WINE: A CRY FOR ATTENTION

Chris Burgess
University of North Carolina at Chapel Hill
chrisbur@cs.unc.edu

Justin Do
University of North Carolina at Chapel Hill
justindo@live.unc.edu

Kyuyeon Kim
University of North Carolina at Chapel Hill
kyuyeon@live.unc.edu

Rhea Gupta
University of North Carolina at Chapel Hill
rheaa@live.unc.edu

November 22, 2020

ABSTRACT

The transformer model is a powerful tool in many language processing tasks. We apply the same principals of translation to wine variety predictions. We take as input to the encoder the description of the wine and "translate" it to the variety of wine being described. Our model performs well above chance and predicts the wine variety correct with accuracy close to 54%. We believe that using attention mechanisms and treating the categorization as a form of translation to be a valid method of solving language description to label problems.

1 Introduction

Our project is a competition on Kaggle (Wine Reviews). We are provided with information gathered from wine reviews. The features we are focusing on are the textual description of the wine from the taster and the variety of wine (e.g. Pinot Noir, Sauvignon Blanc, etc.). We present a model capable of predicting the variety of wine when given the textual description.

We believe this same model could be used to predict other features such as location and winery. We choose variety because the description of wine focuses most heavily on the sensations of drinking the wine which is most closely linked to the variety.

In this work we propose using a Transformer to solve the task of predicting variety given a description. The reasoning behind this is that one can conceptualize guessing the kind of wine as a translation from description that was given. Given that the input is natural language a multi-headed self attention network has been shown to perform well. [1]

The inspiration for this abstract use of the transformer architecture comes from the use of attention mechanism in solving object segmentation tasks [3]. In their work they showed that transformers are a natural way to solve the problems inherent in older methods of segmentation such as label duplication and occlusion.

2 Model choice and background

We present what we believe to be a novel approach to solve the problem of predicting a variety of wine given on the description. The intuition for choosing the transformer as our model comes from the way a person would process the textual description and try to make a guess. They would attend to words in the description that would lead us to believe the taster was describing a certain kind of wine. Confidence to use a transformer is such an abstract way comes from successful research in object detection frameworks [3].

Below we have an example that exemplifies how intuitive it is to use a language model meant for translating text to predict wine varieties. As you read the passage you will notice how you begin to relate the description to particular wines you may have had or other reviews/descriptions you have heard of wine. Our model won't have tasted wine before but once its done training it will hopefully learn some associations.

What kind of wine is this?

> Ripe and juicy, this perfumed wine has a cool character to go with its black-cherry flavors. Partial wood aging has added spice and a smooth texture to the wine's complexity.
>
> - *Roger Voss*

The wine was described as "ripe", "juicy", "perfumed", "smooth texture". Someone familiar with wine might associate these things with a classic red wine such as a Pinot Noir (which this wine happens to be). Our goal is to take a description like this and learn a dimensionality reduction that is in essence a translation from a description to a variety of wine. The Transformer was chosen as our base model because we want it to learn that words in the description can directly be related to kinds of wine. Self attention should be able to do just that. One hope is that each head should learn to attend to certain kinds of descriptions in the wine and better be able to understand what they mean in terms of the varieties they come from.

We chose to use a transformer library based on the "Attention is all you need" paper [1, 4]. This library is publicly available on GitHub and is meant to be used with the Keras API on a backend of your choosing.

3 Data preparation

The Kaggle challenge focuses on working with the descriptions and the variety. This became apparent once we started working with the data. Many of the features have NULL values or are simply missing. When we attempted to work with other features we had to throw out large portions of our data. As can be seen below our data consisted of many Pinot Noirs. This caused problems in our initial testing where we seemed to be over fitting and the model learned to always predict Pinot Noir. We solved this by increasing the dropout rate to 0.1 while training.

variety	% of dataset
other	45.1
Pinot Noir	10.2
Chardonnay	9.0
Cabernet Sauvignon	7.3
Red Blend	6.9
Bordeaux-style Red Blend	5.3
Riesling	4.0
Sauvignon Blanc	3.8
Syrah	3.2
Rosé	2.7
Merlot	2.4

The library for implementing the transformers was straightforward in how it wanted the data presented. We needed to do the following to each feature:

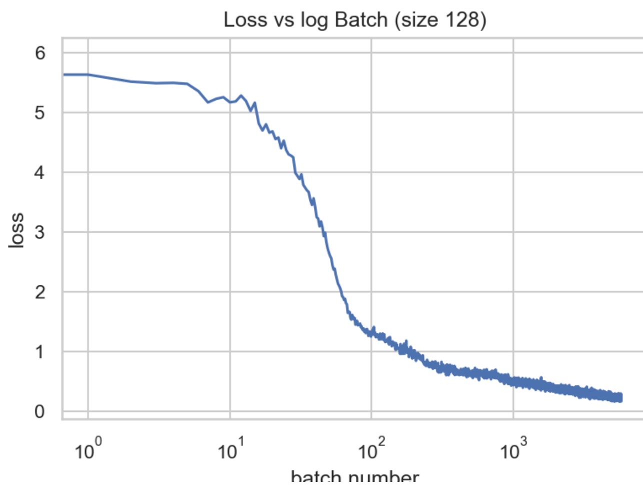
1. Build a list of all tokens in our source and targets. The source is a list of every word that occurs in every description. The target is the variety of wine.
2. Build a token dictionary including the <START>, <END>, and <PAD> tokens. This allows us to map a given word to an index in the dictionary.
3. Map each description to their corresponding token and surround with the start and end token. Add padding where necessary (each input and output needs to be able to contain the maximum length).
4. Separate the features and labels into test and train data sets. We choose to use a test size of 0.05.

4 Model and parameters

We choose to follow the general configuration of the normal size model from the original transformer paper [1]. We chose an embedding dimension for our Q, K, and V [1] vectors of 64. The architecture we had the best results with was four stacked encoders with four stacked decoders. Eight attention heads were chose. More were experimented with but the increase in complexity caused our training to fail. The feed forward network consisted of 256 fully connected neurons. A dropout rate of 0.1. Our batch size for training was 128.

We experimented with only using 2 decode layers as well and saw only small changes in performance. If we stacked too many layers the model learns that it should only predict "Pinot Noir" because it is the most probable with it making up a large portion of the data. We were unable to determine if this was due to under or over fitting. In our initial testing we had to limit the dataset as well as train for only short time.

Training time had a large impact on performance (as expected) and we saw a jump from 23% accuracy to 50% from 1 to 3 epochs and finally 54.95% with 6 epochs. This is not a lot of time spent training but we are working with limited hardware. The limited hardware also means we were unable to make our models any larger. Currently we are working with about 10 million trainable parameters. Anything larger than this and the GPU we are using (a 1070) runs out of memory during the training process.



5 Results

With our final trained model we were able to achieve an accuracy of 54.82%. This is well above chance when predicting the wine variety. The model could perform better if there was more data. When it comes to language translation training data sets only having 130,000 examples is not ideal. This result also impressed us because of the high amount of dimensionality reduction involved in going from a large textual description down to a single variety.

After the references is an example of some of our test data and the predictions the model made. You will notice that some of the descriptions include the kind of wine that was being tasted. This is not explicitly taken advantage of in the design of the model so we believe it to be fair game. There is also an adversarial example that mentions a "Cabernet" when in actuality the wine is a "Red Blend".

A fun side effect of using a language model to predict the wine variety is that we could enable it to come up with new wine varieties if it is forced to not choose the most probable symbols. When experimenting with choosing a top_k value of greater than 1 the model would output made up varieties of wine that were a combination of different wines. Some of them seemed quite tasty. We would love to try a "Cabernet Sauvignon-Carmenère" or "Bois Blend Red Riesling Grosso Nebbiolo Grosso".

[1] Descriptions of the Query, Key, and Value vectors can be found in the original transformer paper [1]. They are essentially how the self attention layer tracks which parts of the sentence are important. Combining the Query + Key from the encoder with the Value from the decoder is how the translation takes place.

Accuracy We evaluated the performance of our model by having it predict values from our test dataset that the model did not have a chance to train on. We consider the model correct if it prints exactly the output that we expect from the category. This is a challenge because we have trained a language model that is capable of writing sentences (where the vocabulary is just wine types). These accuracy numbers are from trying to predict 5000 wine descriptions. Having the model print more than one guess and comparing if it predicted any words correctly yielded much better results but we considered this to be unrealistic and did not use that metric.

Model	Accuracy	# params (M)
random guess	0.0014	
most probable guess	0.1137	
4 encoder 2 decoder	0.4739	10.4
4 encoder 4 decoder	**0.5482**	10.5

References

[1] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *arXiv preprint arXiv:1706.03762v5*, 2017.

[2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805v2*, 2019.

[3] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. *arXiv preprint arXiv:2005.12872v3*, 2020.

[4] Keras Transformer Library https://github.com/CyberZHG/keras-transformer

Description	Actual Variety	Predicted Variety
Made with Sangiovese and a touch of Colorino and Canaiolo, this easygoing wine offers pretty aromas of violet and red cherry. The soft, enjoyable palate doles out crushed raspberry, clove and a t...	Red Blend	Red Blend
Campfire, animal, horseradish and berry aromas make for a slightly difficult nose. This Malbec feels pinched, with tannic pull. Salty black-fruit flavors finish similar, with popping acidity that...	Malbec	Malbec
Scorched earth, toasted nut, espresso, cured meat and baked plum aromas lead the nose on this concentrated wine along with a hint of new leather. The big, chewy palate doles out mature black cher...	Sangiovese	Sangiovese
Here's a soft, fruity and sweet Cabernet, rich and spicy in raspberry, cherry and blackberry jam. Lots of new oak adds smoky vanilla and caramel notes. It's delicious, but pretty direct, lacking ...	Red Blend	Cabernet Sauvignon
Aromas of coconut, espresso and spiced plum steeped in spirits lead the nose on this full-bodied red. The bold palate delivers oak extract, dried black cherry, sweet vanilla and roasted coffee be...	Aglianico	Red Blend
Bright, generous and straightforward, Soave Classico San Michele makes a perfect pairing partner to white asparagus or spring peas. The wine delivers pretty aromas of honey, peach and grapefruit....	Garganega	Garganega
Deep violet red to the eye, this wine offers fragrances of raspberry and vanilla. It has a silky mouthfeel as well as a pleasantly easy-drinking quality, with flavors of cherry, Dr Pepper, pomegr...	Shiraz	Rhône-style Red Blend
Apple and citrus notes signal freshness, with a seam of bright acidity pervading the palate, giving a tart aspect to the wine. The fizz is soft and the body is light, making for a simple, refresh...	Sparkling Blend	Pinot Gris
This ripe, smoothly textured wine is rich with apricot and white-peach fruits. Light wood flavors add spice and fill out the texture. With its generous weight, the wine is going to be full and ri...	Chardonnay	Chardonnay
Bold, ripe fruit and chocolate notes open the nose of this blend of Merlot, Cabernet Franc and Petit Verdot. The wine is layered and rich, with lingering tones of tobacco and sweet spice. ...	Red Blend	Bordeaux-style Red Blend
This funky rosé is made from Cabernet Sauvignon, Merlot and Syrah, and there's color, spice aromas and weight dictated by that grape mix. Heavy plum and cherry flavors are fruity but flat in feel...	Rosé	Rosé
Lightly herbal, spicy aromas of tobacco, forest floor, juniper and pepper accent core blackberry, cassis and cherry aromas. This Cabernet is tight and mildly tannic, but not fierce or demanding. ...	Cabernet Sauvignon	Cabernet Sauvignon
Austere and refined, this elegant sparkler opens with aromas of toasted bread crust, walnut and a hint of oak. Made entirely with Pinot Noir, the linear palate delivers yellow apple, lemon zest a...	Pinot Nero	Sparkling Blend
This rusty-colored Merlot is burnt smelling, with oregano and tomato-driven aromas. The palate feels dried out and raw, while flavors of burnt grass, roasted plum and tomato finish with an oxidiz...	Merlot	Merlot
This wine combines oaky sweetness with good, plummy fruit flavors and a lightly grippy texture. It's tasty and approachable in style. ...	Cabernet Sauvignon	Pinot Noir