16th Edition of the AgroStat Conference (AgroStat 2022)

Jun 17, 2022 12:00:00 AM
Lyon
France

Table of contents

Application of Deep Learning to phenotypic traits extraction from wheat kernels

Alioune Dieye , Jacques Le Gouis , Bruno Viallis , Fabien Le Couviour , Benoit Jaillais *† [1]

[1] Statistiques Sensométrie Chimiométrie (StatSC) – Statistiques, Sensométrie, Chimiométrie (StatSC), INRAE – Rue de la géraudière, F-44300 Nantes, France

Introduction

Plant phenotyping corresponds to the identification of effects resulting from interactions between genotype and environmental conditions. Among fast and non-destructive technologies, multispectral imaging, combining image and spectrum, is the best technology to measure and quantify a large amount of phenotypic information in a single analysis.

Methodology

A phenotyping robot "Phenotim" with a multispectral LED imaging system as detector was used to increase the high throughput of the robot (about 500 images per 24-hour period). Two UNet Deep Learning models were trained with a random selection of images coming from various lots, and were applied to all multispectral images of cut grain in order to obtain segmented images: one corresponding to the whole grain and the second to the grain without the peripheral layers.

From the segmented images, phenotypic traits were estimated on 80 sample lots, constituted by 20 accessions $\times$ 2 locations $\times$ 2 years. The number of grains per lot was 500 minimum. The traits were the following ones for each grain and for each batch of samples: dimensions of the cut grain as length and width, the depth of the crease, the thickness of the peripheral layers and the vitreousness.

Results

The optimized image processing procedure based on deep learning on the one hand and GPU computing on the other hand allowed to process a collection of 32,000 multispectral images in two hours. The values of each parameter were estimated; their mean, variance and standard deviation were calculated and automatically saved in an Excel file.

Conclusion

The design and realization of the robot has thus solved the bottleneck problem of phenotyping and has enabled the estimation of variability of each lot.

*Speaker

†Corresponding author: benoit.jaillais@inrae.fr

Keywords: Deep Learning, High, Throughput Phenotyping, Multispectral Image, Wheat

Combined-information criterion for clusterwise elastic-net regression. Application to omic data.

Stéphanie Bougeard * [1], Xavier Bry [2], Thomas Verron [3], Ndèye Niang [4]

[1] ANSES - Laboratoire de Ploufragan-Plouzané-Niort – Agence nationale de sécurité sanitaire de lálimentation, de lénvironnement et du travail, Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail – 22440 Ploufragan, France
[2] University of Montpellier – Université de Montpellier – IMAG, Place Eugene Bataillon, 34095 Montpellier, France
[3] Danais – Danais – 29 rue du Faubourg de Bourgogne 45000 Orléeans, France
[4] CNAM – Conservatoire National des Arts et Métiers (CNAM) – CEDRIC, 292 rue St Martin, 75141 Paris Cedex 03, France

Many research questions pertain to a regression problem assuming that the population under study is not homogeneous with respect to the underlying model. In this setting, we propose an original method called Combined Information criterion CLUSterwise elastic-net regression (CICLUS). This method handles several methodological and application-related challenges. It is derived from both the information theory and the microeconomic utility theory and maximizes a well-defined criterion combining three weighted sub-criteria, each being related to a specific aim: getting a parsimonious partition, compact clusters for a better prediction of cluster-membership and a good within-cluster regression fit. The solving algorithm is monotonously convergent under mild assumptions. The CICLUS method provides an innovative solution to two key issues: the automatic optimization of the number of clusters and the issue of a prediction model. We applied it to elastic-net regression in order to be able to manage high-dimensional data involving redundant explanatory variables. CICLUS is illustrated through a real example in the field of omic data, showing how it improves the quality of the prediction and facilitates the interpretation. It should therefore prove useful whenever the data involve a population mixture as for example in biology, social sciences, economics or marketing.

Keywords: Clusterwise regression, Typological regression, Elastic, net regularization

*Speaker

Optimising the Selection of Temperature Levels in the Experimental Design for the Determination of the Cardinal Parameters for Microbial Growth

Ursula Gonzales-Barron * [1], Mariem Ellouze [2], Vasco Cadavez [3]

[1] Instituto Politécnico de Bragança – Campus de Santa Apolonia, Bragança, Portugal
[2] Nestlé Research – Lausanne, Switzerland
[3] Instituto Politécnico de Bragança – Campus de Santa Apolonia, Bragança, Portugal

The cardinal parameter models (CPMs) are a family of predictive microbiology models describing the effect of intrinsic and extrinsic environmental conditions on microbial growth. In the CPM for temperature, the growth rate μ is defined as a function of the optimum microbial growth rate (μopt) and the maximum (Tmax), minimum (Tmin) and optimum (Topt) temperatures for growth. The objective of this study was to gain knowledge on the planning of an optimised experimental design for accurately estimating the cardinal parameters for temperature of a microorganism.
Simulations were conducted in two phases, in order (i) to optimise the position of the temperature levels by contrasting equidistant versus strategic sampling, the latter based on sensitivity analysis; and (ii) to evaluate the accuracy and precision of the parameter estimates for the best sampling type. Simulations were conducted on "known" parameters of Tmin=-1.28 ºC, Topt=37.30 ºC, Tmax=45.12 ºC and μopt=1.289 h-1 (assumed for *Listeria monocytogenes*), and sampling $\sqrt{\mu}$ at a given temperature from a Normal ($\sqrt{\mu}$T, $\sqrt{\mu}$T × CV) distribution, where μT is the theoretical μ from the CPM, and CV the coefficient of variation set between 0.02 to 0.10 representing the expected experimental error in determining μ. In the simulations, the histograms of the fitted parameters were described by mean/SD, and the parameters' accuracy and precision were assessed as a function of the number of temperature levels and error in μ.
Strategic sampling led to better estimability of parameters with less number of temperature levels than equidistant sampling. For two or three replicates, there was no significant gain in the precision (mean/SD) of all parameters beyond 12 temperature levels yet strategically positioned. Fewer levels underestimated Tmin and overestimated Topt and μopt. The error in μ affected to greater extent the parameters' precision (CVp=100×mean/SE) than the number of temperature levels. According to these simulations Tmax, Topt and μopt can be estimated with mean CVp ranging from 0.35% to 1.80%, whereas Tmin is inherently associated with lower precision (CVp> 20%). However, it can still be accurate if low variation in experimental μ is ensured.

Keywords: Predictive microbiology, sensitivity analysis, strategic sampling

*Speaker

COMBINING INCOMPLETE CYCLIC DESIGN WITH COMPOSITE CENTRAL DESIGN TO OPTIMZE PET FOOD PREFERENCES ASSESSMENT

Julien Rogues * [1]

[1] Symrise – Symrise – zone du Gohelis, France

Paired comparisons are widely used for sensory evaluation. In the petfood industry, they are the golden standard to measure diet palatability and to assess the preferences of cats and dogs.

The Bradley-Terry-Luce model is the most common model to analyze paired comparison data (1). However, the simplicity of this model is counterbalanced by the high number of pairs that each subject needs to evaluate to obtain robust results. Several approaches may be used to optimize paired comparisons sensory evaluation without decreasing the statistical relevance of the analysis. This study aims at evaluating the interest of combining an incomplete cyclic design with a composite central design (2) to decrease the number of paired comparisons used to assess dogs' food preferences.

A central composite design with 3 factors was designed following the incomplete cyclic design from Burton (3). 16 dog food prototypes with various doses of fat, liquid palatant and powder palatant were prepared. The products were then tested with 80 paired comparisons selected among the 120 ones available based on the cyclic design (67% of the tests performed). Trials were launched at Panelis, an expert sensory panel specialized pet food palatability (at least 36 dogs per test).

The analyze of the design consists to an extension of Bradley-Terry model including object-specific covariates. The object-specific covariates are key to identify which are the main drivers of palatability in the central composite design created.

The reduction of data by decreasing the number of pairs compared didn't have a huge impact on conclusion. Indeed, results obtained with only 67% of paired comparisons performed were consistent with expected results (figure 1: response surface at 6% of fat showing the results of 2 factors).

Beyond time and cost saving, this approach could allow pet food sensory scientists to decrease the number of tests performed with animals while keeping a high level of reliability.

*Speaker

figure 1: response surface at 6% of fat showing the results of 2 factors

Keywords: data reduction, sustanaible statistics, paired comparisons, bradley Terry, petfood

Multivariate data analysis and clustering of subjects in a Just about right task

Fabien Llobell *† [1], Evelyne Vigneau [2], El Mostafa Qannari [2]

[1] Addinsoft, XLSTAT – XLSTAT – Paris, France
[2] StatSC, ONIRIS, INRA – Institut National de la Recherche Agronomique - INRA – Nantes, France

Whereas univariate analysis of Just about right (JAR) data is widely used through penalty analysis (Iserliyska et al., 2017), multivariate data analysis is much less so, although it can be very interesting to investigate proximities between products and between attributes. Similarly, cluster analysis of the respondents is common in the vast majority of sensory tests but it is not used for JAR data although the products perception may be very different from one subject to another.
Firstly, we present a strategy of coding JAR data that take account of the proximity among the JAR modalities (*i.e.*, 'not enough', 'JAR', 'too much'). Thereafter, we consider a multi-block setting where each block variables is associated to a respondent. Thus, we can apply several strategies pertaining to multi-block data analysis. We focus herein on an adaptation of CATATIS method (Llobell *et al.*, 2019). This yields a global index of agreement of the panel and weights associated to the respondents that reflect the extent to which they agree with the general point of view of the panel. We also discuss an adaptation of CLUSCATA (Llobell *et al.*, 2019) for the cluster analysis of the subjects based on their product perception.

We illustrate the general strategy of analysis based on a case study.

References

Iserliyska, D., Dzhivoderova, M., & Nikovska, K. (2017). Application of penalty analysis to interpret JAR data–A case study on orange juices. Curr. Trends Nat. Sci., 6(11), 6-12.
Llobell, F., Cariou, V., Vigneau, E., Labenne, A., & Qannari, E. M. (2019). A new approach for the analysis of data and the clustering of subjects in a CATA experiment. *FQ&P*, *72*, 31-39.

Keywords: JAR, Multivariate data analysis, Cluster analysis

*Speaker
†Corresponding author: fllobell@xlstat.com

Analyzing Temporal Dominance of Sensations data with categorical functional data techniques

Caroline Peltier * [1], Michel Visalli [1], Pascal Schlich [1], Hervé Cardot

[1] Centre des Sciences du Goût et de l'Alimentation [Dijon], Chemosens Platform – Université de Bourgogne, AgroSup Dijon - Institut National Supérieur des Sciences Agronomiques, de lÁlimentation et de lÉnvironnement : UMR6265, Centre National de la Recherche Scientifique : UMR6265, Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement : UMR1324, PROBE research infrastructure, PF$_{A}cronymfacility, F-Zip_{c}odeCity, France, AgroSupDijon-InstitutNationalSupérieurdesSciencesAgronomiques, delÁlimentationetdelÉnvironnement : UMR6265, AgroSupDijon-InstitutNationalSupérieurdesSciencesAgronomiques, delÁlimentationetdelÉnvironnement : UMR6265, AgroSupDijon-InstitutNationalSupérieurdesSciencesAgronomiques, delÁlimentationetdelÉnvironnement : UMR6265--UniversitédeBourgogne-Bât.CSG-9EBdJeannedÁrc-21000Dijon, France$

In the last twenty years, Temporal Dominance of Sensations (TDS, Pineau et al., 2009) was extensively used for obtaining temporal sensory profiles of food products. Classical analysis of TDS data relies on either dominance rate (proportion of panelists having chosen a given attribute at a given point in time) or dominance duration (elapsed time between two successive attribute choices). None of these two approaches explicitly takes into account the sequence of attribute (probability to go from to another attribute). Although, the use of Semi-Markov processes has been advocated by Lecuelle et al. (2018) to overcome this limitation, this approach turned out to be quite difficult to use in practice.
Recently, Preda et al. (2021) presented an R package including categorical functional data analysis (CFDA) that is particularly relevant for TDS data. This statistical approach, based on the seminal work by Deville (1982), extends the usual functional data analysis to temporal categorical data. The paper illustrates the relevance of CFDA for TDS data by applying it on both simulated and real TDS data (a study on avocados with 70 subjects, 3 products and 12 descriptors).

CFDA produces a PCA-like map of the sensory evaluations (pairs of subject/product) according to the kinetics of their sensations. The evolution of each descriptor can also be projected on each axis, leading to a temporal interpretation of the individual distribution on the map.

The CFDA outputs can also be used as inputs for further statistical analyses (clustering, discriminant analysis, etc.) allowing temporality to be properly taken into account for answering different sensory science problematics, such as segmentation of consumers or product discrimination both based on temporal perception. Examples of these applications will be presented.

*Speaker

Keywords: CFDA, TDS

A touch of psychology to understand consumers: construction of a psychometric scale for the measurement of techno-affinity

Muriel Noel *† [1,2,3], Sébastien Lê [3], Noémie Lucet [2], Yvonnick Noël [1]

[1] Laboratory of psychology, Cognition, Behavior, and Communication – Université Rennes 2 - Haute Bretagne – Pl. Recteur Henri le Moal, 35000 Rennes, France
[2] DEA-IRA - Research division – Renault Group – Technocentre, 1 Av. du Golf, 78280 Guyancourt, France
[3] Statistics and Computer Science Department – L'Institut Agro Agrocampus Ouest – 65, rue de Saint-Brieuc, 35000 Rennes Cedex, France

To understand consumer characteristics that can influence perception, psychology can be useful. In the context of automotive products, the assumption is that the human-technology relationship influences sensory and hedonic perceptions of Human-Machine Interfaces. Nethertheless, few recent questionnaires allow a quantitative measurement of the relationship with modern technology. In addition, some of them refer to multi-dimensional models. However, the relationship to technology seems to show a single dimension characterized by two extremes: technophilia and technophobia (Osiceanu, 2015). Technophilia would be related to enthusiasm and adaptability for modern technologies. In contrast, technophobia would be related to fear and/or discomfort with modern technologies.
A psychometric approach, based on Item Response Theory (IRT), is proposed for the construction of a unidimensional scale allowing the characterisation of individuals according to their affinity for technologies. Item response models have the property of assuming a functional hypothesis between a latent trait (affinity for technologies), and their objective expression (responses to different questions).

A questionnaire that contains factual questions, varying in terms of content and response scales, is developed. Items refer to exhaustive technological objects and different attitudes: frequency and type of use, knowledge, possession, opinions, etc. Analyses are based on the Partial Credit Models (Masters, 1982; Muraki, 1997). These models have the property of being able to project item and subject parameters onto the same dimension. They make it possible, on the one hand, to identify individuals who have a particular attraction for technology from those who reject it and, on the other hand, to identify the most discriminating items.
The final questionnaire was developed in two versions: a long form, for research purposes, and a short form, which can be used as a consumer screener tool. The purpose of this work is therefore to provide a new statistical methodology for understanding and identifying consumer profiles.

*Speaker
†Corresponding author: muriel.noel7@gmail.com

Keywords: Psychometrics, Technophilia, Technophobia, IRT

Relationships between intensity, Just-About-Right and liking variables using PLS-PM approach

Evelyne Vigneau * [1], Martin Paries , Stéphanie Bougeard

[1] StatSC, ONIRIS, INRA – Institut National de la Recherche Agronomique - INRA – Nantes, France

We consider sensory studies in which consumers are asked to indicate their degree of liking for a set of products and, in addition, to provide information on their sensory perception. Sensory perception are collected using both JAR scales and intensity scales. Such data collection are usually explored using penalty analysis in order to highlight the attributes that have the greatest impact of liking. Penalty analysis makes only use of liking scores and JAR attributes, considered one by one. From another point of view, external preference mapping can be applied for understanding the relationship between liking and intensity scales. In this presentation, we will address another perspective with the aim to get an overview of the relationships between the three types of collected information, using a path modelling approach. It is clear that the focus here is not on product characterisation but on understanding the links between the different types of measures.
The investigated path diagram consists, first, to link, in a pairwise way, a latent variable associated to an intensity attribute to a latent variable associated to an JAR attribute, and next each of the JAR latent variables to the overall liking score. The JAR latent variables are obtained by optimal scaling using the PCAOS approach (Paries et al., submitted). The PLSPM model is adjusted using the mode B for the outer model and a path scheme for the inner model.

A case study on " salsas " presented at a data analysis workshop (Popper, 2004) is considered. The significance of the path coefficients is assessed by bootstrapping. A better understanding of the relationships between the different types of measurement is made possible, either from the intensity to JAR evaluations, either from JAR to liking.

Popper (2004). Food Quality and Preference, 15, 891–899
Paries, Bougeard,Vigneau (submitted). Food Quality and Preference

Keywords: sensory intensity, JAR, liking, path analysis

*Speaker

Use of factorial fractional design for consumer test and comparison between consumer mapping techniques (preference mapping and MDS-unfolding) in the context of coffee product development.

Cassandra Mokdad *† [1], Sébastien Périchon [1], Lucile Waksman [2], Nicolas Pineau [3]

[1] Nestlé Product Technology Center Coffee – Société des Produits Nestlé S.A., Nestlé Product Technology Centre Coffee, Rte de Chavornay 3, 1350 Orbe, Switzerland
[2] Nestlé Product Technology Center Dairy – 29 Quality Rd, Nestle RD Center (Pte) Ltd, Singapour 618802, Singapore
[3] Nestlé Research – Société des Produits Nestlé S.A., Nestlé Research, Rte du Jorat 57, 1000 Lausanne, Switzerland

Design of experiments and product consumer tests are two well-known methodologies in the food industry and heavily used in Nestlé Research & Development. Design of experiments are a powerful tool to guide product development with a limited number of trials, and product consumer testing allows capturing consumer drivers of preference and exploring clusters of consumers.
In the frame of the revision of a soluble coffee-based product, this study combines a design of experiment for product screening, followed by a consumer test to better scope the product space of interest and to identify improvement directions. A fractional factorial design of experiments (24-1) helped the team to structure the coverage of the sensory space of interest according to 4 recipe parameters. From the descriptive sensory profiling ran with 14 panelists on 41 attributes, five samples out of eight were selected to reduce the number of samples for the consumer test, while maximizing the sensory space coverage. To conduct the consumer test, the sample selection was completed with two main competitors and the current recipe.

The data from the consumer test were analyzed in two different ways:

- internal preference mapping on the covariance matrix of the consumer liking data, with sensory information and parameters of the design of experiment used as explanatory variables

- MDS-Unfolding approach based on the same data, using the distance to individual maximum liking score as a measure of distance to optimal product, using the same additional information to
Results will discuss the interpretation differences between both mappings, regarding product po-

*Speaker
†Corresponding author: cassandra.mokdad@rd.nestle.com

sitioning, consumer clustering, and ability to explain liking patterns according to the parameters of the design of experiment and the sensory measures.

Keywords: design of experiments, internal preference mapping, MDS unfolding, sensory, consumer liking

Using ASCA to analyze CATA data

El Mostafa Qannari * [1], Benjamin Mahieu , Pascal Schlich

[1] Ecole Nationale Vétérinaire, Agroalimentaire et de lálimentation Nantes-Atlantique – Statistiques, Sensométrie, Chimiométrie (StatSC) – Nantes, France

INTRODUCTION

CATA is a popular method for "rapid" sensory characterization of the products. It instructs panelists to check the attributes, among a list, that apply to the products under evaluation. Multivariate analysis of CATA data usually consists in performing Correspondence Analysis on the contingency table that cross-tabulates the products and the attributes (Meyners *et al.*, 2013).

METHODOLOGY

CATA data from each subject form a binary table with products as rows and attributes as columns indicating the attributes that are checked for each product. These tables are merged vertically to form a dataset that we denote by **X**.

In a first step, we proceed to the standardization of the columns of **X** by dividing each column (*i.e.*, attribute) by the square root of the proportion of *1s* in this column. The second step is to decompose **X** (after standardization) according to the principle of ANOVA-Simultaneous principal components analysis (ASCA). ASCA is based on the decomposition of matrix **X** according to its sources of variations namely, the "main effect", the "respondents' effect" and the "products' effect" in a similar fashion that ANOVA performs for a single attribute. We denote the resulting matrices of effects by **Xr** for respondents and **Xp** for products.

RESULTS

PCA on matrix **Xr** enables us highlighting those respondents that have a tendency to check certain attributes more than the rest of the panel or vice versa. PCA performed on matrix **Xp** makes it possible to depict the configuration of the products and characterize them by means of the attributes. In effect, this method of analysis is tightly linked to MR-CA t (Mahieu et al., 2021). We show that it is superior to CA for the analysis of CATA data.

CONCLUSIONS

We propose a general strategy to analyze CATA data using the ASCA method. In addition to the usual descriptors by products analysis, it yields a descriptors by respondents analysis to highlight respondents having specific patterns of citations. Furthermore, the ASCA decomposition could be useful in presence of repeated evaluations to investigate the interaction effect between respondents and products by adding the interaction effect in the ASCA decomposition.

*Speaker

Keywords: Check, All, That, Apply (CATA), ANOVA, Simultaneous principal components analysis (ASCA)

SensoMap: an interactive tool for sensory analysis develop with R {golem}

Rim Aboulaasri *† [1], Marie-Héloïse Bardel [2], Emmanuelle Mauger‡ [3]

[1] Rim Aboulaasri – ITM STATS : DataScientist Consultant – r.aboulaasri@itm-stats.com, France
[2] Marie-Héloïse Bardel – CHANEL PB : Headof Neuroscience Department – marie-heloise.bardel@chanel.com, France
[3] Emmanuelle Mauger – CHANEL PB : SeniorData Scientist – emmanuelle.mauger@chanel.com, France

SensoMap is an interactive app that aims to map products based on their monadic sensory scores. This tool stems from the need to semi-automate the statistical process behind the design of sensory mappings.
SensoMap was implemented with R and using shiny. The app was modularized by using {golem} package and integrates the conventional methodology of mapping construction with the best practices of shiny application development. The methodology to build mappings relies on several statistical steps. The first step consists in outliers' identification using the standard deviation range method. The second step is to identify sensory attributes that discriminates products by adjusting linear models through ANOVA. The third main step applies a dimensional reduction to adjusted mean sensory data using PCA (Principal Component Analysis). The last step aims to classify products by applying a HCA (Hierarchical Cluster Analysis) on the products factorial coordinates. The app also includes several visualization and data exploration features.

Today, SensoMap is used as an internal tool by CHANEL Neurosciences team to build mappings in a semi-automated and more efficient way. This interactive tool is a time saver and it centralizes sensory mappings.

The use of {golem} facilitates the app's deployment in a package and its dockerization.

Keywords: sensory, mapping, ANOVA (ANalysis Of Variance), PCA (Principal Component Analysis), HCA (Hierarchical Cluster Analysis), shiny, {golem}, app modularization

*Speaker
†Corresponding author: rim.aboulaasri@gmail.com
‡Corresponding author: emmanuelle.mauger@chanel.com

List of participants

- Aboulaasri Rim
- Albert Isabelle
- Accueil Techni'sens
- Ben Debba Camille
- Bertrand Noémie
- Bougeard Stéphanie
- Brandolini-Bunlon Marion
- Bremaud Damien
- Cadavez Vasco
- Calmels Marion
- Chanel Eugénie
- Cheng Yue
- Claeys-Bruno Magalie
- De Walsche Annaïg
- Elisa Palmas
- Gonzales-Barron Ursula
- Govaerts Bernadette
- Grégoire Aymeric
- Guille Adrien
- Jaillais Benoît
- Lê Sébastien
- Le Bé Jean-Vincent
- Llobell Fabien
- Mahieu Benjamin
- Mauger Emmanuelle

- Mokdad Cassandra
- Mousavi Zeinab
- Niang Ndeye
- Noel Muriel
- Pedrosa Mariana
- Peltier Caroline
- Pineau Nicolas
- Qannari El Mostafa
- Ricci Michele
- Rimek Frank
- Rogues Julien
- Schlich Pascal
- Selmi Nour
- Sergent Michelle
- Teillet Eric
- Verron Thomas
- Vigneau Evelyne

Author Index