Exploring the use of Latent Semantic Analysis (LSA) to investigate wine sensory profiles

Poster · September 2024

DOI: 10.13140/RG.2.2.32030.96328

3 authors:

Gonzalo Garrido-Bañuelos
Stellenbosch University
36 PUBLICATIONS **242** CITATIONS
SEE PROFILE

Mpho Mafata
Stellenbosch University
31 PUBLICATIONS **101** CITATIONS
SEE PROFILE

Astrid Buica
AstraZeneca
120 PUBLICATIONS **827** CITATIONS
SEE PROFILE

Exploring the use of Latent Semantic Analysis (LSA) to investigate wine sensory profiles

Gonzalo Garrido-Bañuelos[1,2], Mpho Mafata[1,3] & Astrid Buica[1]

[1] School of Data Science and Computational Thinking – Stellenbosch University, Stellenbosch, South Africa.

[2] Innovate Solutions, Dublin, Ireland.

[3] CREST: Centre for Research Evaluation, Science and Technology, Stellenbosch University, South Africa.

BACKGROUND & AIM

The rise in global temperatures, and the use of frost-resistant grape cultivars (e.g., Solaris), has led to more favourable vine growing conditions in countries like Sweden. However, little is known about the sensory space of Solaris wines, the potential different wine styles.

Our work **explored the use of Latent Semantic Analysis (LSA) as a tool to investigate the sensory space of different products using text data.** LSA generates different ***Topics.*** These topics represent the term-to-term associations which could potentially represent different dimensions. The outcomes of LSA were compared to conventional data strategies used to analysed text data such as Multiple Correspondence Analysis (MCA).

The study focused on Swedish Solaris (SOL) wines and their fit with international Albariño (AL) and Chenin Blanc (ChB) wines.

EXPERIMENTAL DESIGN & DATA STRATEGY

I

Relationship between LSA topics

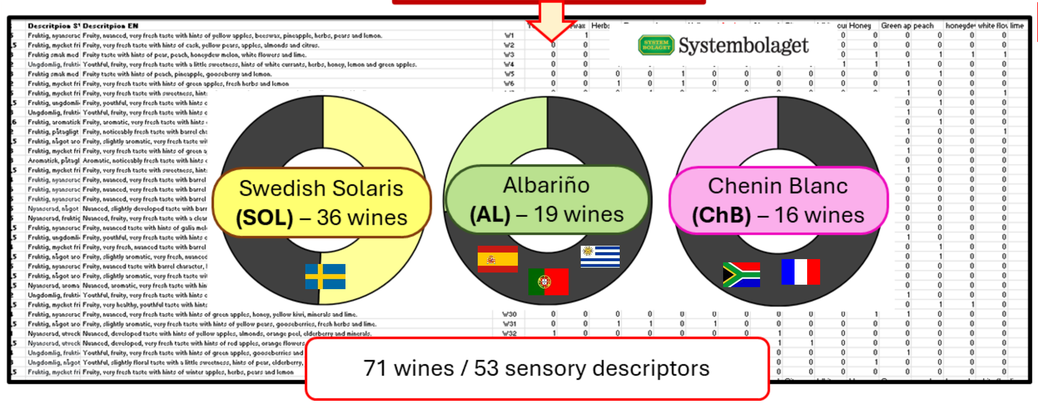
Latent Semantic Analysis (LSA)

Multidimensional Scaling (MDS)

Agglomerative Hierarchical Clustering (AHC)

Data collection

Nuanced, fruity, very fresh taste with clear barrel character, hints of red apples, browned butter, mineral, spices, nuts and lemon. Served at 8-10°C to tasty dishes of fish, for light meat, or for lighter dishes of lamb.



71 wines / 53 sensory descriptors

II

Multiple Correspondence Analysis (MCA)

AHC

RESULTS

I

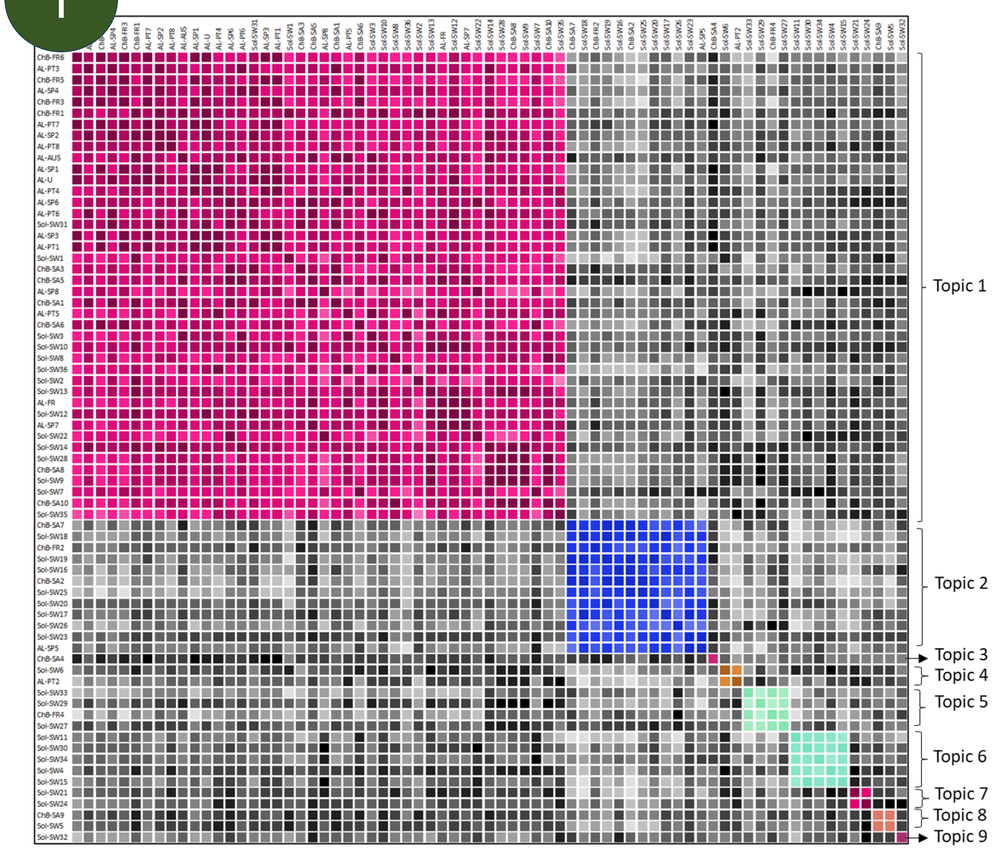


Figure 1. Wine-to-wine correlation matrix from LSA

Table 1. LSA wine topics and associated terms

Topics	Number of documents ('Wines')	Number of terms ('Sensory attributes')	Variability (%)	Terms ('Sensory attributes')
Topic 1	42	12	27.4	Pear, Herbs, Peach, Mineral, Lime, Yellow pear, Apple, Beeswax, Gooseberry, White currant, Smoky mineral, Kiwi
Topic 2	12	14	10.4	Oak, Nut, Orange, Yellow plum, Apricot, Butter, Vanilla, Spices, Red apples, Cardamom, Marzipan, Dried pineapple, Hazelnut, Watermelon
Topic 3	1	5	6.6	Lemon, Pineapple, Grilled lemon, Macadamia nuts, Nectarine
Topic 4	2	2	5.5	Galia melon, White peach
Topic 5	4	5	4.9	Citrus, Dried apricot, Passion fruit, Apple blossom, Blackcurrant leaves
Topic 6	5	3	4.55	Green apple, White flowers, Yellow kiwi
Topic 7	2	4	4.0	Honey, Elderberry, Mango, Bergamot
Topic 8	2	3	3.9	Yellow apples, Grapefruit, Almond flowers
Topic 9	1	5	3.3	Honeydew melon, Almond, Rosehips, Orange flower, Orange peel

LSA creates a document-to-document (*i.e.*, wine-to-wine) correlation matrix (Figure 1) based on the degree of similarity from the different term-to-term associations – **Topics** (Table 1).

- Most wines (**54** out of 71) from all three cultivars were associated to the first two topics.
- SOL wines (**36**) were associated to multiple Topics
  - Low typicality
  - Potential for multiple wine styles
- Opposite example, most AL wines (18 out of 19) were associated to Topic 1.
  - 'Pear' and 'lime' in Topic 1 were found in 89% and 68% of AL wines, respectively.
- Relationship between Topics was investigated using MDS and AHC – 2 clusters were found.
  - Cluster 1: Topic 1 and Topic 3
  - Cluster 2: Topic 2 and all the other Topics (4-9)

II

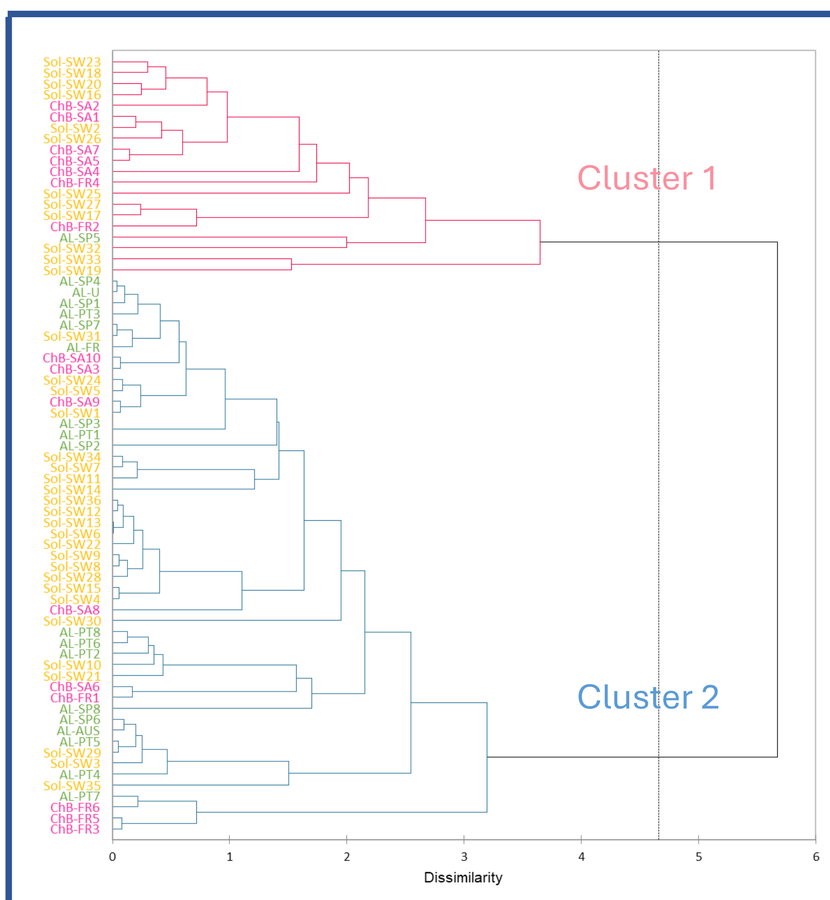


Figure 2. Dendrogram on MCA dimensions

- All wines associated to Topic 2 are in Cluster 1.
- SOL (yellow) ad ChB (pink) wines were found in both Cluster 1 and 2 .
- Most AL wines (green) are in Cluster 2.

CONCLUSION

- Similar results obtained with LSA and MCA
  - MCA balances the relationship between samples and attributes.
  - LSA emphasizes the relationship between attributes (Topics).
- Study shows no evidence of a Solaris typicality based on current data
- Solaris wines shared sensory features with both Albariño and Chenin Blanc.

SCHOOL FOR DATA SCIENCE & COMPUTATIONAL THINKING

INNOVATE SOLUTIONS

crest Centre for Research on Evaluation, Science and Technology