Journal of Physics: Conference Series **2858** (2024) 012002 doi:10.1088/1742-6596/2858/1/012002

Support vector machine based on the quadratic unconstrained binary optimization model

Qiwei Wang*

Shandong University of Science and Technology, College of Mathematics and Systems Science, Qingdao, 266000, China

*Corresponding author's e-mail: 2938035936@qq.com

Abstract: Support vector machine (SVM) is a powerful supervised machine learning model that is often used in binary classification algorithms. As Moore's Law approaches its theoretical limits and the demand for machine learning to handle large-scale, high-dimensional data analysis intensifies, the necessity of adopting non-traditional computational approaches becomes evident. Quantum computing, in particular, emerges as a vital solution for the effective training of SVM models, providing capabilities beyond those of classical computing systems. To solve the above problems, a QUBO (quadratic unconstrained binary optimization) model is proposed to transform the SVM machine learning model into a quadratic unconstrained binary optimization problem so that they can be effectively trained on the D-Wave platform using adiabatic quantum computer. The results show that the QUBO model can transform the SVM model into a simple quadratic programming problem, which makes it suitable for adiabatic quantum computer processing. When processing large-scale and high-dimensional data, this transformation shows a natural advantage and significantly improves computational efficiency. The application potential of this transformation technology is huge in the medical field.

1. Introduction

Support Vector Machine is a generalized linear classifier that classifies data using supervised learning. Its decision boundary is the maximum margin hyperplane obtained from learning samples[1]. High-dimensional data has brought challenges to traditional SVM, so a method of embedding traditional mathematical models into quantum computers has become a research hotspot, and the representative related studies in recent years are described below.

Vapnik et al.[2] proposed SVM in 1995. Since then, the research on SVM has improved the classification effect and accuracy to a certain extent, but the optimization effect in processing large-scale data is not obvious. In 1982, Richard Feynman[3] proposed the potential of quantum "parallelism" to make computations more efficient. In the early 21st century, Farhi et al.[4] proposed adiabatic quantum computing and quantum approximate optimization algorithms. Aharonov et al.[5] showed that adiabatic quantum computing is equivalent to standard quantum computing. Kadowaki and Nishimori[6] proposed a quantum annealing algorithm based on adiabatic quantum computation to solve the combinatorial optimization problem. Anthony et al.[7] proposed a quadratic unconstrained binary optimization problem and obtained binary variables to minimize the objective function. Glover et al.[8] proposed Quantum Bridge Analytics I: A Tutorial on Formulating and Using QUBO Models, which is currently the most widely used optimization model in the field of quantum computing and unifies various combinatorial optimization problems.

The traditional SVM training process is reduced to a convex quadratic programming problem. With the increase of training samples, the training space and time will increase dramatically, so the SVM training model is difficult to apply to large-scale datasets. In order to further improve the learning efficiency of the SVM model to meet the needs of various large-scale data processing, this study introduced the QUBO model to transform the SVM training model into a Hamiltonian operator energy expression that can run on adiabatic quantum computer and then generate QUBO matrix. Compared with forming the QUBO matrix in the process of problem transformation, transformation is easier, and the ground state representation of Hamiltonian is the best solution to the problem. In addition, this study proposes to construct SVM with kernel function in QUBO framework to process nonlinear separable large-scale data on adiabatic quantum computers. Finally, the research method in this paper is applied to the medical field to identify autistic children. In general, medical images and other data in the medical field usually have a high dimension, while the traditional SVM algorithm uses violent exhaustion to solve the problem in theory, but it may not be realized due to high computational cost. This research provides a method for the application of quantum computers to machine learning, so this topic has research value.

QUBO model is a quadratic programming problem used to describe optimization problems, and the goal is to minimize a quadratic objective function[9]. Adiabatic quantum computers can efficiently solve the QUBO problem[10]. Therefore, converting SVM into QUBO form and solving it with adiabatic quantum computers can greatly improve the efficiency and performance of the algorithm. Compared with binary bits in traditional computers, qubits can be in multiple states at the same time, so using adiabatic quantum computing to train SVM models greatly improves computational efficiency and speed.

2. Related word

2.1. Support vector machine

The ultimate goal of support vector machines is to find the hyperplane with the maximum distance from both types of data points, so as to maximize the distance between the two types of data points. Suppose the training data X consists of n sample points, as follows:

$$\mathrm{X} = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \tag{1}$$

where $x_i \in \mathrm{R}^d$ is the ith row vector in X and $y_i \in \{-1,1\}$ is the class label to which x_i belongs. The SVM mathematical model is as follows[11]:

$$\begin{aligned} &\min \frac{1}{2}\left||w|\right|^2 + \mathrm{c}\textstyle\sum_{i=1}^{n} \xi_i \\ &\mathrm{s.t.}\ y_i[w^{\mathrm{T}}\varphi(x_i) + \mathrm{b}] \geq 1 - \xi_i,\ \xi_i \geq 0, \forall i = 1,2,\dots,n \end{aligned} \tag{2}$$

where w is the weight vector, b is the bias, ξ_i represents a relaxed variable, $\mathrm{c}\sum_{i=1}^{n} \xi_i$ denotes a regular term, c is the regularization parameter, and $\varphi(\bullet)$ signifies a nonlinear mapping from the input space to high-dimensional feature space. The Lagrangian optimization method is used to transform the nonlinear SVM problem into a dual problem, and the Lagrangian function of convex quadratic programming is defined as follows:

$$\theta(w, \mathrm{b}, \lambda, \xi, \beta) = \frac{1}{2}\left||w|\right|^2 + \mathrm{c}\textstyle\sum_{i=1}^{n} \xi_i + \sum_{i=1}^{n} \lambda_i[1 - \xi_i - y_i w^{\mathrm{T}}\varphi(x_i) - y_i \mathrm{b}] - \sum_{i=1}^{n} \beta_i \xi_i \tag{3}$$

where $\lambda = [\lambda_1, \lambda_2, \dots, \lambda_n]^{\mathrm{T}}$ and $\beta = [\beta_1, \beta_2, \dots, \beta_n]^{\mathrm{T}}$ are the vector containing all the Lagrange multipliers, with $\lambda_i \geq 0$, $\beta_i \geq 0$. According to the Karush-Kuhn-Tucker condition for inequality-constrained optimization problems, the gradient of $\theta(w, \mathrm{b}, \lambda, \xi, \beta)$ with respect to w, b, ξ is set to 0. Then, the dual problem is obtained as follows:

$$\begin{aligned} &\max \theta(\lambda) = \textstyle\sum_{i=1}^{n} \lambda_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j y_i y_j [\varphi(x_i) \bullet \varphi(x_j)] \\ &\mathrm{s.t.}\ 0 \leq \lambda_i \leq C \\ &\textstyle\sum_{i=1}^{n} \lambda_i y_i = 0\ ,\ \ \forall i = 1,2,\dots,n \end{aligned} \tag{4}$$

where C represents a constant. The role of the mapping function $\varphi(\bullet)$ in the algorithm is entirely implemented through the inner product $[\varphi(x_i) \bullet \varphi(x_j)]$, from which a kernel function can be introduced as follows:

$$\mathrm{K}(x_i, x_j) = [\varphi(x_i) \bullet \varphi(x_j)] \tag{5}$$

where $\bullet$ is the inner product operation of $\varphi(x_i)$ and $\varphi(x_j)$. Bring Eq. (18) into Eq. (17) as follows:

$$\max\theta(\lambda) = \sum_{i=1}^{n} \lambda_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i\lambda_j y_i y_j \mathrm{K}(x_i, x_j) \tag{6}$$

Common kernel functions are linear, polynomial, Gaussian radial basis kernel functions, as follows:

$$\mathrm{K}(x_i, x_j) = x_i{}^{\mathrm{T}} x_j$$
$$\mathrm{K}(x_i, x_j) = (x_i{}^{\mathrm{T}} x_j + 1)^m$$
$$\mathrm{K}(x_i, x_j) = e^{\frac{||x_i - x_j||^2}{\partial^2}} \tag{7}$$

The parameters m and ∂ are selected, and appropriate constants are chosen based on the specific problem at hand.

2.2. The QUBO model

The The QUBO model is a combinatorial optimization model utilized for solving quadratic function and binary variable optimization problems. It comprises a linear part and a quadratic part, as follows[12]:

$$H_{QUBO}(x) = \sum_{i \in n} a_i x_i + \sum_{i,j \in n} b_{ij} x_i x_j = x^{\mathrm{T}} Q x, x_i \in \{0,1\} \tag{8}$$

The constants a_i and b_{ij} are arbitrary, and n represents the number of decision variables. The matrix Q is of type n*n and contains all the data necessary for solving the problem. The matrix Q is defined as follows: Q_{ij} equals to a_i when i equals to j, and Q_{ij} equals to b_{ij} when i is less than j. All other elements of Q are 0, and b_{ij} equals to b_{ji}. Additionally, x^{T} represents the transpose of the vector x.

3. Transformation of the SVM model

By transforming the QUBO model, the SVM training model is converted into a binary unconstrained quadratic function. The corresponding coefficients between binary variables are then calculated, and the Q matrix conforming to the QUBO form is constructed. This transformation effectively turns the SVM training model into an optimization problem for quadratic unconstrained binary approximate solution, which aligns well with D-wave adiabatic quantum computing capabilities.

3.1. QUBO form of SVM model

The QUBO model is typically utilized for solving minimum problems. To transform the dual problem into solving the minimum problem, rewrite equation (6) to make it applicable to the QUBO model:

$$\min\theta(\lambda) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i\lambda_j y_i y_j \mathrm{K}(x_i, x_j) - \sum_{i=1}^{n} \lambda_i$$
$$\forall i = 1,2,\ldots,\mathrm{n}，\ \lambda_i, \lambda_j \geq 0 \tag{9}$$

The QUBO model can only handle the 0,1 binary decision variable, while λ is any constant greater than zero, hence the introduction of the S-dimensional precision vector $\boldsymbol{p} = [p_1, p_2, \ldots, p_S]^{\mathrm{T}}$ is defined. Each element in $\boldsymbol{p}$ is an integer power of 2 and is positive. Additionally, S binary decision variables λ_{is} are introduced for each Lagrange multiplier. These variables indicate whether the ith Lagrange multiplier chooses the sth element in $\boldsymbol{p}$. If the ith Lagrange multiplier chooses the sth element in $\boldsymbol{p}$, then λ_{is} equals 1; otherwise, it equals 0. This representation allows for expressing the Lagrange multiplier λ_i using both the vector $\boldsymbol{p}$ and the binary decision variables λ_{is} as follows:

$$\lambda_i = \sum_{s=1}^{S} p_s \lambda_{is} \tag{10}$$

The symbol p_s denotes the sth element in the precision vector $\boldsymbol{p}$.

$$\lambda_i\lambda_j = \sum_{s_1=1}^{S}\sum_{s_2=1}^{S} p_{s_1}\lambda_{is_1} p_{s_2}\lambda_{js_2} \tag{11}$$

Journal of Physics: Conference Series **2858** (2024) 012002 doi:10.1088/1742-6596/2858/1/012002

λ_{is} is a primary variable, and the QUBO model is utilized for quadratic variables. However, as x_i is a binary variable with values of 0 or 1, so $\lambda_{is} = {\lambda_{is}}^2$, then can be expressed as follows:

$$\lambda_i = \sum_{s=1}^{S} p_s {\lambda_{is}}^2 \tag{12}$$

Each data in the X training dataset is a D-dimensional vector, where d represents the number of features in the training dataset. By converting the form of the inner product between the sample data into the form of variable coefficients and incorporating Equations (12) and (13) into Equations (9), the SVM training model can be transformed into an energy expression known as the Hamiltonian operator through the QUBO model transformation. When the SVM model introduces linear kernel function, polynomial kernel function, and Gaussian radial basis kernel function, the energy expression of the Hamiltonian operator is as follows:

$$\min\theta(\lambda) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{s_1=1}^{S}\sum_{s_2=1}^{S}(p_{s_1}\lambda_{is_1})(p_{s_2}\lambda_{js_2})\, y_{[i][d+1]}y_{[j][d+1]}\left(\sum_{d_1=1}^{d} x_{[i][d_1]}x_{[j][d_1]}\right) - \sum_{i=1}^{n}\sum_{s=1}^{S} p_s {\lambda_{is}}^2 \tag{13}$$

$$\min\theta(\lambda) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{s_1=1}^{S}\sum_{s_2=1}^{S}(p_{s_1}\lambda_{is_1})(p_{s_2}\lambda_{js_2})\, y_{[i][d+1]}y_{[j][d+1]}\left(\sum_{d_1=1}^{d} x_{[i][d_1]}x_{[j][d_1]} + 1\right)^m - \sum_{i=1}^{n}\sum_{s=1}^{S} p_s {\lambda_{is}}^2 \tag{14}$$

$$\min\theta(\lambda) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{s_1=1}^{S}\sum_{s_2=1}^{S}(p_{s_1}\lambda_{is_1})(p_{s_2}\lambda_{js_2})\, y_{[i][d+1]}y_{[j][d+1]}e^{\frac{\sum_{d_1=1}^{d}(x_{[i][d_1]}x_{[j][d_1]})^2}{\partial^2}} - \sum_{i=1}^{n}\sum_{s=1}^{S} p_s {\lambda_{is}}^2 \tag{15}$$

The Hamiltonian energy expression serves as a bridge between classical computation and adiabatic quantum computation, with the optimal solution achievable through the quantum annealing algorithm on the D-Wave platform.

Note: By defining the precision vector $\boldsymbol{p}$, the constant Lagrange multiplier λ is represented by coefficients and binary variables, which reduces the complexity of conversion. The inner product operation of$\varphi(x_i)$,$\varphi(x_j)$is transformed into the coefficient form of variables $\sum_{d_1=1}^{d} x_{[i][d_1]}x_{[j][d_1]}$, and then it is deformed according to the selected kernel function. The relationship between the characteristic variables is calculated, and the Q matrix in the QUBO model is generated.

3.2. SVM training model hyperplane solution and prediction

The linear divisibility of data is achieved in a high-dimensional space, and an optimal decision boundary is obtained by constructing a separation hyperplane. The hyperplane is as follows:

$$w^{\mathrm{T}}\varphi(x) + \mathrm{b} = 0 \tag{16}$$

The value of the weight w of the SVM model is obtained from Equation (14), but the high-dimensional mapping dimension of $\varphi(x)$ can be infinite, therefore, $w^{\mathrm{T}}\varphi(x)$ is directly obtained according to the hyperplane, and then the value of b is obtained as follows:

$$w^{\mathrm{T}}\varphi(x) = \sum_{i=1}^{n}\lambda_i y_i \varphi(x_i)\varphi(x) = \sum_{i=1}^{n}\lambda_i y_i \mathrm{K}(x_i, x) \tag{17}$$

$$\mathrm{b} = \frac{\sum_{i=1}^{n}[y_i - \sum_{j=1}^{n}\lambda_j y_j \mathrm{K}(x_i, x_j)]}{n} \tag{18}$$

In order to be more robust, b is the average value.

The binary decision variable λ_{is} value has been computed by the quantum computer on the D-wave platform, resulting in the following SVM model hyperplane:

$$\sum_{i=1}^{n}\sum_{s=1}^{S} p_s\lambda_{is}\, y_i \mathrm{K}(x_i, x) + \frac{\sum_{i=1}^{n}[y_i - \sum_{j=1}^{n}\sum_{s=1}^{S} p_s\lambda_{js} y_j \mathrm{K}(x_i, x_j)]}{n} = 0 \tag{19}$$

The SVM training model predicts a sample. When entering the prediction sample x_c, the category label y_c is as follows:

$$y_c = sign\left(\sum_{i=1}^{n}\sum_{s=1}^{S} p_s\lambda_{is}\, y_i \mathrm{K}(x_i, x) + \frac{\sum_{i=1}^{n}[y_i - \sum_{j=1}^{n}\sum_{s=1}^{S} p_s\lambda_{js} y_j \mathrm{K}(x_i, x_j)]}{n}\right) \tag{20}$$

By transforming the QUBO model, the SVM training model is converted into a binary classification machine training model suitable for quantum computer training. This allows for obtaining the classification hyperplane and predicting unknown label data, thereby enhancing the efficiency of training multidimensional binary classification data in machine learning.

4. Experimental analysis

4.1. Data description

The R-fMRI data utilized in this study were obtained from the Open Access Autism Brain Imaging Data Exchange (ABIDE) database[13]. Detailed demographic information analysis of R-fMRI data[14] is presented in Table 1.

Table 1. Demographic information of the participants.

	ASD	TD	p-value
Gender (M/F)	38/7	36/11	0.3428
Age(years)	11.1±2.3	11.0±2.3	0.7773
FIQ	106.8±17.4	113.3±14.1	0.0510

According to Table 1, as the value of P is greater than 0.05, there are no statistically significant differences in age, gender, and IQ between the two groups. According to the Automatic Anatomical Labeling (AAL) atlas[15], the brain space was divided into 116 regions of interest (ROI). After pre-processing, we obtained the average value of functional magnetic resonance imaging (fMRI) for 116 ROIs, i.e., the data of [170,116] per image. The cumulative contribution rate of the first 12 principal components of the image data reaches about 0.8, which contains the vast majority of the information, and the first 12 principal components of a single image are visualized, as shown in Figures 1 and 2:

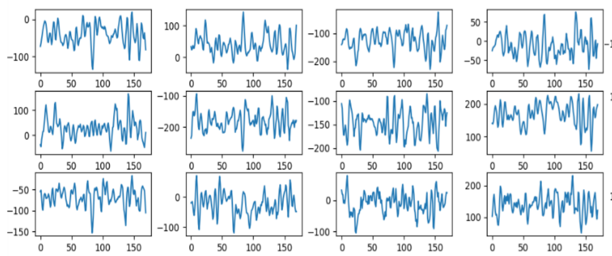

Figure 1. Visualization of the ASD dataset.

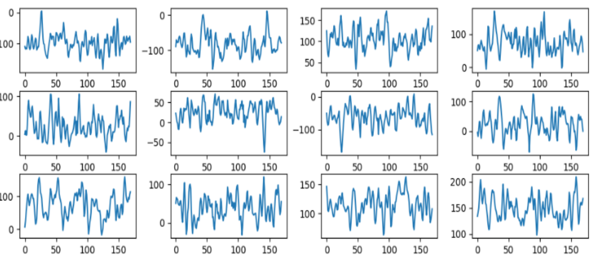

Figure 2. Visualization of the TD dataset.

4.2. Model performance evaluation

Data from 46 images were randomly selected as the training set X and 10 as the test set (with a training-to-test ratio of 8:2). In the preparation of experimental data, samples from individuals with autism spectrum disorder were designated as Positive samples, and samples from individuals with normal development were designated as Negative samples. The model performance was evaluated using the evaluation index $A_{accuracy} = \frac{T_{TP}+T_{TN}}{T_{TP}+T_{TN}+T_{FP}+T_{FN}}$ for detection accuracy.

4.3. Analysis of experimental results

The autoencoder method is used to further select the features of the R-fMRI data, and the irrelevant features and the features with low correlation are removed without reducing the classification accuracy, and the parameter S is selected as 10. Take the linear kernel function as an example, and create the corresponding model through the Python open-source library PyQUBO. The model compiled by PyQUBO can serve as input for D-Wave sampler. The annealing method is then utilized to train the optimal solution for binary λ_{is}, as follows:

$$\lambda_{00} = 1, \lambda_{01} = 0, \lambda_{02} = 0, \lambda_{03} = 0, \lambda_{04} = 1, \lambda_{05} = 1, \lambda_{06} = 1, \lambda_{07} = 0, \lambda_{08} = 0, \lambda_{09} = 1 \ldots$$

The value of λ_i can be obtained from formula (26) as follows:

$$\lambda_0 = 625, \lambda_1 = 687, \lambda_2 = 760, \lambda_3 = 467, \lambda_4 = 690, \lambda_5 = 986, \lambda_6 = 282, \lambda_7 = 514 \ldots$$

Due to the substantial volume of variable data, we will not display each individual piece here. The hyperplane of the SVM is obtained, and the prediction is made by the test set data, $A_{accuracy} = 0.8$, the classification performance is better. This method solves the shortcomings of traditional SVM brute-force solutions in the face of high-dimensional data. The model presented in this paper only shows that SVM can be run on a quantum computer, so as to solve the problem that traditional SVM models cannot be run directly on a quantum computer, and the experiment is only to illustrate this.

5. Conclusion

In order to improve the performance of SVM in processing high-dimensional complex data, this study employs the QUBO model to enhance the traditional SVM model and adapt it for quantum computer operation. The QUBO model acts as a bridge between traditional algorithms and quantum algorithms. This model is utilized in predicting autism in children, incorporating kernel functions into the training process and reducing space complexity. The optimal solution is attained through quantum annealing methods, integrating medical field issues into quantum computing.

References

[1] Zhihua, Z. (2016) Machine learnin. Tsinghua University Press, Beijing. http://www.tup.tsinghua.edu.cn/booksCenter/book_06402702.

[2] Cortes, C. and Vapnik, V. (1995) Support-vector networks. Machine learning, 20: 273-297. https://springer.longhoe.net/article/10.1007/Bf00994018.

[3] Feynman, R.P. (1982) Simulating physics with computers. Int J Theor Phys, 21：467–488. https://doi.org/10.1007/BF02650179.

[4] Farhi, E., Goldstone, J., Gutmann, S., et al. (2001) A quantum adiabatic evolution algorithm applied to random instances of an NP-complete problem. Science, 292(5516): 472-475. https://www.science.org/doi/abs/10.1126/science.1057726.

[5] Aharonov, D., Van Dam, W., Kempe, J., et al. (2008) Adiabatic quantum computation is equivalent to standard quantum computation. SIAM review, 50(4): 755-787. https://epubs.siam.org/doi/abs/10.1137/080734479.

[6] Kadowaki, T. and Nishimori, H. (1998) Quantum annealing in the transverse Ising model. Physical Review E, 58(5): 5355. https://journals.aps.org/pre/abstract/10.1103/PhysRevE.58.5355.

[7] Anthony, M., Boros, E., Crama, Y., et al. (2017) Quadratic reformulations of nonlinear binary optimization problems. Mathematical Programming, 162: 115-144. https://springer.longhoe.net/article/10.1007/s10107-016-1032-4.

[8] Glover, F., Kochenberger, G., Hennig, R., et al. (2022) Quantum bridge analytics I: a tutorial on formulating and using QUBO models. Annals of Operations Research, 314(1): 141-183. https://springer.longhoe.net/article/10.1007/s10479-022-04634-2.

[9] Nüßlein, J., Gabor, T., Linnhoff-Popien, C., et al. (2022) Algorithmic QUBO formulations for k-SAT and hamiltonian cycles. Proceedings of the genetic and evolutionary computation conference companion, 2022: 2240-2246. https://dl.acm.org/doi/abs/10.1145/3520304.3533952.

[10] Martins, L. N., Rocha, A. P. and Castro, A. J. M. (2021) A QUBO Model to the Tail Assignment Problem. ICAART (2), 2021: 899-906. https://www.scitepress.org/PublishedPapers/2021/102596/102596.pdf.

[11] Naiyang, D. and Yingjie, T. (2009) Support Vector Machines-Theory, Algorithms and Extensions. Science Press, Beijing.

https://book.sciencereading.cn/shop/book/Booksimple/show.do?id=B9F7A368AD6C447F4A7A423FCB10E0831000.

[12] Zaman, M., Tanahashi, K. and Tanaka S (2021). PyQUBO: Python Library for Mapping Combinatorial Optimization Problems to QUBO Form. IEEE Transactions on Computers, 71(4), 838-850. https://arxiv.org/abs/2103.01708.

[13] ABIDE. (2016) Autism Brain Imaging Data Exchange. https://fcon_1000.projects.nitrc.org/indi/abide/abide_I.html.

[14] Martino, A.D., Yan, C. G., Li, Q., et al. (2014) The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Molecular psychiatry, 19(6): 659-667. https://www.nature.com/articles/mp201378.

[15] Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., et al. (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage, 15(1): 273-289. https://www.sciencedirect.com/science/article/pii/S1053811901909784.