

# Discussion on “Big data mining for industry 4.0: the new pathway for the digital and green ‘twin’ transition”

---

Antonio Canale

University of Padova

antonio.canale@unipd.it

1222·2022  
**800**  
ANNI



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

September, 21-23 2022

- Paradigm shift also in many other contexts (the complexity of the world translates also in complex data)
- Novel technologies lead to novel challenges
  - Medical diagnosis
  - Neurosciences and brain imaging
  - Social Sciences
- Last few years have been an exciting period for statisticians
  - I graduated almost when the notorious “sexy job” quote was said (*“The sexy job in the next 10 years will be statisticians”* H. Varian, 2009)
  - *“Every time the amount of data increases by a factor of ten, we should totally rethink how we analyze it”* (J. Friedman, 1997)

- Paradigm shift also in many other contexts (the complexity of the world translates also in complex data)
- Novel technologies lead to novel challenges
  - Medical diagnosis
  - Neurosciences and brain imaging
  - Social Sciences
- Last few years have been an exciting period for statisticians
  - I graduated almost when the notorious “sexy job” quote was said (*“The sexy job in the next 10 years will be statisticians”* H. Varian, 2009)
  - *“Every time the amount of data increases by a factor of ten, we should totally rethink how we analyze it”* (J. Friedman, 1997)

- Paradigm shift also in many other contexts (the complexity of the world translates also in complex data)
- Novel technologies lead to novel challenges
  - Medical diagnosis
  - Neurosciences and brain imaging
  - Social Sciences
- Last few years have been an exciting period for statisticians
  - I graduated almost when the notorious “sexy job” quote was said (*“The sexy job in the next 10 years will be statisticians”* H. Varian, 2009)
  - *“Every time the amount of data increases by a factor of ten, we should totally rethink how we analyze it”* (J. Friedman, 1997)

- Paradigm shift also in many other contexts (the complexity of the world translates also in complex data)
- Novel technologies lead to novel challenges
  - Medical diagnosis
  - Neurosciences and brain imaging
  - Social Sciences
- Last few years have been an exciting period for statisticians
  - I graduated almost when the notorious “sexy job” quote was said (*“The sexy job in the next 10 years will be statisticians”* H. Varian, 2009)
  - *“Every time the amount of data increases by a factor of ten, we should totally rethink how we analyze it”* (J. Friedman, 1997)

- Colosimo et al. (2021, J. Qual. Tech.) is an example where using off-the-shelf methods (after a data preprocessing that ignores complexity) leads to a loss of information
- There is a dark side of this: parsimony in data management
- *Dietary data analysis*
  - A cliché in multivariate statistics
  - PCA, matrix factorizations, projections, embedding
- Maybe the whole big data story is just the matter of reasonable data reduction? Big in size or in complexity?

- Colosimo et al. (2021, J. Qual. Tech.) is an example where using off-the-shelf methods (after a data preprocessing that ignores complexity) leads to a loss of information
- There is a dark side of this: parsimony in data management
- *Dietary data analysis*
  - A cliché in multivariate statistics
  - PCA, matrix factorizations, projections, embedding
- Maybe the whole big data story is just the matter of reasonable data reduction? Big in size or in complexity?

- Colosimo et al. (2021, J. Qual. Tech.) is an example where using off-the-shelf methods (after a data preprocessing that ignores complexity) leads to a loss of information
- There is a dark side of this: parsimony in data management
- *Dietary data analysis*
  - A cliché in multivariate statistics
  - PCA, matrix factorizations, projections, embedding
- Maybe the whole big data story is just the matter of reasonable data reduction? Big in size or in complexity?

- Colosimo et al. (2021, J. Qual. Tech.) is an example where using off-the-shelf methods (after a data preprocessing that ignores complexity) leads to a loss of information
- There is a dark side of this: parsimony in data management
- *Dietary data analysis*
  - A cliché in multivariate statistics
  - PCA, matrix factorizations, projections, embedding
- Maybe the whole big data story is just the matter of reasonable data reduction? Big in size or in complexity?

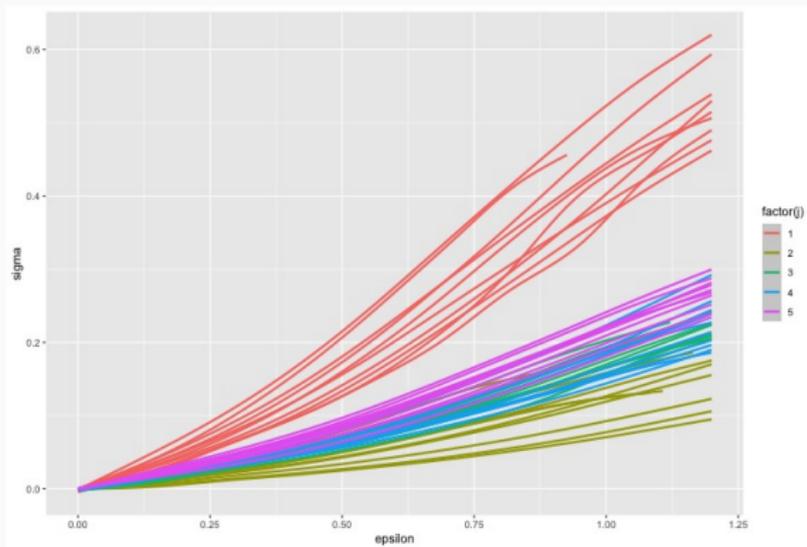
- Colosimo et al. (2021, J. Qual. Tech.) is an example where using off-the-shelf methods (after a data preprocessing that ignores complexity) leads to a loss of information
- There is a dark side of this: parsimony in data management
- *Dietary data analysis*
  - A cliché in multivariate statistics
  - PCA, matrix factorizations, projections, embedding
- **Maybe the whole big data story is just the matter of reasonable data reduction? Big in size or in complexity?**

- Analyzing video of welding/production allows quicker/more efficient anomaly detection
- See Colosimo & Grasso, (2018 JQT); Yan, Grasso, Paynabar & Colosimo - (2022, IISE Trans); Bugatti & Colosimo (2022, J. Int. Manufact.)
- Data stream
- How to decline a dietary data-analysis in a context of data stream?

- Analyzing video of welding/production allows quicker/more efficient anomaly detection
- See Colosimo & Grasso, (2018 JQT); Yan, Grasso, Paynabar & Colosimo - (2022, IISE Trans); Bugatti & Colosimo (2022, J. Int. Manufact.)
- Data stream
- How to decline a dietary data-analysis in a context of data stream?

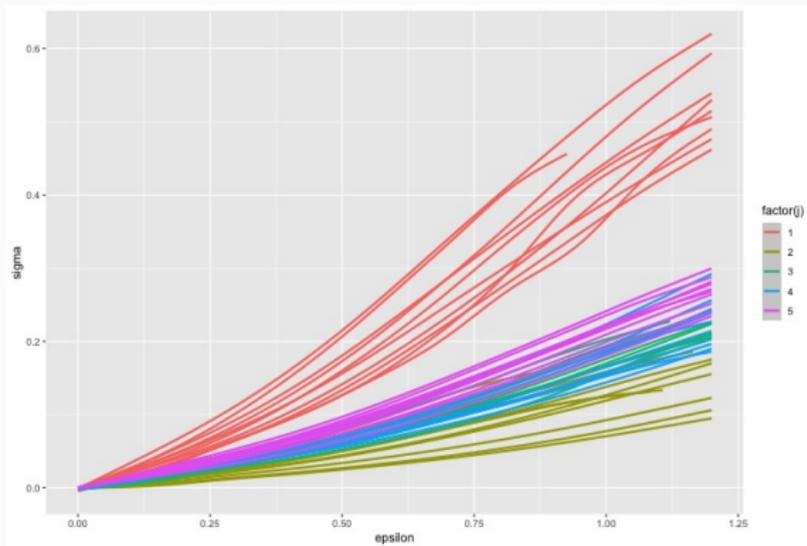
- Analyzing video of welding/production allows quicker/more efficient anomaly detection
- See Colosimo & Grasso, (2018 JQT); Yan, Grasso, Paynabar & Colosimo - (2022, IISE Trans); Bugatti & Colosimo (2022, J. Int. Manufact.)
- Data stream
- How to decline a dietary data-analysis in a context of data stream?

- Analyzing video of welding/production allows quicker/more efficient anomaly detection
- See Colosimo & Grasso, (2018 JQT); Yan, Grasso, Paynabar & Colosimo - (2022, IISE Trans); Bugatti & Colosimo (2022, J. Int. Manufact.)
- Data stream
- **How to decline a dietary data-analysis in a context of data stream?**



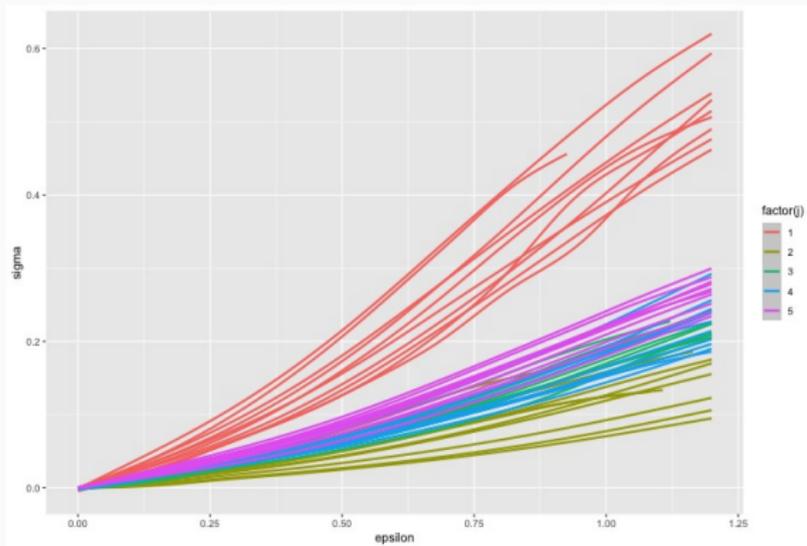
Barbon et al. (2020, Int. J. Molecular Sc.)

- In many contexts, there is huge availability of *prior* information
- Can a Bayesian approach be adopted in these contexts? Which are the possible pros and cons?



Barbon et al. (2020, Int. J. Molecular Sc.)

- In many contexts, there is huge availability of *prior* information
- Can a Bayesian approach be adopted in these contexts? Which are the possible pros and cons?



Barbon et al. (2020, Int. J. Molecular Sc.)

- In many contexts, there is huge availability of *prior* information
- Can a Bayesian approach be adopted in these contexts? Which are the possible pros and cons?

- A lot of modern data analysis is today called machine learning (ML) (or even deep learning and artificial intelligence)
- ML focus is on accurate prediction and not on learning the latent features driving the phenomena
- ML and deep learning often use black-box algorithms
- Is the need of interpretability in contrast with this?



- A lot of modern data analysis is today called machine learning (ML) (or even deep learning and artificial intelligence)
- ML focus is on accurate prediction and not on learning the latent features driving the phenomena
- ML and deep learning often use black-box algorithms
- Is the need of interpretability in contrast with this?



- A lot of modern data analysis is today called machine learning (ML) (or even deep learning and artificial intelligence)
- ML focus is on accurate prediction and not on learning the latent features driving the phenomena
- ML and deep learning often use black-box algorithms
- Is the need of interpretability in contrast with this?





- A lot of modern data analysis is today called machine learning (ML) (or even deep learning and artificial intelligence)
- ML focus is on accurate prediction and not on learning the latent features driving the phenomena
- ML and deep learning often use black-box algorithms
- **Is the need of interpretability in contrast with this?**



- Big data: Big in size or in complexity?
- Green transition: How to decline a *dietary data-analysis* in a context of data stream?
- Existing knowledge: Can a Bayesian approach be adopted in these contexts? Which are the possible pros and cons?
- Interpretability: Is the need of interpretability in contrast with the current trends in ML and deep learning?
- Thanks!

- Big data: Big in size or in complexity?
- Green transition: How to decline a *dietary data-analysis* in a context of data stream?
- Existing knowledge: Can a Bayesian approach be adopted in these contexts? Which are the possible pros and cons?
- Interpretability: Is the need of interpretability in contrast with the current trends in ML and deep learning?
- Thanks!