

An NLP-Based Scotch Whisky Recommender Agent

Robert Soane

14302983

June 13, 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 distilleries 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 approximate whisky production process:³

- Grains germinate to develop sugars
- Sugars extracted to produce *wort*
- Wort fermented
- Beer distilled
- *New make* aged

The Law

Legal definition of Scotch:^a

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

^alegislation.gov.uk/ukxi/2009/2890/regulation/3/made

³K. Jacques, T. Lyons, and D. Kelsall, *The Alcohol Textbook* 4th edition, 4th ed. Nottingham University Press, 2003.

Single, Malt, Blends and Grain

- 'Single' \implies One distillery.
- 'Malt' \implies 100% malted barley
- 'Blended' \implies Multiple distilleries
- 'Grain' \implies Any grain, not just malt.

Flavour Dimensions

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.

⁴G. 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

⁵J. 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 Methods⁷

⁷<https://hrcak.srce.hr/140857>

Graph Based KE

- Co-occurrence graph
- Each node represents a candidate keyword
- Centrality measures are used to select keywords
- RAKE⁸, Eigencentrality⁹

Eigencentrality

Each node's centrality proportional to all adjacent nodes.

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

⁸S. 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>

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

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:¹⁰

- 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.

¹⁰https://link.springer.com/chapter/10.1007/978-3-642-59789-3_14

The Agent and the Environment

- **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.

The dataset was scraped from the Master of Malt website using a python script.¹¹

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.

¹¹<http://masterofmalt.com/>

- The Scotch lexicon leads to poor lemmatization attempts
- WordNet : 'peated' → '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

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. Unsuitable 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 similarity used to rank whiskies based on similarity to IV

Cosine Similarity

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

The Matrix Equation

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

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.

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.

User Ratings

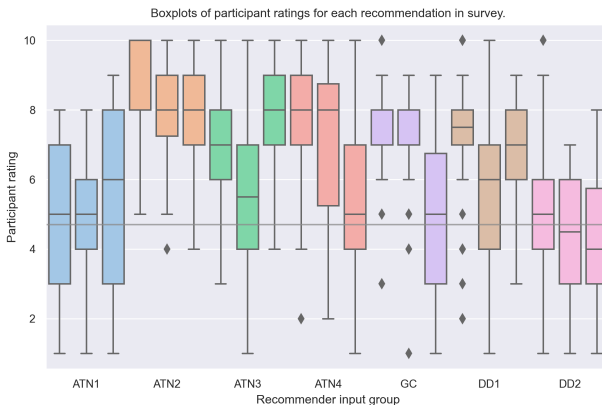


Figure: Boxplots of participant ratings for each recommendation, grouped by sample input. The grey line indicates the mean baseline rating.

- 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?