

Learning with small data: what can be inferred from small samples?*

Serge Dolgikh^{1,†}, Oksana Mulesa^{2,3,*} and Volodymyr Sabadosh^{3,†}

¹ National Aviation University, Lubomyra Huzara 1, Kyiv, Ukraine

² University of Presov in Presov, Presov, Slovakia

³ Uzhhorod National University, Universytetska St 14, Uzhhorod, Ukraine

Abstract

The challenge of factor analysis with small datasets is encountered commonly in problems and domains where the amount of data available for analysis may not be sufficient to assure its confidence according to common statistical methods and criteria. While it has been approached from many directions and with different methods, in this work we first proceed to formally define the problem of small data analysis from the information-theoretical perspective as that of “insufficient sampling”: the amount of data below the threshold required for a confident limit on the error of generalization. Below this “minimal sampling” threshold, generalization of the method cannot be assured with statistical confidence. In this work, we discuss approaches in the analysis of small data and establish the conceptual logical framework that incorporates formulation of early hypotheses and verification of their consistency based on iterative sampling. While the conclusion of our analysis is that the problem of insufficient sampling does not have a general solution in all cases, the approaches outlined and discussed here can be instrumental in many practical problems and applications.

Keywords

Small data, statistical analysis, factor analysis, prototype analysis, small sampling

1. Introduction

Methods of factor analysis, statistical and more recently, those based on the methods and models in Machine Learning have proven widely successful and effective in the analysis of complex data of different types and from many sources in a wide range of applications.

In the established practice, conventional methods of data and factor analysis require certain prior information about the object or phenomenon being studied, such as annotations with known categories, assumptions about the character and type of the distribution and others. Several well-established results in the fields of statistics and computer science indicate that formal statistical confidence in the ability of such methods to learn general characteristics in the data, “to generalize” is guided by certain minimal amount of the data, or size of the sample that is specific to the method being applied.

On the other hand, the challenge of factor analysis with small datasets is encountered commonly in problems and domains where the amount of data available for analysis may not be sufficient to assure its confidence according to common statistical methods and criteria. While it has been approached from many directions and with different methods, the principal challenge of the confidence in the statistical significance of such results, based on the theoretical and experimental results, remains open.

In this work we first define the problem of small data analysis from the information-theoretical perspective as that of “insufficient sampling”: where the amount of data, or the size of the sampling

IDDm'24: 7th International Conference on Informatics & Data-Driven Medicine, November 14 - 16, 2024, Birmingham, UK

* Corresponding author.

† These authors contributed equally.

✉ sdolgikh@kai.edu.ua (S. Dolgikh); oksana.mulesa@uzhnu.edu.ua (O. Mulesa); vsabadosh@gmail.com (V. Sabadosh)

ORCID 0000-0001-5929-8954 (S. Dolgikh); 0000-0002-6117-5846 (O. Mulesa); 0009-0006-9933-444X (V. Sabadosh)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

of the unknown distribution is below the minimal threshold required for a confident limit on the error of generalization imposed by the results in theoretical computer science. Below this “minimal sampling” threshold generalization of the method cannot be assured with statistical confidence.

We discuss approaches in the analysis of small data and establish the conceptual logical framework that incorporates formulation of early hypotheses and verification of their consistency based on iterative sampling. While the conclusion of our analysis is that the problem of insufficient sampling does not have a general solution in all cases, the approaches outlined and discussed here can be instrumental in practical problems and applications.

2. Prior Work

The problem domain of factor analysis [1] with small data, that is, determining characteristics patterns, relationships and trends can entail natural and hardly avoidable tensions and challenges related to the very framework of such studies, where the size of the sample may not be large or representative enough to ensure statistical significance of the findings by conventional methods of statistical analysis (hence, the problem of insufficient sampling). On the other hand, early examination of patterns and trends in the emerging data can be beneficial in the novel scenarios and where bodies of confidently annotated data necessary for application of conventional methods of factor analysis simply may not yet have been accumulated and compiled [2].

The challenges that arise from attempts to use conventional methods of pattern analysis [3] have been examined and discussed at length in the literature. Among the well-known ones can be mentioned the strong dependency of the learning success on various training parameters; stability and reproducibility of the results between different samplings; issues with generalizing, i.e., the consistency of the results across different samplings; overfitting; and others [4,5]. Consequently, in many cases, the problem can be characterized as that of a general stability and confidence of learning: the results produced with methods of similar types with the same data can lack consistency and statistical significance according to accepted standards. That, in its turn, can make comparison of different methods, approaches and models less reliable, as it may not be known with sufficient confidence whether the reported results reflect an essential advantage of the method, or an artifact of the experiment.

Numerous attempts were made to approach the problem of stability of learning with small data with case-specific methods and approaches, including: ensemble methods [6,7]; methods adjusted to small data analysis as Radial-Basis Function (RBF) networks [8], prototype learning, including ensemble-based approaches [9] and others [10,11]. However, though some of the results showed promise in specific cases, their general applicability could not be assured due to specialized design, architecture and critical assumptions. In addition, the very methods intended for verification of stability, consistency and generalization, such as cross-validation themselves can be challenged for accuracy and consistency in the scenarios with small-sized training datasets.

In another perspective and a direction of addressing the challenge of insufficient sampling offer models and methods developed in the domain of self-supervised and unsupervised learning [12]. These methods can be instrumental and demonstrated successful ability to interpret and resolve the underlying conceptual structure of the data, regardless of its specific type, thus having general if not near-universal applicability. Their effectiveness was demonstrated in a number of applications, including with complex real-world types of data [13,14].

The potential for these methods in the challenge of small data analysis stems from the fact that they commonly do not require massive prior knowledge of the domain and can be used with “raw” data that does not have, or not yet has confident annotations. In the cases and applications where the constraint limiting the effective size of the training data is that of the prior knowledge, that is, annotations rather than raw data itself, using these methods can offer additional insights and perspectives for the analysis as we discuss further in this work.

To examine and address the challenges outlined in this section and the cited studies, we first attempt to formalize the concept of small data based on the established results in the information

theory and theoretical computer science [15]. Considered in the plane of the question, which samplings are sufficient and not, for the confident generalization of the methods trained with them, a question well researched in the field, the case of small data analysis can be defined as a sector of general factor analysis where the sample available for training is below the minimal threshold necessary for the confident generalization.

In this context of interpretation of the problem of small data, cases and scenarios of the distribution of the data points in the samplings can be considered that can offer essential insights with regards to the conclusion of the analysis and confidence of its findings.

3. The Problem of Small Data as Insufficient Sampling

We consider a general case of some data that describes a process in time or an object, phenomenon, etc., that is obtained as a sampling of a presumably, unknown distribution D , $W = \{ P, F \}$ where $P = \{ p \}$: points of observation that describe “the domain”, such as individual cases/subjects; individuals or social groups and so on; $F = \{ f \}$, the observable factors. Each data point $p \in P$ is described by a set of observable factors $f_p = F(p)$.

Next, we consider the problem of establishing a relation R between certain factor(s) of interest, $K(p)$ that characterizes the data points in the domain, and the observable factors of the data points, up to a certain degree of confidence that can be ascertained in a number of ways, such as evaluation of statistical significance and other methods.

$$K(p) = R(F(p)). \quad (1)$$

The relationship above, “the factor relationship”, is a formulation of the classical problem of factor analysis [1,16]: finding or establishing, within certain criteria of precision, generality and confidence, a relationship R between certain factors of interest related to the domain and the observable factors of the data points in the domain, based on a sampling, or “data” W .

In approaching this problem, it is common to employ certain methods or models that can be seen as effective means of finding or determining the relationship in (1). It is an established in computer science agreement supported by several results, that most if not all of the known methods, at least in the more specific field of supervised learning, require certain minimal amount of prior knowledge for effective, within the specified criteria of accuracy, confidence and generalization, learning. It can be formulated as the relationship between the accuracy of the method m and its ability to generalize (as the essential characteristics of the effectiveness) and the size of the known sample with which it is trained. One of the expressions of this relationship is the Vapnik-Chervonenkis factor $C(m)$ [17]. This factor limits the minimum size of the data W that can provide a confident bound on the error of the generalization of the method.

Then, in the cases where the size of the training sample is below the minimal threshold for a given method, the generalization of the method cannot be assured. In other words, it can overfit by failing to reproduce the level of accuracy achieved with the training sample, with a different or general sample of data.

This brief discussion leads to the formulation of the problem of small data: where the size of the known, “training” sample W is lower, possibly significantly, than the minimal required for confident generalization, what knowledge, if any, can be inferred about the domain it is a sample of? In this setting one needs to consider the case that is opposite of that of standard, conventional supervised learning.

$$| W | \lesssim C(m); \text{ or } | W | \ll C(m). \quad (2)$$

Note that (2) also presents a formal definition of the “smallness” of data, which is dependent on the method of representation.

In this work we will attempt to offer some insights into this problem.

3.1. Sampling: Relevance, Representativity and Descriptiveness

For an illustration of the problem of sampling, let us consider an example. Suppose the horizontal axis represents the distribution sampled by data $X = \{ x \}$, whereas the vertical axis measures the factor of interest: $y = K(x)$. We consider several possible scenarios of the composition of a small sample S defined as discussed earlier, relative to the general characteristics of the distribution and the factor relationship (1).

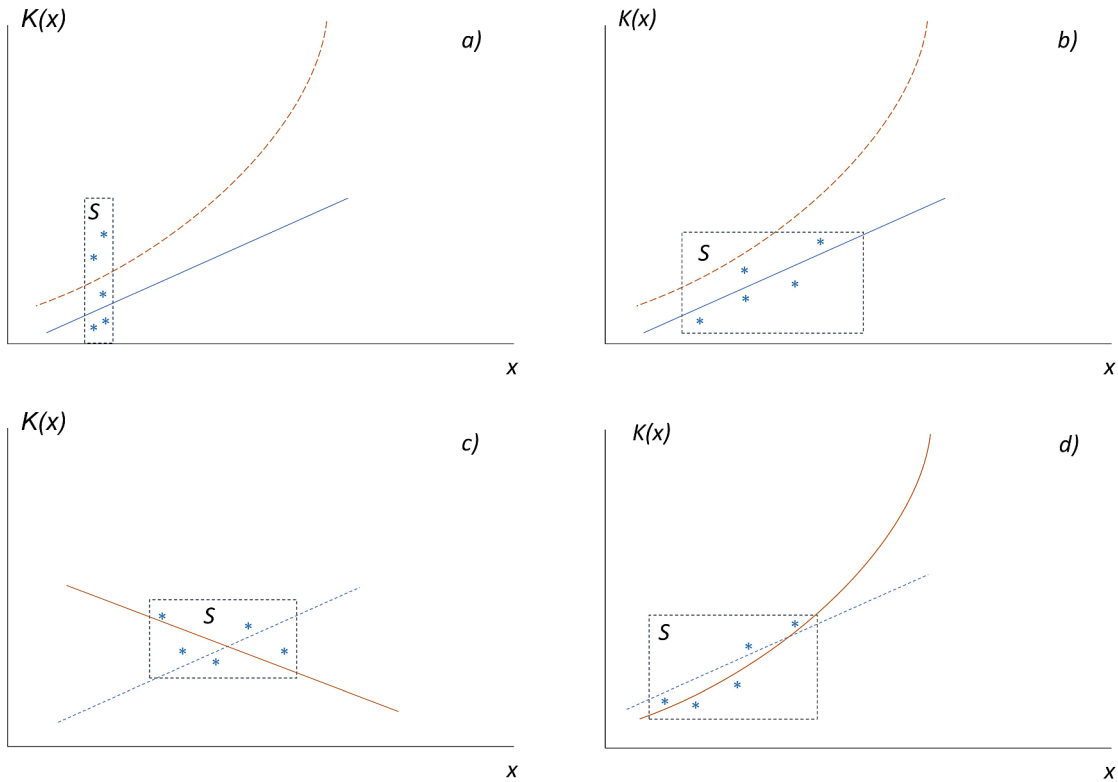


Figure 1: Scenarios in small sampling.

An immediate conclusion that can be derived from the examples above is that there is no general universal solution to the problem of small data, that is, any combination of *sampling*; *observable factors* and *the factor relationship*. Some combinations may not have a solution, whereas others can have essential limits or constraints on the generality and accuracy of the approximation of the factor relationship.

3.1.1. Relevance

The scenario shown in the diagram *a*), Figure 1 is an example of the case that can be referred to as “irrelevant sampling”: the samples do not cover any meaningful range of the variation of the unknown distribution D . Though it may seem obvious in the diagram with the minimal set of descriptive factors, in practical cases and applications the relationship between the observable factors, often of a large quantity, and the informative factor(s) that are correlated with and therefore, describe the variation in the distribution can be challenging to identify. Then, a small sample S , representativity of which across the informative domain of variation of D cannot be assured, can fall into this category, sampling stochastic variation of the factor of interest in a very narrow interval of the distribution, or even artifacts of measurement and/or statistical error.

Clearly, no meaningful relationship (1) can be established in this case based on the sampling S , and the problem of factor analysis with a small sample as a given input to the problem does not have a solution in this case.

3.1.2. Representativity and Sufficiency

A different scenario: b), Figure 1 demonstrates that under certain conditions, small sampling can offer some insights into the character of the unknown distribution, perhaps as a formulation of initial hypotheses, into the character of the factor relationship sought in the formulation of the problem (1). The caution that has to be exercised here is that it is only one possibility among several competing options, given that statistical confidence of the approximation of the factor relationship could not be established, as discussed earlier in (2).

For these reasons it cannot be assumed to be the case automatically and needs to be justified either by additional research and argumentation; or by collection of more data and ongoing verification up to the point where the significance of the hypothesis can be substantiated quantitatively at the acceptable formal level.

Next, let us compare the scenarios b) and d), Figure 1. Both samples can be used in formulation of the initial hypothesis on the character of the factor relationship, such as linear approximation. However, one can observe that the sampling in the latter scenario, d) represents only a limited range in the distribution of the informative variable, that is insufficient to determine the character of the underlying factor relationship correctly and with confidence: indeed, extending it (i.e., the sampling) to the range toward the greater x would have had a significant impact on the approximation of the relationship in this case. Then, it can be determined that the sample was insufficiently representative with respect to essential characteristics of the distribution in the range of its variation that resulted in an error of approximation of the factor relationship.

A similar conclusion can be inferred from comparing the cases b) and c): whereas the effective range of variation in the sampling in the former case can be seen as sufficient at least for a formulation of the hypothesis, the sampling in the latter case does not allow to distinguish between the alternative hypotheses with any confidence. Then, this case can be classified as that of insufficient representativity of the sampling as well.

This observation underlines the challenges and constraints of confidence, accuracy and generality that can be attributed to insufficiency of small samples to provide confident approximations of the factor relationship (1).

3.1.3. Descriptiveness

As noted earlier, in practical cases and applications the distribution D is commonly not known precisely and its sampling is expressed by a large set of observable factors. Let us assume for a moment that there exists another variable: informative or “latent” factor(s), $l(p)$ that can be calculated from some observable factors that are not necessarily the ones the data is sampled in: $\tilde{f}(p)$, and are in a good correlation with the factor of interest $K(p)$. Then, an explicit relationship:

$$l(p) = \tilde{f}(p), \quad (3)$$

along with a sampling of D in the observable factors \tilde{f} would provide a solution to the problem of factor analysis (1). The challenge is of course that in practice most often neither of the relationship (3), and the “effective” observable factors \tilde{f} are known at prior, and have to be found, calculated or approximated by some method. In fact, even the existence of such informative factors for any possible factor of interest cannot be assured.

Then the observations made above for the sampling scenarios in can be fully applied to mappings from observable factors to the informative ones. Because the effective observation factors are not usually known, a small sampling expressed in some set of observable factors can in fact represent a case of irrelevant sampling, mapping to a small region of the variation of the sought distribution. Such a scenario can be classified as a “descriptiveness” problem: the chosen set of observable factors is insufficient to capture the essential parameters of the distribution of the problem D , and may cause irrelevance of samplings expressed in them to the problem.

This observation again reinforces the conclusion that not every combination of a sampling, observable factors and the problem has a solution as defined in (1).

3.2. No General Solution in Small Data Problem

Based on the examples considered in this section and their analysis, one can summarize it with a substantiated conclusion that the formal problem of factor analysis with small data as defined in (1), (2) *does not have a solution* in the general case; that is, any combination of the sample; method; and the choice of observable or descriptive factors that describe data points in the sampling set.

Therefore, in each particular case a specific analysis and verification of the aspects of representativity, descriptiveness and sufficiency of the sampling must be performed as outlined in this section, before proceeding to the analysis of the factor relationship (1) by the formal means and methods.

4. Methods in Small Data Analysis

Following the discussion earlier in this work, we will attempt to outline certain approaches to analysis of small samplings. It does not claim the breadth or comprehensive nature of the coverage that can be found in the reviews cited earlier and other literature. As was noted, insufficient statistical confidence is a characterizing staple in working with small datasets. It means working in a frame where one can never ignore other possibilities, such as irrelevant sampling and other cases where a solution to the problem of factor analysis with a small sampling may not exist; and interpret all results as a hypothesis rather than confident finding. On this basis we will consider two broad families of methods that can be applied in the analysis of small data, without exclusion or supposition of limitation of applicability of different approaches.

4.1. Statistical Analysis, Regression and Multivariate Factor Analysis with Small Data

Methods of statistical analysis can be used to determine significant factors of influence, most commonly, via calculation statistical correlation between certain observable factors $\{f\}$ and the factor of interest K . In the simplest form it can be expressed as:

$$C_f = \text{Corr}(W(f), K), \quad (4)$$

where $W(f)$: the column of the data W at factor f , $\text{Corr}(a,b)$: correlation factor (such as correlation coefficient of the vectors a, b) [1,18].

For an illustration let us return to the example of small data distributions and initial hypotheses in Figure 1. Considering the scenario d) in the above, one can formulate the initial hypothesis or “trend” of the distribution (K, x) as a parametric function $K = (a_1, a_2, \dots, a_k)(x)$, such as a linear function, $a_1 x + a_2$ if linear regression is applied.

Now, as always in the framework of working with small data, the formal statistical significance of the hypothesis based on the initial data may not be sufficient for a confident conclusion on its validity. This is why one has to take into account the concept of “future data”.

Indeed, a common observation requires that a valid hypothesis would gain significance with accumulation of new data all the way to formal threshold of confidence; on the other hand, irrelevant or spurious hypothesis may not be aligned with the new data, in other words, *irrelevant hypothesis lacks predictive power*. Then, one can propose two directions of the consistency analysis of the formulated early hypothesis based on the new data:

1. Trend drift: how does the new data impact the trend obtained from the initial analysis? Is it consistent with it (low trend drift) or does it cause significant change of the trend?
2. Error or margin drift: how does the addition of the new data affect the error, for example, standard deviation of the data from the initial trend?

Again, we will attempt to illustrate these approaches with scenarios of distributions of samplings, the initial and future ones. Shown in Figure 2 shown are two possible scenarios of “future” samplings, related to the cases considered earlier. It will be assumed that “future” samples, S_{2a} and S_{2b} were obtained at later point than the initial small set, S_1 .

In the first scenario, with the new sample S_{2a} one can calculate the new trend: $t_{2a} = (a_1^{2a}, a_2^{2a})$ assuming the linear factor relationship, and compare it to the initial one. A significant difference between the initial and subsequent trends can indicate incompatibility of the relationship calculated with the fuller dataset with the initial hypothesis.

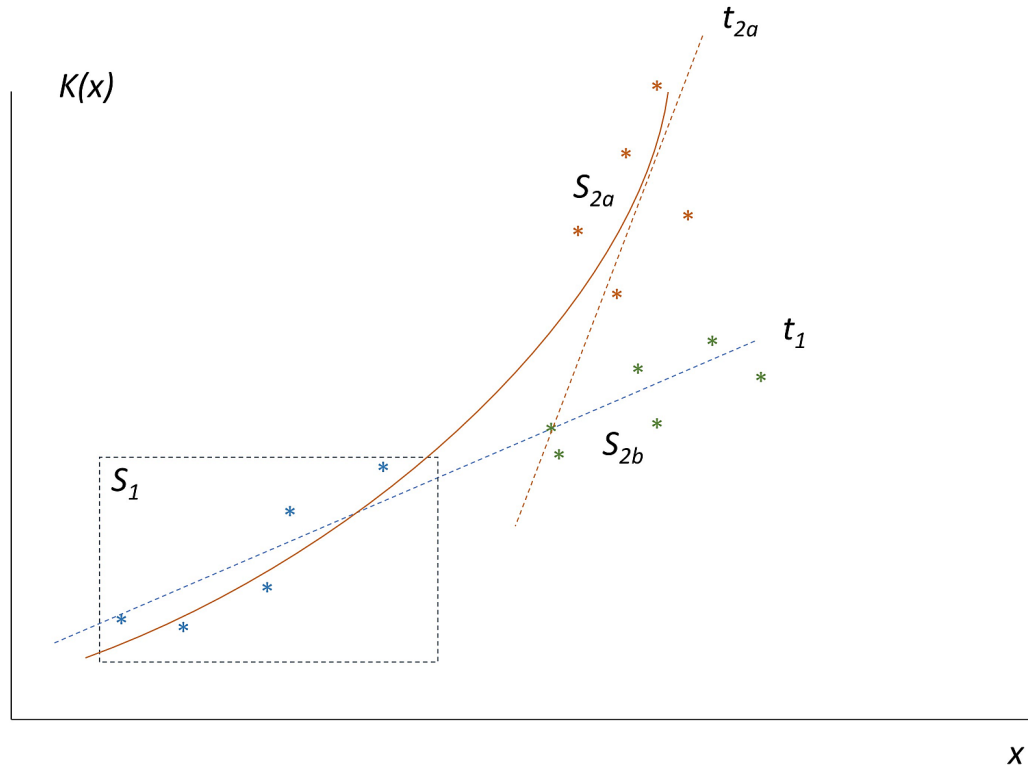


Figure 2: Iterative approach in estimating statistical confidence.

The same conclusion can be obtained from the error drift analysis. Indeed, in this scenario one would observe an increase in the average error with the addition of new data. That observation would not be compatible with the correctness of the initial hypothesis.

This brief illustration of course confirms the guidelines made as the general logical framework of factor analysis with small data: any initial indications can be taken only as an initial hypothesis subject to verification with more representative samplings.

In the second scenario, S_{2b} , one can observe that both the initial trend and the error are stable and consistent. This observation, verified with several independent samples can result in the determination of the validity of the initial hypothesis under certain criteria of statistical confidence.

It can be noted in conclusion of this section, that methods of statistical analysis considered here can be applied effectively with the type of samplings (data) described by a large number of factors of different characteristics, types and formats i.e., multivariate heterogeneous observable factors type of sampling/problem.

4.2. Generative Prototype Analysis

As mentioned in the introduction section, this approach is based on the observation that methods of unsupervised generative learning can be instrumental in the analysis of the structure of the data, expressed in observable parameters, without the need for known association with the factors of interest (for example raw, unannotated data). These methods can be effective with the types of data

described by a large number of similar observable factors, indicating a possibility of a strong redundancy in the observable factors (multiple homogeneous observable factors) [19,20].

The advantage of these methods in application to the problem of small data analysis stems from the possibility that while prior data in the problem can be limited, it may not necessarily be so for the general, non-annotated data. Then, certain analysis of the structure of the general data that can be performed and offer additional informative insights into the composition of the sample.

For an illustration of the approach, let us consider the scenario described in [9], where an observable distribution of the aforementioned type was modeled by a dataset of images of geometric shapes (Figure 3).

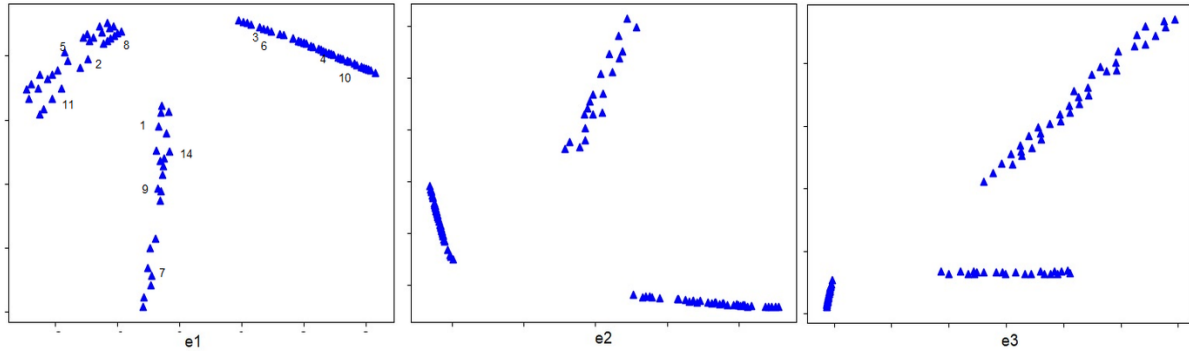


Figure 3: Latent distribution of multifactorial homogeneous data and latent cluster analysis (from [9]).

In this example, different groups of samples characterized by closer similarity in the general distribution are modeled by the types of the shape in the images. It corresponds to the case where an unknown general distribution is described by a large number of observable factors that have approximately equal weight or significance in describing the object or entity in the distribution (hence, homogeneous multifactorial description of the problem).

Then, as described in the cited work, in many cases, a decomposition of the general sample D into a collection of general types $T(D)$ can be determined with sufficient confidence by the methods of unsupervised ensemble learning that do not require, generally, much or any prior knowledge about the distribution. Such a decomposition can then reduce the problem of factor analysis by the population, that is inherently constrained by the size of the sample in the case of small data, to that of the analysis by the general type or cluster [21].

A structure of general types (population clusters) can be seen clearly in the diagram above, along with the regions of their distribution in the informative latent space.

While a decomposition of this type may not immediately solve the problem of factor analysis, it can offer additional perspectives of study as and when more confidently known data becomes available, such as:

- Cluster correlation analysis: correlation of the general types or population clusters, $T(x)$ with the factor of interest, $K(x)$. For example, some population clusters can show higher (or lower) statistics of the distribution of the factor of interest than across the entire population. In this case, the correlation hypothesis of the factor K with the characteristics of the clusters can be proposed and studied, until a confident conclusion can be reached.
- If, on the contrary, a significant correlation of the factor of interest with the population clusters could not be observed with a growing body of verified data, it may point at the possibility of a descriptiveness problem with the observable factors (Section 3.1.3) that may not be sufficiently detailed or “granular” to differentiate the hidden factor(s) that are correlated with the factor of interest. In that case, the conditions of the study may need to be corrected, possibly by addition of more descriptive observable factors.

- Intra-cluster analysis can be useful in detecting marginal or irrelevant samplings. For example, if a distribution of the factor of interest in the same population cluster shows unexplainable variance it can point either to the case of irrelevant sampling; or again, insufficient descriptiveness of the observable factors of the analysis.

Thus, in the approach outlined in this section, application of methods of unsupervised generative learning, of which many types and models have been developed, linear and non-linear, can offer new additional perspectives and insights into the problem of factor analysis with small data.

5. Conclusions

In this study, we approached the problem of factor analysis with small datasets as that of insufficient sampling, in the view of formal information-theoretical requirements for bounded generalization error.

Formulating it in this way allowed first, to define formal criteria for the “smallness” of the data that are not universal but rather, specific problem and method. The second essential conclusion stemming immediately from this framework of analysis is that the problem of factor analysis with small data does not have a general solution in all cases, that is, any combination of sampling, the choice of observable factors and the method of analysis.

These conclusions can have profound significance for working with small data, especially in novel problems, scenarios and situations where large sets of confidently annotated data simply may not exist. In this domain of analysis, one may not expect the conclusions to reach the level of firm statistical confidence but rather be considered early hypothesis to be verified by other methods and/or with more data as and when it becomes available. One has to be aware of the cases that were identified and described here, where the problem may not have a solution, such as irrelevant sampling; insufficient representativity and descriptiveness. Working within the framework of these conclusions and guidelines can improve the effectiveness of formulating early hypotheses and avoid expensive pitfalls.

The challenge of approaching the problem of factor analysis with samplings that may not be sufficiently large to ensure statistical confidence with the conventional methods can be described as the trade-off between shortening the span of the research process, especially in the formulation of early hypotheses and the confidence of its conclusions. The observations, relationships and hypotheses found in the early phase of the cycle will have to be verified and confirmed with larger sets of data as and when they become available. Being aware of the limitations and conditions of working with small data identified and discussed in this work can ensure the cumulatively positive effect of such studies, in the formulation of early hypothesis and offer valuable insights for further examination.

New and novel problem areas where large bodies of knowledge have not yet been accumulated emerge in today’s science with regularity and frequently. We hope that the approaches and conclusions in working with early samplings of such problems developed in this work will be of value to the research community in data science and factor analysis.

Declaration on Generative AI

1. **Tools and services:** GenAI tools were not used in preparation or editing of this work.
2. **Tools’ contributions:** GenAI tools were not used in preparation or editing of this work.

References

- [1] R. L. Gorsuch, Factor Analysis, 2nd. ed., Chronicle Books, San Francisco, 1983.
- [2] E. B. Hekler, P. Klasnja, G. Chevance et al., Why we need a small data paradigm, BMC Medicine, 17 (1) (2019) 133.

- [3] J. A. Richards, Supervised classification techniques. Remote Sensing Digital Image Analysis, Springer, Berlin, Heidelberg 247–318 (2013).
- [4] P. Cunningham, J. Carney, S. Jacob, Stability problems with artificial neural networks and the ensemble solution, *Artificial Intelligence in Medicine*, 20 (3) (2000) 217–255.
- [5] G. Forman, I. Cohen, Learning from little: comparison of classifiers given little training, in: *Proceedings of PKDD 2004*, volume 19, pp. 161–172.
- [6] D. Opitz, R. Maclin, Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research*, 11 (1999) 169–198.
- [7] Y. Chen, D. Li, X. Zhang, J. Jin, Y. Shen, Computer-aided diagnosis of thyroid nodules based on the devised small-datasets multi-view ensemble learning, *Medical Image Analysis*, 67 (2021), 101819.
- [8] I. Izonin, R. Tkachenko, I. Dronuyk et al., Predictive modeling based on small data in clinical medicine: RBF-based additive input-doubling method, *Mathematical Bioscience Engineering*, 18 (3) (2021) 2599–2613.
- [9] S. Dolgikh, Modeling of small data with unsupervised generative ensemble learning, in: *Proceedings of the 5th International Conference on Informatics and Data-Driven Medicine (IDDM-2022) Lyon France 2022*, CEUR-WS.org volume 3302 pp. 35–43.
- [10] A. Elisseeff, T. Evgeniou, M. Pontil, Stability of randomized learning algorithms, *Journal of Machine Learning Research*, 6 (2005) 55–79.
- [11] P. Xu, X. Ji, M. Li, et al., Small data machine learning in materials science. *NPJ Computational Materials* 9 (2023) 42.
- [12] R. Krishnan, E. Rajpurkar, E. J. Topol, Self-supervised learning in medicine and healthcare. *Nature Biomedical Engineering*, 6 (2022) 1346–1352.
- [13] Y. Bengio, A. Courville, P. Vincent, Representation Learning: a review and new perspectives, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35 (2012) 1798–1828.
- [14] W. Liu, Z. Wang, X. Liu et al.: A survey of deep neural network architectures and their applications, *Neurocomputing*, 234 (2017) 11–26.
- [15] L. Gondara, Medical image denoising using convolutional denoising autoencoders, in: *Proceedings of the 16th IEEE International Conference on Data Mining Workshops (ICDMW)*, Barcelona, Spain, 2016, pp. 241–246.
- [16] S. Dolgikh, O. Mulesa, Covid-19 epidemiological factor analysis: identifying principal factors with Machine Learning, in: *Proceedings of the 7th International Conference "Information Technology and Interactions" (IT&I-2020) Kyiv Ukraine 2020*, CEUR-WS.org volume 2833 pp. 114–123.
- [17] V. N. Vapnik, A. Y. Chervonenkis, On the uniform convergence of relative frequencies of events to their probabilities, *Theory of Probability & Its Applications*, 16 (2) (1971) 264.
- [18] H. Wendland, *Scattered data approximation*, Cambridge University Press, 2005.
- [19] M. Biehl, B. Hammer, T. Villmann, Prototype-based models in machine learning, *WIRE's Cognitive Science* 7(2) (2016) 92–111.
- [20] S. Dolgikh, Categorization in unsupervised generative self-learning systems, *International Journal of Modern Education & Computer Science*, 13 (3) (2021) 68–78.
- [21] M. Ester, H.-P. Kriegel, J. Sander et al., A density-based algorithm for discovering clusters in large spatial databases with noise, in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96) 1996*, pp. 226–231.