

„New“ Techniques for Site Engineering:
Kano and Jobs-To-Be-Done Revisited

Daniel Baier, Franziska Sophie Kullak, Herbert Woratschek

Nowadays, many online-fashion shops as well as the stationary trade are permanently confronted with the question of how to improve their service processes. Therefore, site engineering models help companies to develop or improve customer-centered services. The Kano model for customer satisfaction is one out of many quantitative methods where hypothetical testing of innovations can be applied. But does the Kano model really address profound customer needs? Does its application lead to new innovation potential? A possible remedy of the problem may be the in practice recently well discussed qualitative idea-generating method *Jobs-To-Be-Done (JTBD)*. The aim of this study is to explore how *JTBD* could be applied for innovation potential in fashion retail. Human beings hire to ensure the best possible way of addressing certain functional, personal or social jobs according to their situation and the specific products or services which they require. By means of a qualitative study (N=14) personal and social jobs which human beings try to accomplish when doing offline and online shopping were identified and innovation potential has been derived. The results show that in the case of shopping in stationary trade, personal and social jobs like taking time-out from daily duties or maintaining social contacts are of prime importance. In contrast, the purpose behind online shopping seems to be primarily purchasing products that fulfil a functional need. The main potential for innovation in offline fashion retail lies in autonomous shopping, whilst online restructuring of online shops, based on specific needs of customers should be considered to enhance the overall shopping experience.

Components, Cluster (and Correspondence): A data driven approach to consumer behaviour segmentation

Oliver A. Gansser, Karsten Lübke

FOM Hochschule für Oekonomie & Management

Dortmund

Marketing departments usually prefer to cluster customers by predetermined patterns or types to work out marketing strategies that fit in a segment-specific matter. But the theoretical, methodological and empirical background of derived segmentations offered by marketing research institutes is most often rather vague.

With data of more than 20,000 face-to-face interviews with consumers in Germany we aimed at a stable, simple and informative segmentation model based on values (Schmidt et al., 2007) and consumer styles (Sproles and Kendall, 1986) since these may influence the buying process. Therefore a Principal Component Analysis with item selection followed by a Cluster Analysis is conducted for both sets of components. Both methods are optimized concerning the prediction stability (Josse and Husson, 2012; Tibshirani and Walther, 2005). By combining these data analytic methods a consumer behaviour segmentation is proposed which may be easy to handle for researchers and marketers as well. Moreover the results are checked in the following years.

Singular Least Squares and the Analysis of Simple Conjoint Experiments
Andreas Geyer-Schulz
Karlsruher Institut für Technologie

For experimental designs with singular design matrices, the solution by general least squares (with a Moore-Penrose inverse) has been proposed by several authors. However, no consistent interpretation of the mathematical solution (a point and a linear affine operator) has been given. In this contribution, we give a consistent interpretation of this solution in analogy to the well known technique of the measurement of velocity with regard to a reference coordinate system: When the reference coordinate system changes, the change in the reference coordinate system is expressed by a linear affine operator.

Guiding supervised classification algorithms via semantic knowledge
Hans Kestler
Ulm University

Classification of influences for SNS text data regarding radiation

Takafumi Kubota¹, Hiroyuki A. Torii², Michikuni Shimo³, and Kazuko Uno⁴

Abstract

This paper is a report about the research which was gathered on how to classify influences for SNS text data regarding radiation. As for the research background, after the accident of TEPCO's Fukushima Daiichi Nuclear Power Station in March 2011, a lot of information about the keyword relating to "the radiation" has been taken up in the media, and a group of people who receives the media has been becoming uneasy and dissatisfied about the information. To solve this issue, the Japan Health Physics Society started a website of questions and answers which related radiation (radiation QA). Within this, the experts have been answering towards the questions that the people asked (mainly the people in metropolitan area and the Fukushima prefecture). The goal of this study is to quantify the influence of radiation QA by classifying radiation QA and other posts in Tweet data. It is also the goal to see how spreading tweets of influencers tweet.

Acknowledgement

This work was supported by Research on the Health Effects of Radiation organized by Ministry of the Environment, Japan. "Creating Effective Scientific Crisis Communication Methods: Big data analytics to assess the ripple effects of how scientist disseminated information regarding radiation effects after March 2011"

¹Tama University kubota@tama.ac.jp, ²The University of Tokyo, ³Fujita Health University, ⁴Louis Pasteur Center for Medical Research

References

The Japan Health Physics Society (2014), Expert's Answers: Living of radiation Q and A (In Japanese), URL: <http://warp.da.ndl.go.jp/info:ndljp/pid/8699165/radi-info.com/> (2017-08-01)

Keywords

TEXT MINING, TWEET DATA, RADIATION

Applying tensor decompositions to author name disambiguation of common Japanese researcher names

Kei Kurakawa¹ and Yasumasa Baba²

Abstract

In the digital age of scholarly information environment, there has existed author name ambiguity problem of databases and digital libraries, which has not been solved sufficiently for many decades. Two major approaches have been taken to this problem until recent years, i.e., machine learning based name disambiguation and assigning identifiers to authors mainly by hand. Ferreira et al (2012) provided the taxonomy of automatic name disambiguation methods, which focuses on the method of classification, clustering and external knowledge types. ORCID is an emerging approach of assigning identifiers to researchers dealt with international academic publishers and scholarly stakeholders.

We aim at looking around automatic methods of author name disambiguation, and seeking another possible method among others. Our research focuses on tensor decompositions that were recently under the spotlight as a technique for data mining, data analysis, and data science (Kolda et al, 2009). Tensor represents n-dimensional array of data and can be decomposed in several ways, in which popular decompositions are CP and Tucker decompositions. Hitchcock proposed the idea of the polyadic form of a tensor as the sum of a finite number of rank-one tensors in 1927, and Carroll and Chan, and Harshman proposed independently the same idea as is called CP (CAN-DECOMP/PARAFAC) in 1970. Tucker proposed the Tucker decomposition in 1966. In recent years, Nickel et al (2011) proposed RESCAL, a Tucker decomposition of third-order tensor which subjects to the constraint that two factor matrices are the same and the other factor matrix is identical, in the context of relational learning. We adopt the tensor decompositions to a three dimensional array of data consisting of bibliographic attributes so as to de-

1. National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda, Tokyo, Japan kurakawa@nii.ac.jp · 2. The Institute of Statistical Mathematics, 10-3 Midoricho, Tachikawa, Tokyo, Japan baba@ism.ac.jp

rive a feature of author from a bibliographic record identified to a specific author.

In our experiment, we prepared a set of bibliographic records with identical label of author, extracted from the biggest Japanese grants database KAKEN which is characterized by researcher numbers and lists of publications related to the numbers and names. We used the most common 23 Japanese names referring to the 23 clusters with each shared name, which consist of bibliographic records of researchers. Then, we applied tensor decompositions, i.e., CP and RESCAL to each cluster to extract author feature vectors for each bibliographic record in the cluster, and use k -means to distinguish them according to the Euclidian distance of the feature vectors. The author number k is given in this experiment. To compare the performance of tensor decomposition to other models, we adopt LDA (latent dirichlet allocation) that derives a topic distribution for each document in a document set. The document and its topic refer to a bibliographic record and its author feature, respectively. As a result, RESCAL and LDA showed the same performance of clustering such that average f-measures on purity and inverse-purity metrics of RESCAL and LDA are 0.541 and 0.532, respectively. RESCAL was faster than LDA in the experimental data clusters. In addition, we analyzed the relation between cluster size and execution time and recognized that as it becomes larger the preprocessing time to set the data into tensor slices can not be neglected because its computational complexity is $O(n^2)$.

References

- FERREIRA, A. A., GONCALVES, M. A. and LAENDER, A. H. F. (2012): A brief survey of automatic methods for author name disambiguation. *ACM SIGMOD Record*, 41(2), 15. doi:10.1145/2350036.2350040
- KOLDA, T. G. and BADER, B. W. (2009). Tensor Decompositions and Applications. *SIAM Review*, 51(3), 455-500. doi:10.1137/07070111X
- NICKEL, M., TRESP, V. and KRIEGEL, H.-P. (2011). A Three-Way Model for Collective Learning on Multi-Relational Data. *28th International Conference on Machine Learning*. 809-816.

Keywords

AUTHOR NAME DISAMBIGUATION, TENSOR DECOMPOSITION, CP, RESCAL, KAKEN DATABASE

Ensemble methods in classification and clustering
Berthold Lausen
University of Essex

We review methods to use ensembles of selected classifiers to achieve classification rules with increased accuracy (Gul et al. 2016, Khan et al. 2016). Feature selection methods are often used as preprocessing method. For example after preprocessing microarray data with 500 000 probes and 22 125 features (probesets) which represent genes, we use a proposal to improve feature selection of microarray data based on a proportional overlapping score (Mahmoud et al. 2014).

We discuss ensemble methods for cluster analysis. Using ensemble concepts Stoyanov (2015) developed an R package to use hierarchical clustering as preprocessing for k-means clustering. In addition we mention proposals to use nonparametric and parametric bootstrap resampling and distance based variance estimation (Felsenstein 1985; Lausen & Degens, 1988; Degens, Lausen & Vach 1990) to derive a statistical evaluation of clusters.

Degens, P.O., Lausen, B., Vach, W. (1990), Reconstruction of phylogenies by distance data: Mathematical framework and statistical analysis, Lecture notes in biomathematics 84, 9-42.
Felsenstein, J. (1985), Confidence limits on phylogenies: an approach using the bootstrap, *Evolution*, 783- 791

Gul, A., Perperoglou, A., Khan, Z., Mahmoud, O., Miftahuddin, M., Adler, W., Lausen, B. (2016), Ensemble of a subset of kNN classifiers, *Advances in Data Analysis and Classification*, online first

Khan, Z., Gul, A., Mahmoud, O., Miftahuddin, M., Perperoglou, A., Adler, W., Lausen, B. (2016), An ensemble of optimal trees for class membership probability estimation. In: Wilhelm, A., Kestler, H. A. (eds.), *Analysis of Large and Complex Data*, Springer-Verlag Berlin.

Lausen, B., Degens, P.O. (1988), Evaluation of the reconstruction of phylogenies with DNA-DNA hybridization data, in: Bock, H.H. (ed.), *Proceedings First conference of the international federation of classification societies (IFCS)*, North Holland, Amsterdam, 367-374.

Mahmoud, O., Harrison, A.P., Perperoglou, A., Gul, A., Khan, Z., Metodiev, M., Lausen, B. (2014), A feature selection method for classification within functional genomics experiments based on the proportional overlapping score, *BMC Bioinformatics* 15 (1)

Stoyanov, K. (2015), Hierarchical k-means clustering and its application in customer segmentation. Master dissertation, Department of Mathematical Sciences, University of Essex, UK.

Screening for ordinal class structures

Ludwig Lausser*, Lyn-Rouven Schirra, Florian Schmid, Robin Szekely), and Hans A Kestler
Ulm University

Multi-class classifiers systems are typically trained to identify patterns that are associated to the class labels of the objects. Other information such as semantic relationships among the class labels is often neglected. Nevertheless, a feature representation that is assumed to fulfill a semantic relationship should reflect the assumed dependencies. A constraint classifier system that respects the semantic relationship should be able to detect the corresponding patterns.

In this work, we discuss ordinal relationships of type $\text{small} < \text{medium} < \text{large}$ as an example for semantic relationships. Here, a certain ordering of the class labels is assumed. We present the CASCADES algorithm, which is fast enough to screen exhaustively through all permutations of class labels in order to identify candidate orderings. The CASCADES algorithm utilizes two new error bounds on the class-wise sensitivities of ordinal classifier cascades, which can be applied as early stopping criteria. As an example, we show an application out of the field of analyzing gene expression profiles.

Meta-Analysis with Symbolic Data Approach

Masahiro Mizuta

Abstract

Meta-Analysis is a common statistical analysis based on multiple scientific studies. Symbolic Data Analysis (SDA)[1] was proposed 1980's and the field of SDA are extending continuously. We have reported that SDA is good for dealing with Big Data[2].

If we assume that the studies are regarded as concepts, SDA may be powerful tool for Meta-Analysis. In my talk, we will introduce a framework of *symbolic data approach* for Meta-Analysis using interval valued data or distribution valued data.

References

- 1 L. Billard, E. Diday (2006) *Symbolic Data Analysis: Conceptual Statistics and Data Mining*. John Wiley & Sons.
- 2 M. Mizuta (2016). Post Big Data may be Mini Data. *Proceedings of Hokkaido University and Korea University 4th Workshop in Statistics*.
- 3 M. Mizuta(2017). Meta-Analysis and SDA, *Abstract of 6th International Workshop Symbolic Data Analysis 2017*, Ljubljana, Slovenia

Keywords

Mini Data, Distribution Valued Data, Homogeneity and Heterogeneity

Advanced Data Science Lab., Information Initiative Center, Hokkaido University, N11, W6, Kita-ku, Sapporo 060-0811, Japan mizuta@iic.hokudai.ac.jp

Study on Globalization of Japanese Companies Based on Text Mining

Atsuhō Nakayama¹ and Komoda Fumio²

Abstract

The purpose of the present study is to select the optimal unit of document for extracting relationships among each word by text mining on the basis of the purpose of the study. One of the problems in the text mining is to set the unit of the document to be extracted relationships among each word. It can be extracted without losing meaningful relationships among each word if the relationships is extracted from a large unit of document like a whole sentence. But there might arise a problem that many meaningless relationships among each word are extracted as noise. Conversely, the noise can be reduced if we analyze a small units of document delimited by period. However, a problem might arise in that meaningful relationships among each word are missed. Therefore, it is important to select the optimal unit of document for extracting relationships among each word in using the text mining.

We analyzed the magazine article that concerns the internationalization of Japanese firm in the present study. There is no doubt that one of the reasons for the current slump in Japanese companies and the loss of competitiveness lies in the response to internationalization. By applying text mining to journal articles related to internationalization of Japanese companies for 15 years, we will clarify what Japanese companies are taking for internationalization. Then, in order to solve the problem, we examine to set the optimal unit of document for extracting relationships among each word in using the text mining. In the present study, we selected articles of "Nikkei Business" magazine written in Japanese as text mining material. The reason is that the magazine is suitable for knowing the overall trend of each era. The magazine discusses strategies, technology trends, consumer needs, macroeconomic conditions, and so on with regard to problems and tasks that Japanese companies face in each era for all industrial sectors without major bias. In the present study we deal with global strategies of Japanese companies, so we extracted all articles that included the word "国際" (international) or "グローバル" (global) from the articles published from 2001 to 2015 of the same magazine. As a result, 1,408 articles were obtained.

We used morphological analyses such as tokenization, stemming, and part-of-speech tagging to separate the words. Then, we set up the following three units of documents to extract the occurrence frequency of words and the co-occurrence frequency between words, and discuss what kind of each result of text mining will produce and verify in the present study.

- Full sentence unit (article units)

Tokyo Metropolitan University, 1-1 Minami-Ohsawa, Hachioji-shi, 192-0397, Japan atsuhon@tmu.ac.jp ¥ and Saitama Gakuen University, 1510, Kizoro, Kawaguchi-shi, Saitama, 330-0831, Japan techtra@ae.auone-net.jp

- A sentence separated by periods, and five sentences preceding and following the sentence (five sentence units)
- Sentences separated by periods (sentence units)

Since the total of articles containing the word "国際" (international) or "グローバル" (global) is 1,408, the number of article units is 1,408. In addition, the number of sentences separated by periods is 146,866 (the means that the number of periods is 146,866). The number of types of words co-occurrence increases as the unit of document increases, the types of words co-occurrence are the largest in sentence units, followed by five sentence units and article units. The same is true for the total co-occurrence frequency number of words. The important thing is how the co-occurrence relationship changes by changing the units of document such as full sentence units, 5 sentence units, and sentence units.

Three matrix of ratios of co-occurrence frequency number of words was calculated for each of full sentence units, 5 sentence units, and sentence units. These matrices were analyzed by multidimensional scaling. Judging from the difference in each analysis result, the unit of the optimum document used for text mining became clear.

Key words: DATA MINING, GLOBALIZATION, JAPANESE COMPANY, MULTIDIMENSIONAL SCALING, TEXT MINING

Diagnostic Classification of Students' Knowledge States based on Their Item Response Data

Yuan Sun¹, Shunya Inoue², and Yi Sun³

¹ National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, Japan
yuan@nii.ac.jp

² Tokyo Kasei University, 1-18-1 Kaga, Itabashi-ku, Tokyo, Japan
inoues@tokyo-kasei.ac.jp

³ School of Electronic and Communication Engineering, University of China
Academy of Science, China sunyi@ucas.ac.cn

Abstract. As technology has improved over the years we have seen a significant change in learning environments and ways how students learn. The concept of adaptive learning and personalized learning has gained traction among educators. Instead of traditional classroom education, personalized learning promote students initiate their own learning and provide varied resources and learning path for different students based on their knowledge and abilities to help students meet their learning goals. In our studies, we focus on applying for Diagnostic Classification Models (DCMs) (Rupp, Templin and Henson, 2010) to diagnose students' knowledge states and ultimately aim for supporting personalized learning. A key issue in DCMs is to correctly specify the so-called Q-matrix introduced by Tatsuoka (1983), a Boolean matrix which associates the test items and attributes required to solve the items by students which a diagnostic test intends to assess. Currently, a Q-matrix had to be constructed manually by experts in the particular domain, which work requires a lot of time and labor, and also depends on the experts' subjectivity and their personal skills. When the domain or content of the tests is broader, it is an extremely difficult task for experts to specify attributes and the Q-matrix manually. During the past decade, the problem of how to map test items into latent skills based on students' test responses has become a hot topic in psychometrics and in educational data mining. In our previous study, we proposed new algorithms to learn both the Q-matrix and students' knowledge states automatically from the students' ideal item response matrix on the basis of the Boolean Matrix Factorization (BMF) technique (Sun, Ye, Sun and Inoue, 2014). Considering of students' real item response including noises, in this study we apply BFM to solve the deterministic Q-matrix learning problem and propose a recursive algorithm to find an approximate solution, based on the DINA model incorporating guessing and slip parameters (de la Torre, 2008), while solving the uncertainty parameters analytically using maximum likelihood estimation.

References

- Rupp, A., Templin, J. and Henson, R. A. (2010): *Diagnostic measurement: theory, methods, and applications*. Guilford Press.
- Tatsuoka, K. K. (1983): Rule space: an approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, 20(4), 345-354.
- Sun, Y., Ye, S., Inoue, S. and Sun, Y. (2014): Alternating recursive method for q-matrix learning. *Proceeding of the 7th International Conference on Educational Data Mining (EDM 2014)*, 14-20.
- de la Torre, J. (2008): An empirically-based method of Q-matrix validation for the DINA model: Development and applications. *Journal of Educational Measurement*, 45, 343-362.

Keywords

Q-MATRIX, DIAGNOSTIC CLASSIFICATION MODEL, DATA-DRIVEN, ITEM RESPONSE DATA, BOOLEAN MATRIX FACTORIZATION

Customer Segmentation with Qualitative Comparative Analysis

Yuki Toyoda

Abstract

The purpose of this research is to propose a segmentation analysis method that is considering the diversity of customer's brand evaluation. Segmentation focusing on the diversity of individual customers is an important issue for brand management. However, there was a problem that the traditional approach of segmentation using preference regression can not necessarily consider this diversity. For that problem, we use QCA (Qualitative Comparative Analysis). The reason for using QCA is as follows.

- QCA is able to clarify complicated evaluation structure by using logical expression.
- Minimization of the logical expression makes it possible to clarify the core evaluation structure.
- QCA is useful for analyzing a small number of sample data. In this case, the number of data is the number of brands (nine brands) in order to analyze for each customer.

In this study, we use survey data on 9 coffee shops brands (ex. Starbucks, Tully's Coffee etc.) and apply QCA to this data to clarify the evaluation structure of each customer. And then we categorize customers as segmentation by the difference in the evaluation structure.

References

RAGIN, Charles C. (1987): *The Comparative Method. Moving Beyond Qualitative and Quantitative Strategies*. Berkeley, Los Angeles and London: University of California Press.

Hosei Business School of Innovation Management: Shin-Hitokuchizaka bldg. 3-3-9, Kudan-Kita, Chiyoda-ku, Tokyo, 102-0073, Japan. yuki.toyoda.ds@hosei.ac.jp

- RAGIN, C. (2000), *Fuzzy Set Social Science*, The University of Chicago Press, Chicago and London.
- RAGIN, C. (2008), *Redesigning Social Inquiry: Fuzzy Sets and Beyond*, The University of Chicago Press, Chicago and London.
- THIEM, A. and DUSA, A. (2013a), *Qualitative Comparative Analysis with R: A User's Guide*. New York:Springer.
- THIEM, A., and DUSA, A.(2013b), QCA: A Package for Qualitative Comparative Analysis. *The R Journal* 5 (1):87-97.

Keywords

BRAND IMAGE, BRAND PREFERENCE, CUSTOMER SEGMENTATION, QCA(Qualitative Comparative Analysis), CLUSTERING

Omni-Channel Overview -Business Trends and Research Reviews in Marketing Science-

Hiroyuki Tsurumi

Abstract

Our research has the following two objectives. The first objective is to overview the business trends of Omni-Channel. The second objective is to review Omni-Channel research in marketing science and to clarify the research topic.

Omni-Channel is "an integrated sales experience that melds the advantages of physical stores with the information-rich experience of online shopping" (Rigby, 2011). Since the success by US Department Store "Macy's", various retailers began to adopt the concept into retail business. The reason for the success of the Omni-Channel is the progression of "Customer Omni-Channelization" (Okutani, 2016). The customer Omni-Channelization will inevitably promote the retailer Omni-Channelization.

Verhoef (2012) categorized the research on Omni-Channel (including Cross-Channel) into the following three categories.

- (1) Impact of channels on performance
- (2) Shopper behavior across channels
- (3) Retail mix across channels

We also adopted the frame of Verhoef (2012) and reviewed the papers reviewed in Verhoef (2012) and the papers published after Verhoef (2012).

As a result of the review, we clarified research topics. For example, in previous studies, the impact of Omni-Channelization is analyzed based on data. However, the situation varies widely from case to case. After classifying retailers with similar environments, we should measure the impact of Omni-Channelization. In the future, we will rebuild the frame and increase the number of papers to review.

References

- OKUTANI, T. (2016): Omni-Channelization of consumers and decision-making process: The future and problems of retail industry brought by Mobile Device. *Japan marketing journal*, 36(2), 21-43(in Japanese).
- RIGBY, D. (2011): The Future of Shopping. *Harvard Business Review*, 89(12), 65-76.
- VERHOEF, P. C. (2012): Multi-channel Customer Management Strategy. In: V. Shankar, and G. Carpenter (Eds.): *Handbook of Marketing Strategy*. Cheltenham: Edward Elgar Publishing Limited, 135-150.

Key words: MARKETING SCIENCE, RETAIL, OMNI-CHANNEL

Gini index penalized canonical covariance analysis for three-mode three-way data

Jun Tsuchida¹ and Hiroshi Yadohisa²

Abstract

Three-mode three-way data are observed by using the same set of objects and variables in different sources or conditions, with such data often obtained as a set of multivariate matrices. For example, panel data often include values for the same objects and variables at different times. These data are often organized as a three-way array. For three-mode three-way data, researchers have proposed several methods [1][3]. Given two three-mode three-way data sets, we often investigate two types of factors: common factors that show the relationships between the two data sets, and unique factors that represent the uniqueness of each data set. In many cases, the number of variables and conditions is large. Thus, the interpretation of the factors is difficult. This has motivated us to propose a method for investigating common factors and unique factors simultaneously and adopt sparse estimation using the Gini index penalty function. The Gini index is one of the measures of sparsity [2]. Using the Gini index, we could obtain the estimated value of the factors, which is easy to interpret.

References

- [1] Harshman, R. A. (1970). Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA Working Papers in Phonetics*, **16**, pp. 1-84.
- [2] Hurley, N. and Rickard, S. (2009). *Comparing measures of sparsity*, IEEE Trans. on Information Theory, vol. **55**, pp. 4723 - 4741.
- [3] Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, **31**, pp. 279-311.

Keywords

ALTERNATIVE LEAST SQUIRES, MAJORIZING ALGORITHM, SPARSE ESTIMATION,

Graduate School of Culture and Information Science, Doshisha University
jt.tabakosangyo@gmail.com · Department of Culture and Information Science, Doshisha University
hyadohis@mail.doshisha.ac.jp

Lorenz Curves, Gini Index and ... Bacteria

Alfred Ultsch

University of Marburg

Societies of bacteria seem to be able to solve important problems studied in economics. The 'Dilemma of the Common Goods' is an example, which seems to be solved by bacterial films. There is a conjecture that some bacteria colonies implement concurrency for goods (nutrients) in form of a market. The question is: how can one recognize, whether such societies implement a market. The talk will present an answer to this question which refers to agent based swarms and draws on the properties of non symmetrical and heavily skewed distributions and the associated Lorenz curves.

Data Science – Is the impact of Statistics significant?

Claus Weihs

Technical University Dortmund

In this talk we substantiate our premise that statistics is the most important discipline to provide tools and methods to find structure in and to give deeper insight into data. Hence, we will address the impact of statistics on Data Science.