An NLP-Based Scotch Whisky Recommender Agent

Robert Soane

June 14, 2021

The Scotch Market

The Scotch whisky industry is massive.

- 75% of Scotland's food and drink exports.
- Over 20% of the UK's food and drink exports.[1]
- Over 130 disilleries in Scotland.[2]

[1] https://www.scotch-whisky.org.uk/insights/facts-figures/

[2] https://whiskytastingcompany.com/blogs/news/how-many-whisky-distilleries-are-in-scotland

Distillation

The Law

Legal definition of Scotch:[3]

- Aged in oak casks
- Minimum age $= 3$ years
- ABV $\geq 40\%$
- Entire production process in Scotland

[3] legislation.gov.uk/uksi/2009/2890/regulation/3/made

Whisky and Words

Various whisky-specific words are used.[4] [5]

Peat

Smoky flavour imparted on Scotch from peat fires used dry malt/grain.

Sherry

Descriptor for flavours imparted on Scotch from aging in casks previously used to age sherry.

---

[4] G. N. Bathgate, "The influence of malt and wort processing on spirit character: the lost styles of Scotch malt whisky," *Journal of the Institute of Brewing*, vol. 125, no. 2, pp. 200-213, 2019

[5] J. Mosedale and J.-L. Puech, "Wood maturation of distilled beverages", *Trends in Food Science & Technology*, vol. 9, no. 3, pp. 95-101, mar 1998. Available: https://linkinghub.elsevier.com/retrieve/pii/S0924224498000247

Statement of Problem

This project sought to apply NLP techniques to produce a recommender agent and ascertain whether NLP techniques applied to Whisky tasting notes can power an effective recommender agent.

Mapping Text to a Vector Space

- Need a model to transform text.
- Common models transform the text to a vector space.[6]
- Root form transformations - stemming, lemmatization etc.
- Bag of Words & Word2vec

[6] https://ieeexplore.ieee.org/document/6786458

Keyword Extraction

- To build BoW need keywords
- Various methods can be used
- TF-IDF
- Graph Based Methods[7]

[7] https://hrcak.srce.hr/140857

Graph Based KE

- Co-occurence graph
- RAKE[8], Eigencentrality[9]

Eigencentrality

Each node's centrality proportional to all adjacent nodes.

- Adjacency matrix A
- i^{th} node's score x_i
- $x_i = \frac{1}{\lambda} A_{ij} x_j$
- $\mathbf{A} \cdot \underset{\sim}{x} = \lambda \underset{\sim}{x}$

---

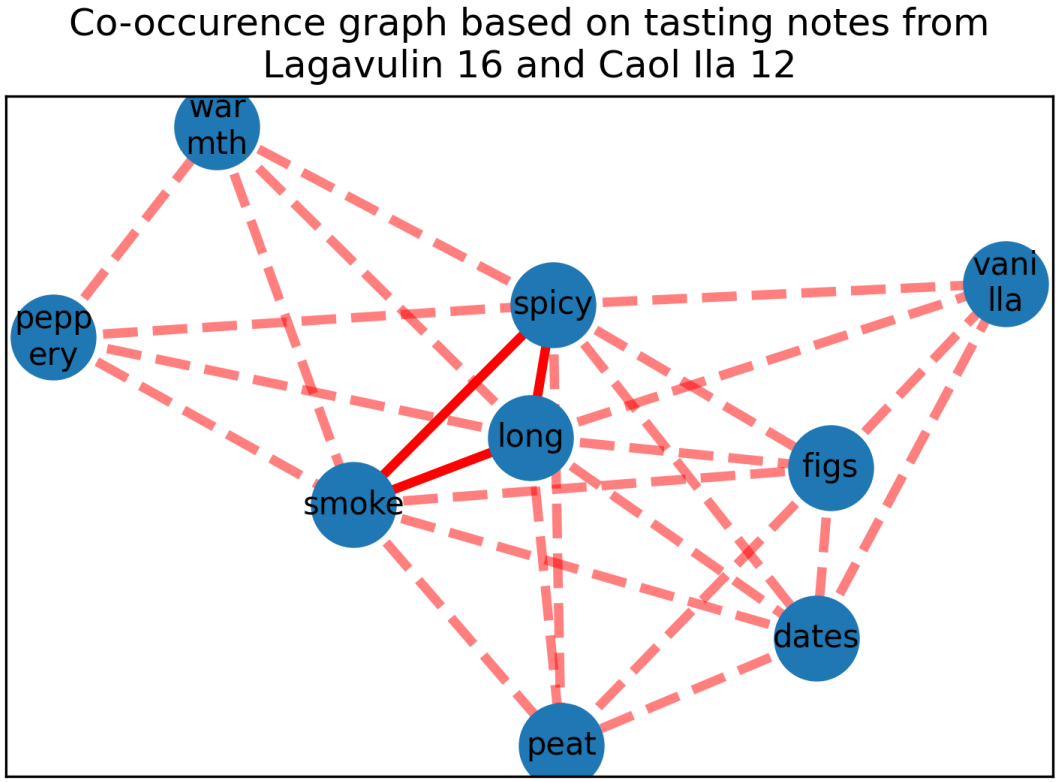
[8] S. Rose, D. Engel, N. Cramer, and W. Cowley, "Automatic keyword extraction," *Text Mining: Applications and Theory*, pp. 1-277, 2010

[9] https://linkinghub.elsevier.com/retrieve/pii/S0378873307000342

Co-Occurence graph

Sample tasting notes:[10]

- "*Long, peppery, spicy warmth, smoke.*"
- "*Long, spicy, figs, dates, peat smoke, vanilla.*"



[10] Adapted from masterofmalt.com's listings for Caol Ila 12 and Lagavulin 16

Graph Based KE

- Co-occurence graph
- RAKE[11], Eigencentrality[12]

Eigencentrality

Each node's centrality proportional to all adjacent nodes.

- Adjacency Graph A
- i^{th} node's score x_i
- $x_i = \frac{1}{\lambda} A_{ij} x_j$
- $\mathbf{A} \cdot \underset{\sim}{x} = \lambda \underset{\sim}{x}$

[11] S. Rose, D. Engel, N. Cramer, and W. Cowley, "Automatic keyword extraction," *Text Mining: Applications and Theory*, pp. 1-277, 2010

[12] https://linkinghub.elsevier.com/retrieve/pii/S0378873307000342

Recommender Agents

- **Collaborative Filtering (CF):** Based on users habits.
- **Content Based (CB):** Based on details about instances in data set.

AI Applications to Whisky

Large gap in the research

- Small number of CF and CB agents have been produced.
- Predominantly based on distinct features, or the users entire profile.

Classification of Single Malt Whiskies

One piece of research into clustering of Scotch using tasting note data:[13]

- Used tasting notes to cluster 80 whiskies
- Very much aimed at finding specific details of flavour dimensions
- Interesting research, working with industry

This project aims instead to produce an autonomous agent to recommend whiskies.

[13] https://link.springer.com/chapter/10.1007/978-3-642-59789-3_14

The Agent and the Environment

Defining the System

- **The Agent:** Recommends based on user input and tasting notes. Capacity to autonomously update database.
- **The Environment:** The user input, and an online spirits shop.

Web Scraping

The dataset was scraped from the Master of Malt website using a python script.[14]
The agent update function scrapes the new additions to the website. IDs are created using an MD5 hash - the update function stops updating after it achieves a number of duplicates.
The model is re-trained after each update.

[14] http://masterofmalt.com/

KE and Lemmatizing

- The Scotch lexicon leads to poor lemmatization attempts
- WordNet : 'peated' $\rightarrow$ 'peated'
- Whisky tasting notes are very feature rich
- Default lemmatizer and KE methods performed poorly
- Implemented an Eigencentrality KE method in Python

KE and Lemmatizing

- The Scotch lexicon leads to poor lemmatization attempts
- WordNet : 'peated' $\rightarrow$ 'peated'
- Whisky tasting notes are very feature rich
- Default lemmatizer and KE methods performed poorly
- Implemented an Eigencentrality KE method in Python

Table: Times of TF-IDF, RAKE and eRAKE with various lemmatizers in seconds.

	TF-IDF	RAKE	eRAKE
U	*1247*	*0.441*	-
WN	*1461*	*130.2*	-
WL	*1209*	*4.947*	*47.6*

Note: eRAKE was only applied to the WhiskyLemmatized corpus as the eRake implementation included the WhiskyLemmatizer.
u: Unlemmatized, *WN*: WordNet, *WL*: Whisky

Clustering

Implemented clustering for sanity check, and qualitative comparison of different KE methods.

- Found that RAKE performed poorly - unsurprising as RAKE is aimed at features with higher relative frequency. Unsuited to tasting notes.
- eRAKE and TF-IDF performed similarly.

Training Process

- eRAKE method lemmatizes tasting notes and performs KE
- Each whisky vectorised
- Dataset transformed to a matrix
- Model created for each of *nose*, *palate*, *finish* and *general*

The Ideal Vector (IV)

- Models queried for each whisky's vectors
- Vectors amalgamated to produce an Ideal Vector (IV)
- Cosine similariy used to rank whiskies based on similarity to IV

Cosine Similarity

- Aim to find cosine of angles between each vector.
- For $\underset{\sim}{u}, \underset{\sim}{v} \in \mathbb{R}^k$, $\underset{\sim}{u} \cdot \underset{\sim}{v} := |\underset{\sim}{u}||\underset{\sim}{v}| \cos\theta$
- All vectors are stored normalised.

The Matrix Equation

$$\begin{pmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ \dots & \dots & \ddots & \dots \\ d_{m1} & d_{m2} & \dots & d_{mn} \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ \dots \\ v_n \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ \dots \\ c_n \end{pmatrix} \tag{1}$$

The Survey

Input	Rationale
ATN1	Replicating a user who has tried and developed tastes for a variety of whiskies available at supermarkets, but hasn't tried much beyond.
ATN2	Replicating a significant partiality towards heavily peated whiskies.
ATN3	Replicating an enjoyment of both peated and sherried whiskies.
ATN4	A user with niche and specific whisky tastes.
GC	Producing recommendations based on general inputs without considering specific tasting notes.
DD1	Dream Dram recommendation from a very peat heavy input.
DD2	Dream Dram recommendation describing a very oily and fruity whisky.

User Ratings

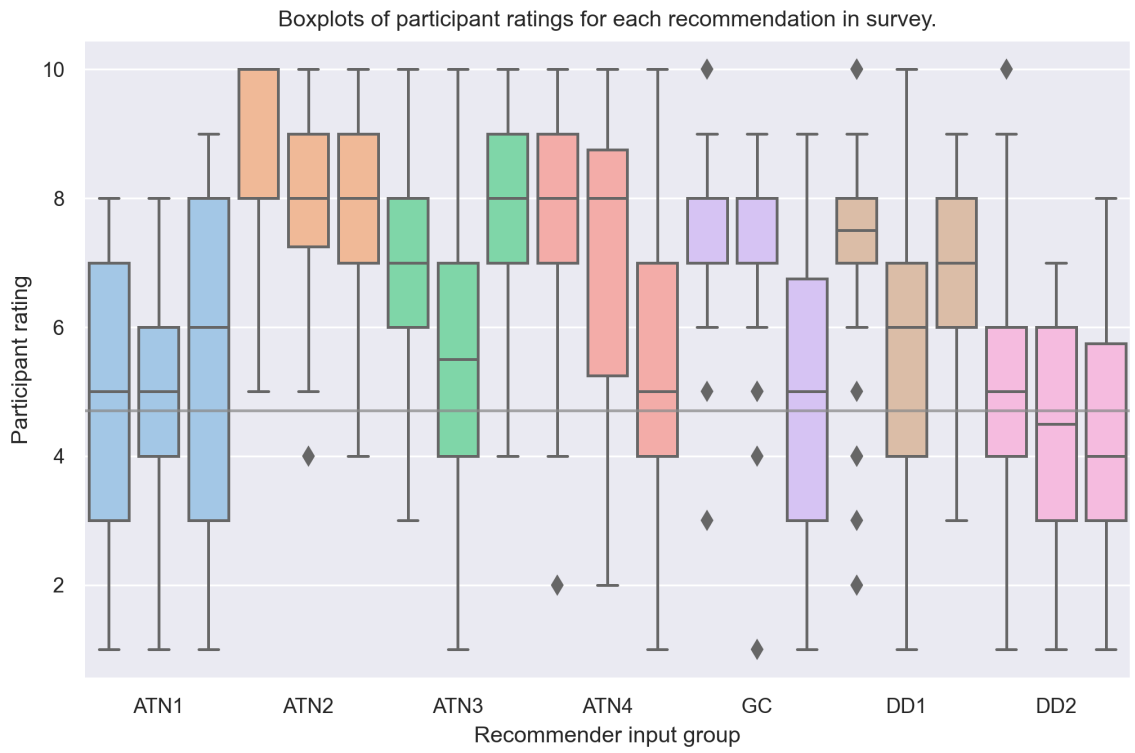


Figure: Boxplots of participant ratings for each recommendation, grouped by sample input. The grey line indicates the mean baseline rating.

T-Test

T-Test Results

- Performed a one-tailed paired t-test on the baseline scores and corresponding recommender scores for each sample.
- Model $M = 6.28$, $SD = 1.13$
- Baseline $M = 4.71$, $SD = 1.15$
- Significant increase in scores at the 5% significance level - $t(12) = 2.22$, $p < 0.05$
- There was a mean increase of 1.57 points.

Conclusions

- Whisky recommender found to work
- More comprehensive evaluation needed

Suggestions for Future Work

- Front End UI
- Further comparisons of KE methods
- Work on whisky clustering
- Work with industry experts to more comprehensively evaluate and improve the agent

Questions

Any Questions?