WORKSHOP: Applying Text Mining Methods for Sensory Evaluation Research

Organizers: Sébastien Lê, Jacob Lahne

Participants: Sébastien Lê, Anne Hasted, Jacob Lahne, Alexiane Luc, Benjamin Mahieu, Leticia Vidal

Facilitator: Jean A McEwan

Many fields associated with "big data" have found great success in analyzing and extracting meaning from unstructured text data. Sensory scientists are increasingly interested in how these methods might be applied to supplement or even in some cases replace traditional methods of data collection. In this workshop, sensory researchers who use text mining in a variety of applications will present 10-minute case studies with an emphasis on interactivity and methods, with a general introduction to the field and an extended intra-panel discussion and audience Q&A. Proposed talks:

Sébastien Lê, *Introduction to text mining for sensory science*.

Jacob Lahne, *Web scraping for sensory research: a case study with cocktails*. Websites that review or describe food products are potentially rich sources of sensory data. However, many sensory scientists find acquiring and parsing web data a challenging barrier. This talk introduces web scraping for sensory data using a case study of a website with ~4500 well-structured cocktail recipes. An interactive workflow for acquiring and parsing this kind of sensory-relevant data is presented (using R), and preliminary results are discussed and demonstrated.

Anne Hasted, *To code or not to code*. A wellbeing study by the European Sensory Network collected text data from over 13,000 respondents in 14 countries. Respondents were asked to write down 4 words related to beauty and feeling good. One of the aims was to investigate cultural differences in articulation of wellbeing. This talk compares results from 5 English-speaking countries obtained using traditional coding methods with those obtained using text analysis software to clean and order the data. We hope to provoke a discussion on the advantages and disadvantages of each approach.

Leticia Vidal, *Automatic text analysis of Twitter data*. Social media is a valuable source to study consumer behaviour. Some years ago, ~50,000 tweets containing the words "breakfast", "lunch", "snack" or "dinner" were retrieved to investigate "what people say when tweeting about eating situations". Content analysis was used, but due to the time-consuming nature of manual coding it was only applied to a subset of 16,000 tweets. In this talk, the use of automatic text analysis tools to gain valuable insights based on the whole dataset will be explored.

Alexiane Luc, *NLP strategies to analyse consumer data using valency: an application to Free JAR data*. This talk presents the analysis of consumer free comment with a constraint of expression, which appears in the systematic use of a JAR structure to describe the tested products. This type of data is also known as Free JAR data. A new algorithm which considers all the specificities of Free JAR consumer data to understand the opinion generated by the tested products among the subjects, will be presented.

Benjamin Mahieu, *Sensory characterization of home-perfumes using Free-Comment as response to open-ended questions*. This talk explores text data from a Free-Comment protocol, in which 90 home-perfume users evaluated 4 home-perfumes. The talk presents data-preprocessing (cleaning and lemmatisation) and addresses the critical issue of grouping descriptors that are “close enough” with an original approach based on a classification using chi-square distance. New statistical tools taking into account dimensionality of the association between words and products will be proposed to enrich Correspondence Analysis.