ISO/IEC 25000 quality measures for A.I.: a geometrical approach

Andrea Trenta
UNINFO UNI CT 504
Technical Committee System and Software Engineering
Italy
andrea.trenta@dataqualitylab.it

Abstract— **In a previous paper [13] we discussed ISO/IEC 25000 application when new quality measures are defined. In the present paper:**
   - some quality issues in A.I. are identified,
   - then known solutions are recalled and
   - new quality measures for A.I. are proposed

Keywords: data quality, eigenvalue, A.I., AI, image recognition, training, ISO, ISO/IEC 25024

I. INTRODUCTION

In this paper new ISO/IEC 25000 quality measures for dataset used in some A.I. applications are proposed based on [7] and [12]. Furthermore, some considerations are developed about the possible specification and extension of the method to any kind of dataset.

II. Quality issues in A.I.

In this paper, the term A.I. is used for simplicity even when referring to Machine Learning.
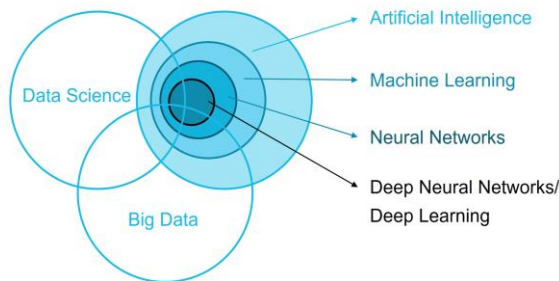

Figure 1 Definitions [14]

Firstly, we consider the A.I. application face recognition, well-known both for the solutions and for the open issues point of view. Among the open issues there is how to understand whether the training dataset is "optimal". To this end, we will explore the measure of completeness characteristic [6] of a set of images supposed to be a training dataset. Note that the measures proposed neither correspond to the measure of the whole A.I. system output results nor to its behavior observation, as it is a purely static measure of the input, although it could be used together with other ones, to evaluate the overall system quality.

The basis of our analysis is the calculation of "eigenfaces" [7], according Karhunen-Loève transformation (PCA) with the following steps:

1. Collect M images with nxn grayscale pixels of faces with similar dimension, light condition, shot, etc.

2. Transform image i (i=1,..M) in a $(n^2 \times 1)$ column vector Γ_i $\{\Gamma_1, \Gamma_2, \dots \Gamma_M\}$;

3. Compute the "average face" $\Psi = \frac{1}{M} \sum_1^M \Gamma_i$

and subtract Ψ to each image and obtain new vectors $\{\Phi_1, \Phi_2, \dots \Phi_M\}$

4. Build the matrix $(n^2 \times M)$ $A = [\Phi_1, \Phi_2, \dots \Phi_M]$ and compute the covariance matrix $(n^2 \times n^2)$ $C = \frac{1}{M} \sum_1^M \Phi_i \Phi_i^T = \frac{1}{M} A A^T$

5. Compute M eigenvalues λ_i (i=1,..M) of matrix $A^T A$ and then eigenvectors of $A A^T$

6. Sort the eigenvalues of C in descending order

7. Choose a number N of eigenvalues, starting from the biggest, in order to represent 95% of their sum η and keep them; the other n^2-N eigenvalues will not be considered

8. Calculate the images dataset as a linear combination of the N eigenvectors defined at step 7

Further steps are defined [7],[12],[13] for face recognition, that is out of the scope of this paper
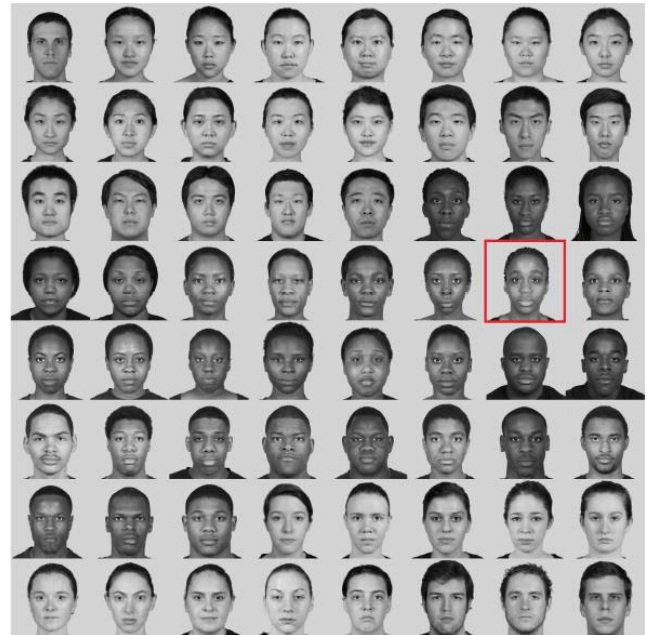

Figure 2 MR2 Face database [11]

III. PROPOSAL

Intuitively, if we want to measure the completeness[1] of an image dataset, we try to answer questions like:

  a. how many similar images are there
  b. how strong is the similarity of some images

Here the proposed measures for (a) and (b):

A.   as some dimensions are eliminated in step 7 above[2], we can measure N\M "PCA space dimension against dataset space dimension"

B.   the more a dataset is orthogonal, the less its images are similar to each other; as a measure of it, we can consider the product of N eigenvalues $\lambda_1 * \lambda_2 * ... * \lambda_N$, that is also the "Determinant of reduced eigenvalues matrix", that in turn is the volume of the hyper-parallelepiped that this matrix represents.

To sum up, with this proposal we reframe the issue of finding an effective data quality (for completeness) measurement function into a well-known geometrical calculation.

IV. FURTHER STUDIES

The steps 1-8 above were proven to be effective in face recognition and are potentially applicable to any dataset. To do this, the vectorization step 2 above shall be applied to any attribute(s) of the dataset: as images were vectorized pixel by pixel[3], similar operation could be performed e.g. for char strings, taking into account possible different lengths that require a further step of normalization.

Further studies are needed to apply the method also to images rotated and translated, that is the most frequent case in A.I. applications (fig.3).
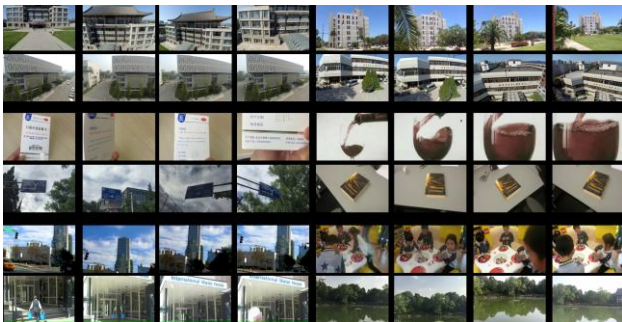

Figura 3 Dataset trial MPEG-CDVA (Compact Descriptor for Video Analysis)– [9] [10]

For the case of a "unsupervised learning", care should be taken in generating the appropriate (i.e. minimum) number of M images, as M appears depending on the kind of dataset (e.g. we expect different M values for rotated or else non-rotated images, for personal names or else tags,…); so, in general, having M>N does not mean that the space is complete, in other words, that every image can be represented, e.g. it could be a new M+1th face that cannot be an acceptable[4] linear combination of existing N images (e.g. a bald face is missing from the dataset of fig.2). If in this case the new M+1th image is added to the training dataset, that corresponds to an "enforced learning". Therefore, some distinction should be made between a machine with a "unsupervised learning" and a machine with "reinforced learning" when evaluating measurements values (A) and (B) over a training dataset.

As bias is critical for the learning dataset quality, the measures (A) and (B) are suggested also for bias[5] measurement, when defining bias as the modification of an ideal fully orthogonal and normalized dataset[6].

V. CONCLUSION

The measures (A) and (B) appear belonging to data quality completeness characteristic [6]. Applying the process described in [8], the measures (A) and (B) can be defined as ISO 25000 conforming measures; they can be considered in SC7 WG6 and SC42 work in progress on A.I., too.

REFERENCES

[1]   ISO/IEC 25010:2011 Systems and Software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models

[2]   ISO/IEC 25012:2008 Systems and Software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Data quality model

[3]   ISO/IEC 25020:2019, Systems and Software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Quality measurement framework.

[4]   ISO/IEC 25022:2016, Systems and Software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Measurement of quality in use.

[5]   ISO/IEC 25023:2016, Systems and Software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Measurement of system and software product quality.

[6]   ISO/IEC 25024:2015, Systems and Software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Measurement of data quality

[7]   Turk, M.A., Pentland, A.P.: Face recognition using eigenfaces. In: Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91. IEEE Computer Society Conference on. pp. 586-591. IEEE (1991)

[8]   D. Natale, A. Trenta: Examples of practical use of ISO/IEC 25000 Proceedings APSEC IWESQ 2019 (CEUR-WS.org, ISSN 1613-0073)

[9]   Compact Descriptors for Video Analysis: the Emerging MPEG Standard 2017 Ling-Yu Duan, Vijay Chandrasekhar, Shiqi Wang, Yihang Lou, Jie Lin, Yan Bai, Tiejun Huang, Alex Chichung Kot, Fellow, IEEE and Wen Gao, Fellow, IEEE

---

[1] It is intended the data quality characteristic "completeness" see [2], [6]

[2] N<M [13]

[3] The method is agnostic respect to the meaning of images (no features extraction and semantic categorization); this is a great simplification that allows to apply the method to other kinds of dataset

[4] i.e. the projection of the M+1th image in the space of faces has an euclidean distance with respect to the other faces below a threshold; in the case the M+1th image is a copy, there is one image in the M-dataset from which the Euclidean distance is 0

[5] For "bias" it is intended the prevalence of some values in an attribute (e.g.: "male" value in "gender" attribute)

[6] Further measures could refer to Hilbert space frame theory

[10] ISO/IEC 15938-15:2019 Information technology — Multimedia content description interface — Part 15: Compact descriptors for video analysis

[11] Strohminger N., Gray K., Chituc V., Hener J., Schein C., Heagins T.B.: The mr2: A multi-racial, mega-resolution database of facial stimuli. Behavior research methods pp. 1-8 (2015)

[12] Sirovich L., Kirby M.: Low-dimensional procedure for the characterization of human faces. Josa a 4(3), 519{524 (1987)

[13] Bagli M. C. (2016) Autovettori e riconoscimento facciale, Università di Bologna, https://amslaurea.unibo.it/12063/

[14] https://www.fokus.fraunhofer.de/en/fame/workingareas/ai