Contents lists available at SciVerse ScienceDirect

Chemical Engineering Research and Design

journal homepage: www.elsevier.com/locate/cherd

Application of dimensionality reduction to visualisation of high-throughput data and building of a classification model in formulated consumer product design

Nicolai Peremezhney[a], *Colm Connaughton*[a], *Gianfranco Unali*[b], *Evor Hines*[c], *Alexei A. Lapkin*[c,*]

[a] *Complexity Doctoral Training Centre, University of Warwick, CV4 7AL, UK*
[b] *Unilever, Research & Development, Merseyside CH63 3JW, UK*
[c] *School of Engineering, University of Warwick, CV4 7AL, UK*

ABSTRACT

Several dimensionality reduction techniques were applied to two data sets of consumer products formulations in order to infer their intrinsic structure and specific product design rules. High throughput experiments were used to generate the data sets of sufficient size. Supervised isometric feature mapping (S-Isomap) was combined with a k-nearest neighbours (*k*-NN) classifier and k-means clustering algorithm to perform categorization of viscosity of new formulations, not used to train the model. We compared prediction results of this approach with several well-established classification models. The results show the accuracy of the S-Isomap based approach to be superior and with a potential for further improvement. Compared with other dimensionality reduction techniques, applying S-Isomap has allowed for a superior visualization of category separation within the formulations, for the data sets used.

Keywords: Formulated products; System properties, Dimensionality reduction; S-Isomap; Classification

1. Introduction

The aim of chemical product design is to find a product that exhibits certain specified behaviour/properties, corresponding to the desired functional properties. Thus, in the area of formulated consumer products the main useful functions, for example the function of 'moisturising' or the function of 'UV protection', are achieved through use of molecules and particles with corresponding physico-chemical properties, e.g., UV absorbance and scattering of TiO_2 micro-particles reducing the flux of the harmful range of UV solar radiation to the skin. These main useful functions in consumer products are accompanied by a varied cocktail of secondary desired functions, such as 'feel', 'smell', 'colour', etc.

Several important technical challenges in the design of formulated products stem from the fuzziness of performance criteria for a number of desired functions, not easily converted to numerical specifications, and their apparent complexity, which does not allow easy prediction of properties based on composition. The latter means that the performance property is often described as a 'system property': a property that *emerges* due to interactions among individual components of a system, i.e., the ingredients of a formulated product in our case. Viscosity is one example of such system property within the context of formulated consumer products: it is often necessary for a formulated product to have a particular rheological behaviour, e.g., shear thinning behaviour or a Newtonian liquid behaviour, which results in a particular 'feel' function. However, our capacity to accurately predict viscosity of formulated consumer products based on composition is quite limited, due to the physical complexity of the system. Thus, we can describe viscosity as a *system property*, not equal to the linear combination of the contributions of the ingredients used. Viscosity is used in the present study as the output performance criteria for characterisation of the formulations. We should note, however, that in our case viscosity is not a

* Corresponding author. Tel.: +44 (0) 2476151101; fax: +44 (0) 2476418922.
E-mail address: a.lapkin@warwick.ac.uk (A.A. Lapkin).
Received 20 February 2012; Received in revised form 4 May 2012; Accepted 14 May 2012
0263-8762/$ – see front matter
http://dx.doi.org/10.1016/j.cherd.2012.05.010