolterink JM, Leiner T, Takx RA, Viergever MA, Išgum I: An automatic machine learning system for coronary calcium scoring in clinical non-contrast enhanced, ECG-triggered cardiac CT In: Medical Imaging 2014: Computer-Aided Diagnosis, p. 90350E International Society for Optics and Photonics (2014)
14. Wolterink JM., Leine T., Viergever MA., Išgu I.: Automatic coronary calcium scoring in cardiac CT angiography using convolutional neural networks In: Nava N., Hornegge J., Well WM., Frang AF. (eds.) MICCAI 2015. LNCS, vol. 9349, pp. 589–596. Springer, Cham (2015). 10.1007/978-3-319-24553-972
15. Wolterink JM, Leiner T, de Vos BD, van Hamersvelt RW, Viergever MA, Išgum I: Automatic coronary artery calcium scoring in cardiac CT angiography using paired convolutional neural networks. Med. Image Anal 34, 123–136 (2016) [PubMed: 27138584]
16. Xie Y, Cham MD, Henschke C, Yankelevitz D, Reeves AP: Automated coronary artery calcification detection on low-dose chest CT images. In: Proceedings of SPIE, vol. 9035, p. 90350F (2014)

Fig. 1.

Regions of Interest extracted around the heart. Each column corresponds to a case. Rows are the axial, coronal and sagittal planes that cross the central point of the ROI. Coronary Calcifications are identified by bright voxels found in the coronary arteries. They have highlighted with green circles. Numbers below each image correspond to the Agatston score value calculated using the full volumetric information. Please note that there are other bright voxels in the image that correspond to bone structure or calcifications that are not present in the coronary arteries. Such structures should be rejected by the algorithm. (Color figure online)

Fig. 2.

Left: Distribution of database for each Risk Group, Right: Risk Groups Agatston score range as defined in [3].

Fig. 3.
Convolutional network diagram.

Fig. 4. Correlation results for the loss functions in Table 1. In each row, the first two plots represent the correlation plotted in linear and logarithmic scale respectively. Red dots were used for incorrect risk group result and blue dots depicted a correct group classification. The last plot shows the confusion matrix of risk groups, where the color of each cell represents the relative prevalence within the column. The numbers of the first column correspond to the loss functions described in Table 1. (Color figure online)

Table 1.

Loss function comparative results, Pearson ' ρ ' and Spearman ' s ' correlation coefficients in linear and logarithmic scale, and Risk Group Accuracy ' $RGAcc$ '. ' Z ' correspond to a normalization value, 5000 in this case. We have omitted ' s ' correlation coefficient in logarithmic scale since it is equivalent to ' s ' in linear scale due to the fact that a logarithmic transformation preserves ordering.

	Loss function	ρ	$\rho(\log)$	s	$RGAcc$
(1)	$\frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - x_i}{Z} \right)^2$	0.932	0.664	0.843	55.2%
(2)	$\frac{1}{n} \sum_{i=1}^n \left \frac{y_i - x_i}{Z} \right $	0.938	0.657	0.903	59.5%
(3)	$\frac{1}{n} \sum_{i=1}^n (\log(y_i + 1) - \log(x_i + 1))^2$	0.902	0.620	0.949	78.2%
(4)	$\frac{1}{n} \sum_{i=1}^n \log(y_i + 1) - \log(x_i + 1) $	0.916	0.631	0.949	81.7%
(5)	Categorical cross entropy	N/A	N/A	N/A	75.7%
(6)	Categorical regression	N/A	N/A	N/A	81.1%