SOMM: INTO THE MODEL — AI FOR WINE PROFESSIONALS AND ENTHUSIASTS

DOTTIE HU

DOT@AI-SOMM.COM

Find more interactive visualizations, demos, other topics, technical details, and more at https://ai-somm.com

About the Author

Dottie Hu is an AI research scientist in New York City. Her research experience and interests lie in intersdisciplinary research bridging social sciences, computational linguistics, computer vision, and speech. She has published in top conferences and journals in natural language processing, computer vision, speech, and applied statistics including Association of Computational Linguistics (ACL), Empirical Methods in Natural Language Processing (EMNLP), Computer Vision and Pattern Recognition (CVPR), European Conference on Computer Vision (ECCV), International Conference on Computer Vision (ICCV), InterSpeech, and Annals of Applied Statistics (AoAS). She received her PhD from Cornell University in 2019.

She is also a wine professional with credentials including Certified Sommelier by The Court of Master Sommelier, Diploma in Wine with Merit by Wine & Spirits Education Trust, Certified Specialist of Wine by The Society of Wine Educators, and Certified Specialist of Spirits by The Society of Wine Educators. She is working towards the Master of Wine diploma by the Institude of Master of Wine.

0

SECTION

Introduction

Everything about wine appears intricate and complex, and mastering wine appears an ever daunting endeavor with the fast changing landscape of the worldwide wine industry, encompassing a wide range of subjects such as geology, geography, viticulture, viniculture, chemistry, law, marketing, operations, and more, in an almost multi-lingual and largely unstructured form.

From the meticulous handing by experienced sommerliers of delicate aged bottles that have been perhaps scrutinized for provenance at the dinner table, to the selection of distribution and marketing channels possibly subject to the arguably unnecessarily complex three-tier system, from different regimes at bottling that might cause or prevent wine faults in years to come, to numerous experiments done at different stages of *élevage* in the winery to finetune the final product, from intimate decisions on soil treatment, irrigation, vine training and trellising based on vintners' experience, ideals, and terrior, to the processes and experiments revolving around scions and rootstocks in the nursery, it might strike as without doubt that there is perhaps little space for artificial intelligence (AI) at its current state to take hold in the wine industry in the near future.

After all, the thought of ordering a bottle of wine with personal recommendations through a robot at a dinner table, or conversing with Google Home or Amazon Alexa about the intricacies that make Musigny more different than Bonnes Mares than Les Amoureuses, devoid of any human touch of hospitality, would easily make one squirm.

On the other hand, artificial intelligence has made breathtaking breakthroughs in multiple fields in the past decade, not only solving some of the world's most pressing and challenging puzzles, but also penetrating various aspects of our daily lives. AI is making it easier for people to do things every day, whether it's searching for photos of loved ones with a simple query, breaking down language barriers with smart online translators, typing emails with automatic completion, or getting things done with the Google Assistant. AI also provides new ways of looking at existing problems, from rethinking healthcare to advancing scientific discovery. One particularly relevant research theme that is quickly emerging is AI for Social Good, which uses and advances artificial intelligence to address societal issues and improve the well-being of the world. In an excellent review article by the AI and Social Good Lab at Carnegie Mellon University summarized over one thousand relevant academic articles published in top computer science conferences in the following plots by application areas over time:

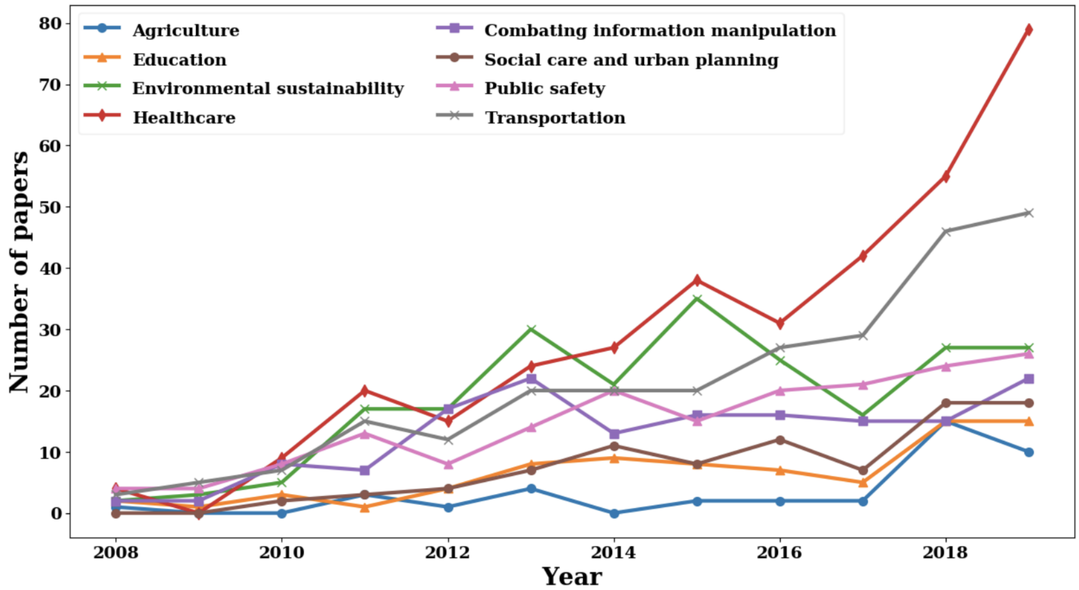


Figure 1: Evolution of AI for Social Good Application Domains [Shi et al., 2020].

With the steady (even exponential) growth of AI technology in various public domains, there is no reason why the wine industry, that overlaps with so many other industries in Figure 1 wouldn't benefit from AI technology. It is my strong conviction that various AI technologies can already resolve many issues of the wine industry in a surprisingly efficient manner, and I am going to show you how in this book in a fun and non-technical way that your parents would understand and hopefully agree.

Objectives of This Book

How could AI assist wine professionals in various aspects of the wine world, perhaps change the wine industry for the better, and ultimately enrich consumers' experience? I try to illustrate by

- examining the essential qualities and responsiblities of wine professionals or sommeliers through the lens of AI, and detailing how to use AI to help with relevant tasks;
- identifying components of the wine industry where AI could potentially improve upon wine professionals, or make their job easier and more efficient at the very least;

- solving challenging problems that have shaken the sommelier circles in recent years;
- laying out future plans to building an ultimate sommelier-in-the-loop AI system for the wine industry.

The Structure of This Book

Chapter by chapter, I will discuss wine-related topics and the challenges therein, followed by to what extent could AI be of help, and in what ways, with introductions to the relevant topics in AI. I try to include as many visualizations, and demos, as possible to make it fun. More *interactive* visualizations, technical versions of the AI parts (in the form of literature review in scientific papers from a previous version of this book), and other topics are available online at https://ai-somm.com.

This book can be read in at least three distinctive ways. First, each chapter provides a visualization that aggregates certain wine knowledge in a (hopefully) creative way. Second, every chapter addresses one of the wine-related challenges commonly faced by wine professionals and/or enthusiasts with a set of AI solutions, introduced in a nontechnical way alongside visualizations, sample results, and demos, when appropriate. Third, every chapter further the discussion, in subsections, of relevant AI methods with a self-contained and relatively non-technical review of method development and evolution over the past decade, while technical reviews are also made available online at https://ai-somm.com. Therefore, each chapter assumes in parallel two themes, one relevant to the wine industry, the other the AI industry. Yet both parts would be self-contained thus no previous knowledge is required to grasp the texts, except a curious mind and a playful heart. In addition, each chapter is in itself self-contained and can be read separately, therefore readers are welcome to read in whatever order you like.

Hopefully, this book makes a unique and novel addition to the wine literature, and the AI literature, while being broad enough in scope to be of use across the wine profession, and perhaps inspire AI applications in other fields as well. Because of the fast-evolving nature of the AI technology (and perhaps the wine industry too), I hope to continuously update the chapters with the newest methods and introduce new topics, possibly with a second edition, a third edition, and so on.

A Preview of Chapters

Chapter 1 discusses in-depth about what wine professionals and enthusiasts love (and hate): blind tasting. It has been an essential part of training for wine professionals. However, it does appears that everyone has his or her own unique marker or method, on top of the generally accepted so-called “deductive tasting“. I detail some of the many schools of thought about how to conduct deductive tasting, highlighting their major flaws and inconsistencies, while illustrating how this exact problem corresponds to some of the most classic machine learning methods, which in turn could be used to prevent pitfalls and identify the optimal strategy of deduction.

Chapter 2 gets into the weeds of the vast body of wine knowledge touching on various distinct yet intertwined subjects such as geology, geography, chemistry, viticulture, viniculture, economics, etc. A solid grasp of a large body of wine knowledge is fundamental to being a qualified wine professional, just as how knowledge graphs are fundamental to various AI models and their generalizability[1] and flexibility. We recount the important roles knowledge graphs have been playing in modern AI ecosystems, and illustrate with examples how knowledge graphs could be integrated to build question-answering systems like chatbox applications tailored to the wine industry.

Chapter 3 broaches the classic topic of food and wine pairing. Given the textual description of a dish and the identify of a bottle of wine, how could AI methods be used to help determine their compatibility? Given a random food image, how would AI models recommend a wine to pair with, with rationales? Furthermore, given a bottle wine, how could we generate a recipe for a dish that goes well with it, with personal preference customization? We will break down each of the scenarios, and explain AI solutions module by module.

Chapter 4 explores the colorful landscape of wine maps, by comparing various wine map collections and cartography projects. Map-making, or cartography, has long been a labor intensive and time-consuming process that requires extensive and in-depth knowledge of visual design, geography, perception, aesthetics, etc., on the part of cartographers or designers, despite the powerful modern softwares like Adobe Illustrator and ArcGIS that have partially eased the process. When it comes artisanal wine maps that are artistically stylized, however, manual hand-drawing appears inevitable. Could AI help automatically generate artistic maps with style and precision in no time? The answer is yes, yet not without challenges.

Chapter 5 describes the phenomena of flying winemakers, and globe-trotting wine professionals and enthusiasts, and introduces the wine equivalents of the fun game GeoGuesser: VineyardGuesser — given an image of a vineyard, guess where it is located in the world, and CellarGuesser — given an image of a cellar, guess which winery it is! Can you achieve more correct guesses than our AI Guesser? You might be surprised. We will discuss the ins and outs of image geolocalization and how it applies to vineyards and cellars.

Chapter 6 details the fascinating world of grape varieties. Which grape varieties in the world are similar in terms of fruit profile, or structure, or growing patterns? What are the varieties that share something in common with both Riesling and Viognier? To answer such questions and many more, with the help of some of the widely used methods in AI, we produce a comprehensive map of the world's thousands *vitis vinifera*, from which links and associations among grape varieties could be easily identified. Could AI help with grape variety identification in the vineyard with a single photo of the grape vine on the ground? The answers are indeed positive, with the help of fine-grained visual classification applications in computer vision.

Chapter 7 maps out the kaleidoscopic space of (craft) cocktails as a semantic network[2]. What makes a cocktail creative? There is a popular misconception that a great idea strikes from out of the blue, much like the apple that supposedly fell on Newton's head. In fact,

[1]The extent to which these models can be generalized to other domains.

[2]A network of interconnected concepts.

almost every idea, no matter how groundbreaking or innovative, depends closely on those that came before. We analyze the creativity of craft cocktails through the lens of semantic networks and network theory, and provide creative tools and insights for aspiring mixologists. Furthermore, with the help of recent advancement in text generation technologies, we demonstrate how to automatically generate creative cocktail recipes, given minimal inputs.

Chapter 8 examines some of the world's best curated wine lists and explores what makes a great wine list in a data-driven manner. We introduce AI methods particularly adapted to parse a wine list, provide a comprehensive evaluation of any given wine list, and ultimately, generate a wine list given certain constraints such as budget, restaurant theme, perceived creativity, target consumer segments, etc., envisioning the future of AI assistants to wine directors at Michelin-starred restaurants and rustic bistros alike.

Chapter 9 seeks to tease out the causal effects of *Terrior* vs. *Vignerons* on wine quality, as opposed to spurious correlation, by introducing the most classic methods in Econometrics[3] and Statistical Learning[4] of causal inference, as well as their modern renditions in AI research.

Chapter 10 touches on the good old problem of trust-building among supply chain partners in the wine industry. Unsurprisingly, this is by no means a problem unique to the wine industry, therefore we review research efforts and practical insights over the past decade or so on the topics of automatic deception detection, and information concealment detection in text and speech with practical demos as potential solutions to some issues in the wine industry.

Chapter 11 elaborates on the worldwide wine auction scene. What are the optimal strategies for the auctioneer and the customers, respectively? What are some pitfalls corresponding to different mechanism designs from the perspective of customers? How could we induce truth-telling and perhaps greater market efficiency with mechanism design of auctions? In this chapter, we delve deep into the classic game theory and mechanism design that prove wildly relevant in the modern world.

Societal biases could result in social inequity through prejudice (emotional bias), stereotypes (cognitive bias), and discrimination (behavioral bias). The wine industry is no exception. In **chapter 12**, we review how AI could augment or mitigate societal biases in various stages, pinpoint the major sources of algorithmic biases, and highlight the dangers of AI applications if such issues are not taken seriously.

Background Information

Before we start, let us clarify the meaning of artificial intelligence with some background information, as well as several closely related terms you will frequent in this book, and the AI community in general.

With the ever increasing amounts of data in digital form, the need for automated methods

[3] The subject of the application of statistical methods to economic data for meaningful interpretation of economic behaviors and activities.

[4] The sub-field of machine learning drawing from the fields of statistics and functional analysis.

for data analysis continues to grow. The goal of **machine learning (ML)** is to develop methods that can automatically detect patterns in data, and then to use the uncovered patterns to predict future data or other outcomes of interest. Machine learning is thus closely related to the fields of statistics and data mining, but differs slightly in terms of its emphasis and terminology.

Data mining deals with challenges particularly in the areas of data storage, organization and searching.

The learning problems that we consider can be roughly categorized as either supervised or unsupervised, with those in-between termed semi-supervised. In supervised learning, the goal is to predict the value of an outcome measure based on a number of input measures; in unsupervised learning, there is no outcome measure, and the goal is to describe the associations and patterns among a set of input measures. **Statistical learning** brings together many of the important new ideas in learning from data, and explain them in a statistical framework.

The desire of creating machines that think dates back to at least ancient times when the tales of giant bronze robot Talos, artificial woman Pandora and their creator god, Hephaestus, filled the imaginations of people in ancient Greece. When computers first came into being, people had wondered whether they might become intelligent. Today, **Artificial intelligence (AI)** as a field has come a long way with numerous practical applications and active research topics, from intelligent software for automating routine labor, to speed and language understanding, from image and video perception, to scientific diagnoses in medicine, and many more. In the early stages of AI, problems that are intellectually difficult for human beings but relatively straightforward for computers were quickly solved. These are the ones that can be described by a list of mathematical rules. The real challenge to AI remains those easy for humans to perform that are difficult to articulate — those we solve effortlessly and intuitively, such as recognizing faces in images, or recognizing spoken words in a speech. The solution to such tasks — natural for human but challenging for machines, is to allow computers to learn from experience, mostly in the form of data, and understand the world in terms of a web of concepts, the hierarchy of which allows the computers to learn complicated concepts by building them upon simpler ones, just like how humans learn. If we draw graphs of these learned concepts built on top of one another, these graphs are deep with many layers. Therefore, this approach to AI is termed **Deep learning**.

Natural language processing (NLP) is the use of human languages, such as English or Japanese, by a computer. Different than computer languages that were designed to allow efficient and unambiguous parsing, natural language processing commonly revolves around resolving ambiguous and informal descriptions, and includes applications such as question answering (covered in Section 2.2), text generation (covered in Section 3.2, Section 8.2, and Section 7.3), machine translation (touched on in Section 2.1), named entity recognition (touched on in Section 2) and many more.

Speech recognition aims to map an acoustic signal containing a spoken natural language utterance into the corresponding sequence of words intended by the speaker. The **automatic speech recognition (ASR)** task aims to identify an automatic function for that mapping, nowadays mostly based on *deep learning* methods. We will touch on some parts of it in Section 10.1.

Traditionally one of the most active research area for deep learning applications since vision is a task effortless for humans and animals but challenging for computers, **computer vision (CV)** is a very broad field consisting of a wide variety of methods of processing images resulting in an amazing diversity of applications, ranging from reproducing human visual abilities, for instance, recognizing faces, to creating new categories of visual abilities, such as recognizing sound waves from silent videos based on vibrations induced in objects visible therein. Many of the most popular standard benchmark tasks for *deep learning* methods are forms of *object recognition*, covered in Section 3, Section 6 and Section 5, as well as *optical character recognition*, covered in Section 4.4. Computer vision also overlaps with *computer graphics*, surfacing creative and efficient solutions to problems such as repairing defects in images, coloring black and white images, artistically stylize photos, and many more. We will discuss some of these in Section 4.

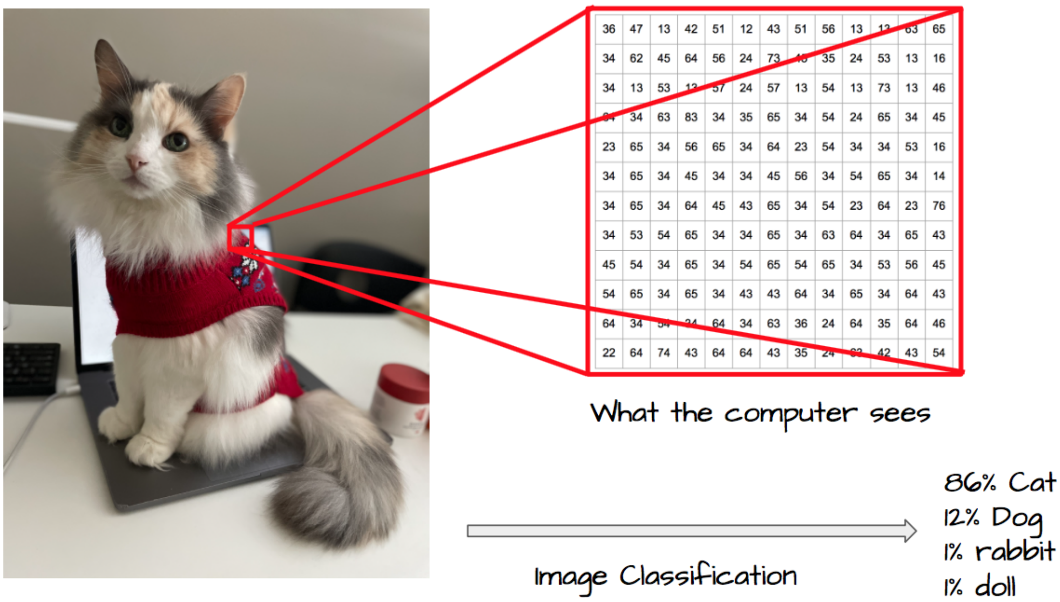
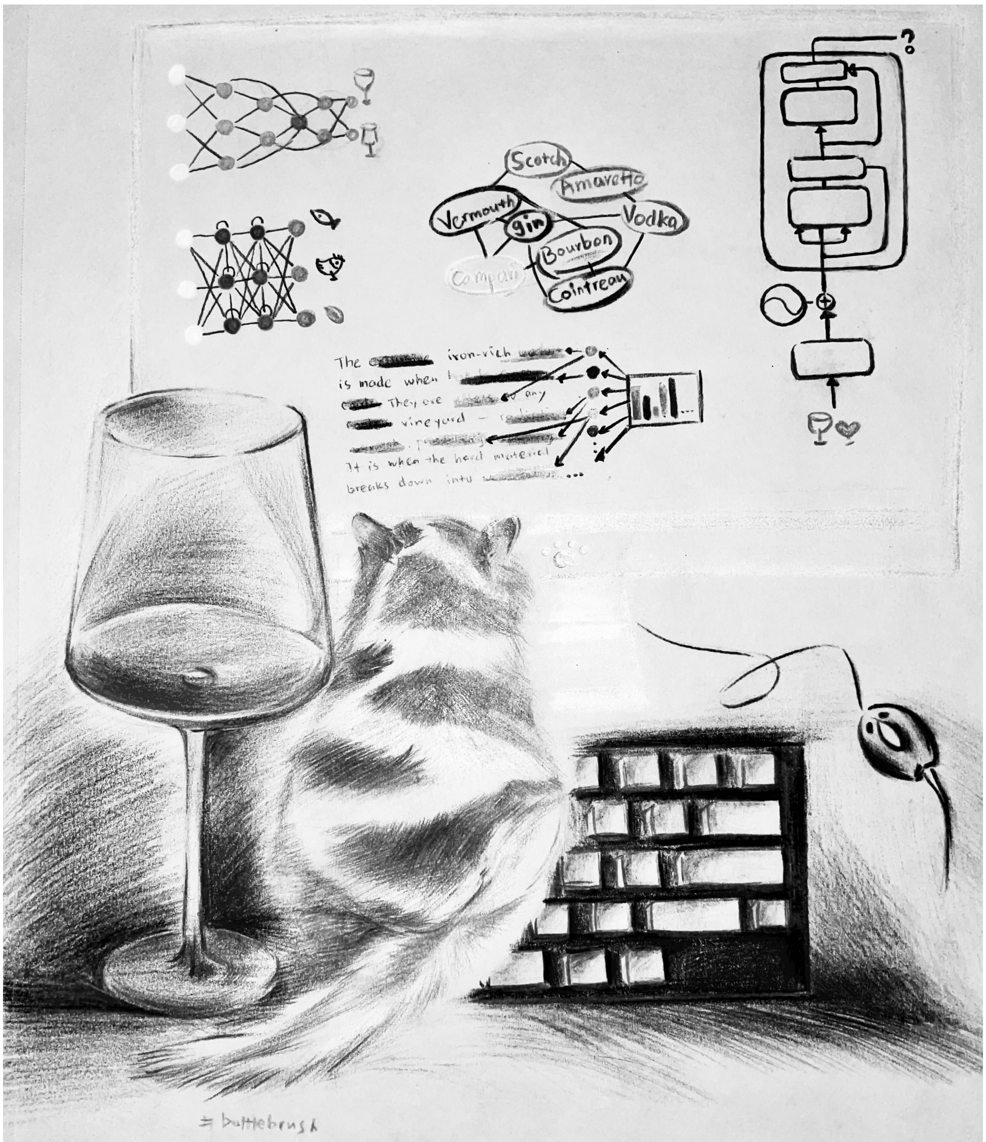


Table of Contents


Scotch
Amaretto
Vermouth
gin
Vodka
Bourbon
Cointreau

SECTION 1

Deductive Tasting

Blind tasting is one of the favorite games among wine enthusiasts. A mysterious bottle of wine wrapped in opaque papers or pouches, poured into a delicate hand-blown crystal glass, revealing its clear deep crimson color. Rich and opulent, pure and high-toned bouquet jumped out of the glass. Beautiful wild red cherries, vigor, fresh, and reverberant. A hint of roses and violets crackling with a touch of white pepper. Stony satiny tannins perfectly balanced with its tension and energy. What is it, you wonder in ecstasy, swirling the elixir gently to take in all its intricate aromatics. Familiar memories conjured up in your mind. *You were wondering within a deep forest, thick with pristine primeval growths. As the humid scent of life wafts from the moss-covered trees. You walked towards the heart of forest in search of solace. Suddenly you noticed a ray of light. You smelt flowers and red fruits that seem out of place in these woods. Unexpectedly the forest opened to a small clear spring, pouring forth like a miracle, like an oasis in the desert. The restoring water appears out of nowhere, and the surface glitters like so many jewels lit by the heavens. Drawn by the beauty, you quietly approached the spring. In that moment, the breeze rippling across the water, delivers to your nostrils the smell of sweet flowers and wild red fruit, so sweet and exalted. Up in the air, a pair of violet butterflies tangling in flight!*[5] There is no other *lieu-dit* in the world that could lead you to a virgin forest like this than Les Amoureuses, the premier cru in Chambolle Musigny, in Cote de Nuits of Cote d'Or, Burgundy. This is George Roumier Les Amoureuses. Vintage... let's say, 1999? You announce to your wine lover friend with whom the bottle is shared, with a somwhat confident smile.

[5] Drops of God, Volume 4. The First Apostle, a.k.a. George Roumier Les Amoureuses. 2001. In the words of Yukata Kanzaki, the fictional world-known wine critic.

Figure 2: An illustration of the imagery that possibly captures G. Roumier Les Amoureuses 2001.

But seriously, blind tasting is one of the essential skills of many wine professionals. For an importer or retailer, to be able to pick out the best quality wine (or the wine most likely to sell) at the most reasonable price contributes directly to the profitability or survival rate of the business. For a wine writer or critic, the capability of correctly judging wine's quality and ageability is very much hinged upon his or her own reputation. For a sommelier, correctly identifying wines blindly not only creates the wow factor for the restaurant[6], but also helps tremendously with building a best wine program given a limited budget.

Therefore, it is no wonder that most rigorous wine study programs include a section on blind tasting in examinations. In The Court of Master Sommelier's tasting exams required to earn the title Master Sommelier, for instance, candidates have to pass an oral example of 25 minutes to identify six wines — three white, three red — correctly in terms of grape variety, region of production, and vintage, by first describing them, one by one, from colors in sight, to aromas on the nose, to flavors or other elements in the mouth, and then concluding with deduced identities. In The Institute of Master of Wine's tasting exams required to obtain the diploma of Master of Wine, as another example, candidates have to sit a written exam of three hours while tasting 12 wines, answering in the form of essays different winemaking techniques or climatic characteristics exemplified in the color, aromas, and tastes of wines, with attempts to identify either the vintage, region, or grape variety, possibly funneling[7]

[6] Like the tales around well-known sommeliers Raj Parr, Fred Dame (exposed in New York Times articles on scandals though), Larry Stone, and the likes.

[7] For instance, if at one point you think the closest you could get with a wine is that it is an Italian red wine due to perhaps its volatile acidity, drying tannins, prominent herbal characters, and an acidic spine. But no clue if it is a Brunello di Montalcino, a Barolo, or an Etna Rosso, you could potentially *funnel* by putting down that you think it could be all three with a slight inclination towards

when uncertainty arises.

COURT OF MASTER SOMMELIERS Americas

DEDUCTIVE TASTING FORMAT

Sight	
Clarity / Visible Sediment	Clear, Hazy, Turbid
Concentration	Pale, Medium, Deep
Color	**White Wines:** Water White, Straw, Yellow, Gold
	Red Wines: Purple, Ruby, Red, Garnet
Secondary Colors	**White Wines:** Silver, Green, Copper
	Red Wines: Orange, Blue, Ruby, Garnet, Brown
Rim Variation	Yes / No
Extract / Staining (Red Wines)	None, Light, Medium, Heavy
Tearing	Light, Medium, Heavy
Gas Evidence	Yes / No
Nose	
Clean / Faulty	TCA, H_2S, Volatile Acidity, Ethyl Acetate, Brettanomyces, Oxidization, Other
Intensity	Delicate, Moderate, Powerful
Age Assessment	Youthful, Developing, Vinous
Fruit	**White:** Citrus, Apple/Pear, Stone/Pit, Tropical, Melon
	Red: Red, Black, Blue
Fruit Character	Ripe, Fresh, Tart, Baked, Stewed, Dried, Desiccated, Bruised, Jammy
Non-Fruit	Floral, Vegetal, Herbal, Spice, Animal, Barn, Petrol, Fermentation
Earth	Forest Floor, Compost, Mushrooms, Potting Soil
Mineral	Mineral, Wet Stone, Limestone, Chalk, Slate, Flint
Wood	None, Old vs New Large vs Small French vs American

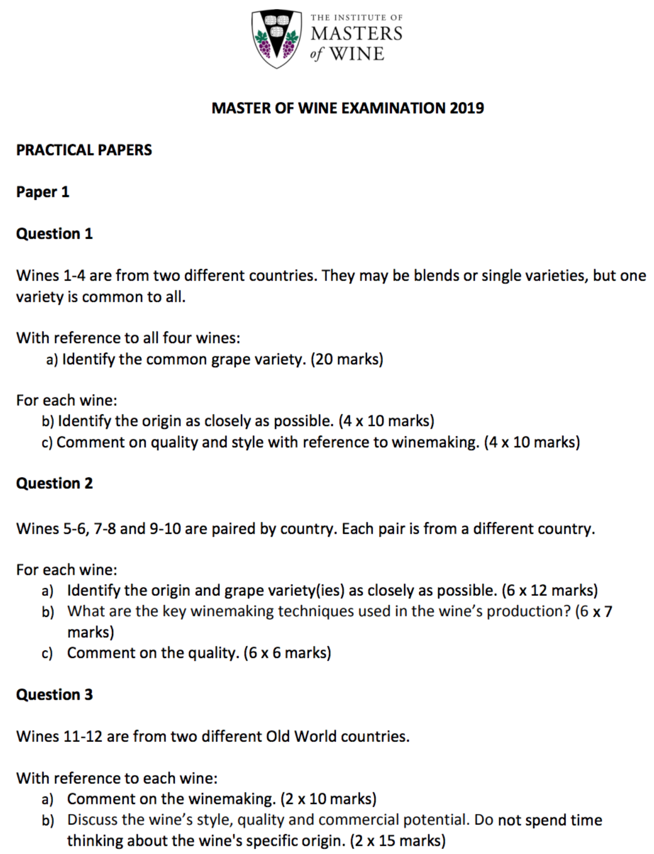
March 2017

Palate	
Sweetness	Bone Dry, Dry, Off-Dry, Medium Sweet, Sweet, Lusciously Sweet
Fruit	**White:** Citrus, Apple/Pear, Stone/Pit, Tropical, Melon
	Red: Red, Black, Blue
Fruit Character	Ripe, Fresh, Tart, Baked, Stewed, Dried, Desiccated, Bruised, Jammy
Non-Fruit	Floral, Vegetal, Herbal, Spice, Petrol, Animal, Barn, Fermentation
Earth	Forest Floor, Compost, Mushrooms, Potting Soil
Mineral	Wet Stone, Limestone, Chalk, Slate, Flint
Wood	None, Old vs New Large vs Small French vs American
Palate	
Phenolic / Bitter (White)	Yes / No
Tannin (Red)	Low, Med-, Medium, Med+, High
Acid	Low, Med-, Medium, Med+, High
Alcohol	Low, Med-, Medium, Med+, High
Body	Light, Medium, Full
Texture	Creamy, Round, Lean
Balance	Does any element dominate?
Length / Finish	Short, Med-, Medium, Med+, Long
Complexity	Low, Med-, Medium, Med+, High
Initial Conclusion	
Possible Grape Varieties	
Old World / New World	
Climate	Cool, Moderate, Warm
Possible Countries	
Age Range	1-3 years, 3-5 years, 5-10 years, 10 years+
Final Conclusion	
Grape Variety / Blend	
Country of Origin	
Region / Appellation	
Quality/ Regional Hierarchy	Grand/Premier Cru, Reserva/Gran Reserva etc.
Vintage	

Figure 3: Tasting format sheet for the Master Sommelier tasting exams.

There have been quite a few different schools of thought regarding how to blind taste, what makes an excellent blind taster, what to look for to improve blind tasting skills, and so forth. One of the most widely accepted approach is *deductive tasting*, possibly popularized by The Court of Master Sommelier and Wine and Spirits Education Trust, which essentially separates the process of blind tasting into two steps: first, describe the wine in terms of color, aroma, and flavor, and structure, as precisely and objectively as possible; second, given the resulting descriptors, logically deduce the identity of the wine without referring back to the liquid in the glass. The first step requires a palate tuned to accurately identify a wide range of aromas and flavors in different forms and levels of doses, from exotic fruits like lychee or tamarind to esoteric flowers like marigold or azalea, from Asian five spices to Comte cheese and Herbes de Provence, from potting soil after an early summer rain, to pencil shavings and graphite, let alone cat urine, dirty socks, wet dogs, barnyard funk, and leather belts. And that is why "licking rocks and eating dirt" are not uncommon perhaps not only among geologists, but also sommeliers — at least those serious about improving palate sensitivity, I guess. It is only when one can objectively identify all the elements in a wine precisely in a consistent manner, can the second step — logical deduction — really shine. In this chapter, I will focus on this second step, the logical deduction. Various advice and toolkits for tuning your palate for the first step have been passed down: constant training with

Etna Rosso due to its volanic characteristics.


THE INSTITUTE OF MASTERS of WINE

MASTER OF WINE EXAMINATION 2019

PRACTICAL PAPERS

Paper 1

Question 1

Wines 1-4 are from two different countries. They may be blends or single varieties, but one variety is common to all.

With reference to all four wines:
a) Identify the common grape variety. (20 marks)

For each wine:
b) Identify the origin as closely as possible. (4 x 10 marks)
c) Comment on quality and style with reference to winemaking. (4 x 10 marks)

Question 2

Wines 5-6, 7-8 and 9-10 are paired by country. Each pair is from a different country.

For each wine:
a) Identify the origin and grape variety(ies) as closely as possible. (6 x 12 marks)
b) What are the key winemaking techniques used in the wine's production? (6 x 7 marks)
c) Comment on the quality. (6 x 6 marks)

Question 3

Wines 11-12 are from two different Old World countries.

With reference to each wine:
a) Comment on the winemaking. (2 x 10 marks)
b) Discuss the wine's style, quality and commercial potential. Do not spend time thinking about the wine's specific origin. (2 x 15 marks)

Figure 4: A sample of Master of Wine tasting exam questions.

wine aroma kits, the sniff and scratch book series by Richard Betts, roasting plain popcorns with different spices — a tip by Jill Zimorski, cooking with a wide range of ingredients and condiments, paying attention to not only the flavors but also the structural elements, the textures, and types and shapes of acidity, to name just a few. Let's assume for now — don't worry, we have solutions to be discussed later for when this assumption is hardly met — that we all have reached the point when we have the perfect palates able to capture the whole spectrum of aromas, flavors, and sensations in a glass of wine.

To logically deduce the identity of any given wine, is to compare the wine in question to stereotypes of wines with known identities in our memory, and find the identity of the most similar stereotype. Therefore, the second step of deductive tasting — logical deduction — can be further divided into two parts.

A first and major part of training for logical deduction, is to build up a robust and comprehensive database of stereotypes of wines of different origins, varieties, vintages, and producers, etc. with objective and subjective descriptors. How would you describe a stereotypical 10-year-old Condrieu from the 2010 vintage? What are the characteristics of a 2013 Aglianico del Vulture in 2020? Luckily, such summmarization tasks are not unique to wine tastings and there is indeed a lot to borrow from the field of *natural language processing* (or more generally, *machine learning*) to accomplish this task in a data-driven manner with much greater precision and capacity than human memorization and manual work. We de-

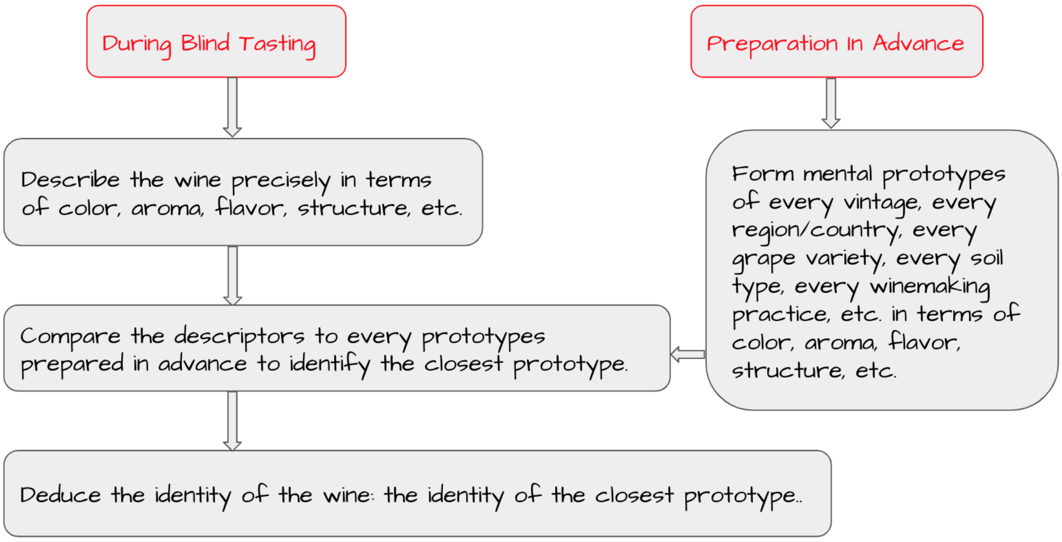


Figure 5: A flowchart of the deductive tasting process.

scribe several such frameworks in Section 1.1.

Secondly, comparing the descriptors of the given wine to those of stereotypes in our collected database in the first step. Humans are notoriously bad in such tasks. For example, in the case of blind tasting, one taster decided to narrow down to only Savennieres, the most “cerebral“ wine producing region, after detecting both Botrytis — honey, marmalade, saffron — and oxidative — almond paste, bruised apple, cheese rind — aromas. However, wouldn’t an aged Montrachet from certain vintages and producers also best exemplify both Botrytis and oxidative notes? One might further confirm or refute the choice of Savennieres with the level, shape, and structure of acidity, as well as unique aromatics since Chenin supposedly is of searingly high and crescendo acidity according to Nick Jackson’s excellent blind tasting book Beyond Flavors, and radiant of fragrances like chamomile, jasmine, honeysuckle, wasabi, and dried stone and tree fruits sometime a touch of tropical too. However, more often than not, one starts to hallucinate certain aromas signature of Savennieres with such an objective in mind, falling victim to the confirmation bias. Blame it on the subjectivity of wine tasting! In another example, one taster might have quickly eliminated varieties like Nebbiolo, Grenache, Pinot Noir, Gamay(, and more indigenous varieties such as Freisa, Ruche, Prie Rouge, Nerello Mascalese, Baga, etc.) simply based on the deep purple color in the glass. However a Hubert Lignier Charmes-Chambertin, a Château de Beaucastel Châteauneuf-du-Pape, or a Yvon Metras Moulin-a-Vent would easily defeat that assumption, in which case the taster would have simply bypassed the correct identities at the initial judgment. The advice of funneling, or enlisting all the possible “grape laterals“ — easily confused or similar grape varieties — has been circulated among some wine study circles. But a Pinot Noir might be similar to Gamay, which might be similar to Nebbiolo in some capacity, which then is similar to Sangiovese (ever mistook a Brunello for Barolo?) or Nerello Mascalese or Xinomavro, and the chain never stops. . . . Such begs the question, is there an optimal or systematic way to move the deduction process consistently towards the correct answer as much as possible? In what steps and based on what characteristics should one eliminate or funnel? For example, Abigail might start with color, then aromatics on the nose, then flavors and finally structural components on the palate, therefore deduce by eliminating most varieties by color, draw initial conclusions based aromatics on the nose and palate, and narrow down to or confirm the final conclusion with the structural components. But Bob perhaps might argue one should use the structural components to come up with a list of initial conclusions and drill down to a few based on aromatics and flavors, and finally confirm with color or quality indicators. Yet another pro Claire might instead use fruit categories and conditions (crunchy tree fruit or jammy stone fruit?) on the nose versus on the palate (if fruit went tart on the palate compared to the nose it might be indicative of certain regions) for initial conclusions, and structural components for final conclusions. Or if Claire is not good at judging the level or type of acidity, she might choose to not rely on structural components as much and use them sparingly. Whose strategies might most consistently lead to the most correct answers in blind tasting sessions? What is the optimal strategy based on one's strengths and weaknesses? For instance, if Claire is confident in her ability to detect spices but lacking in acidity calls, whereas Bob can never detect Rotundone (the chemical compound supposedly responsible for smells of black pepper) but is excellent in accessing fruit aromas and flavors. How should their blind tasting strategies differ to accommodate these strengths and weaknesses? What if we were blind tasting for vintage alone, or variety alone, or country? How would the optimal deduction strategy change according to the target? Intuitively, there might be a much smaller set of characteristics we watch out for if we are trying to decide on the country alone, compared to vintage or variety. Such are exactly what we dive into in Section 1.2 and Section 1.3.

1.1 Summarization

1.2 Decision Tree

1.3 Structured Prediction

2

SECTION

Wine Theory

2.1 Knowledge Graph

2.2 Question Answering

3

SECTION

Food and Wine Pairing

3.1 Food Recognition

3.2 Recipe Generation

Given an image of a plate of food besides a bottle of wine, retrieve/generate the recipe and decide whether its a good pairing

3.3 Recommender Systems

4

SECTION

Cartography

Maps are beloved by the wine world. From my early exposure during the famed course Intro to Wine in college taught by the one and only Cheryl Stanley, to my later studies through the Court, WSET, and (ongoing) MW program, Map drawing as an effective strategy for studying wine theory has been consistently echoed by Cheryl, various Master Sommeliers, Master of Wines, and other senior wine professionals. Some even get into the art and science of cartography – from Alessandro Masnaghetti, a nuclear-engineer-turned mapman, who made groundbreaking maps of wine regions, especially known for the crus of Barolo and Barbaresco, to Deborah Steve De Long, a textile designer and an architect, whose maps are a true labor of love. Like many wine lovers, I greatly apprecaite the exquisite pieces of wine maps with excrusiating details of geology, geography, soil, vine growing, winery information, etc., and collect as many as I can. Starting with the detailed professional maps from the World Atlas of Wine by Jancis Robinson and Hugh Johnson:

To the clear and simple study guides from The Society of Wine Educators:

From GuilSomm:

To Steve De Long:

From Champagne by Peter Liem:

To Jasper Morris Burgundy:

From Bordeaux by Jane Anson and :

To Alex MGA:

From the studious handdrawings for MS study by Jill Zimorski MS that are quite professional looking:

images/atlas.png

Figure 6: A Map from the The World Atlas of Wine Map Set.

To the fabulous watercolor paintings of sommelier James Singh — Children's Atlas of Wine:

My obsession with exquisite wine maps — especially those that perfectly combine dense, precise information and aesthetics knows no bounds. But Mapmaking has been a labor intensive and time-consuming process that requires extensive and in-depth knowledge of visual design, geography, perception, aesthetics, etc., on the part of cartographers or designers, despite the powerful modern softwares like arcGIS and Adobe Illustrator that have partially eased the process compared to manual mapdrawing. I've always lamented how few wine regions James Singh have covered so far with his masterful skills of watercolor mapping. What if, given a basic professional wine map of Burgundy, and a beautiful watercolor map (like Children's Atlas of Wine maps) of another region, say Bordeaux, we could automatically generate a beautifully rendered watercolor map of Burgundy in the style of the Bordeaux map!?

Luckily, computer vision researchers have been working hard on this exact problem — well, images/swe.png

Figure 7: A Map from the The Society of Wine Educators Map Set.

almost! — and with the era of deep learning, the subject of neural style transfer that exploded circa 2015 swept the field with breathtaking results, answering questions like, what would Monet have painted if he saw Degas's ballet dancers, and what Degas would have painted if presented with Monet's garden?
Given the content image on the top left, and the three style images representative of the three artists — JMW Turner's The shipwreck of the Minotaur, 1805-1810..., Vincent van Gogh's Starry Night, and Edvard Munch's The Scream, shown at the left bottom corners of the rest three images, as the respective style images, what L. Gatys and colleagues proposed as the neural style transfer algorithm generated the pleasing results of the accordingly stylized paintings. A brand new era began ever since... How about applying to wine maps? Give me some watercolor artisanal wine maps, in bulk, please! Turns out existing cutting edge computer vision research does have its own share of foes... And in most cases, the algorithm does not do well, especially when it comes to tiny blocks of texts intertwined with complex artistic patterns... But here is a promising first step — uCAST, unsupervised CArtographic

images/.png

Figure 8: A Map from the Map Set.

Style Transfer, by me:
We will discuss the ins and outs of how to make it work, and what the field of neural style trasnfer, along with its close sibling image-to-image translation, is all about in the next part — AI Talk.

4.1 Image-to-image Translation

Image-to-image translation methods have gained significant traction over the past few years ever since the first unified framework based on conditional GANs was proposed by Phillip Isola and colleagues, though the idea dates back at least to image analogies. Such methods require image pairs each of which include one image of the original style, and the other of the desired style, sharing the exact same content preferably perfectly aligned. To relax such a somewhat restrictive constraint for greater practical accessability, unsupervised image-to-image translation methods were proposed for unpaired image datasets by introducing

images/.png

Figure 9: A Map from the Map Set.

additional constraints by either preserving certain properties such as pixel features, semantic features, and latent spaces, etc., or through loss functions to ensure cycle consistency, distance consistency, geometry consistency, etc. Such methods work great as long as within each set, the image styles are consistent and different sets of images don't differ too much in terms of domain. For instance, transfering cats to dogs would probably work well whereas transfering cats to airplanes probably woouldn't. To solve this domain shift problem and enable models to generate diverse styles, multi-domain and multimodal methods (BicycleGAN, MUNIT, DRIT++, StarGAN) have also been introduced to generate diverse images across multiple different domains. The current work is also related to the instance-level image-to-image translation methods, which improve upon the global methods mentioned above in complex scenes. InstaGAN was the first work to tackle instance-level translation. It takes as input objects' segmentation masks in the two domains of interest, and translates between the object instances without altering the background. In contrast, INIT and DUNIT, both instance-level image translation methods, translate the entire image. INIT

images/.png

Figure 10: A Map from the Map Set.

propose to define a style bank to translate the instances and the global image separately. During training, INIT treats the object instances and the global image independently, thus at test time, it does not exploit the object instances, going back to image-level translation. DUNIT propose to unify the translation of the image and its instances, leveraging the object instances at test time. uCAST seeks to translate the entire image as well, by leveraging contrastive learning, CUT proposes a simple patch-based image synthesis approach via maximizing the mutual information between corresponding patches in the input and output images. uCAST deviates in that none above considers paralleled text style transfer which is unique to the task of uCAST.

4.2 Style Transfer

Neural style transfer performs image-to-image translation by synthesizing a novel image by combining the content of one image with the style of another image by matching the

images/.png

Figure 11: A Map from the Map Set.

Gram matrix statistics of pre-trained deep features. Local or semantic style transfer methods emphasize semantic accuracy. For instance, Li et al. match each input neural with the most similar patch in the style image for targeted transfer, Chen et al. preserves the spatial correspondence by way of masking and higher order style statistics to generalize style transfer to semantic-aware or saliency-aware neural algorithms. Gatys and colleagues transfer the complete "style distribution" of the style image through the Gram matrix of activated neurons. Fujun Luan further improves by incorporating a semantic labeling of the input and style images into the transfer procedure so that the transfer happens between semantically equivalent subregions. By introducing a contextual loss, ContextualLoss investigates semantic style transfer without segmentation for non-aligned images. Efforts such as Gu et al. and Huang et al. have been made to synthesize both local and global style transfer methods to enjoy the best of both worlds. None of this body of work distinguishes artistic patterns and artistic texts separately, and when applied to functional artworks, words tend to be illegible, misplaced, or not correspondingly stylized.

images/.png

Figure 12: A Map from the Map Set.

4.3 Font Transfer

Neural font and artistic text synthesis studies apply neural style transfer methods to generate new fonts and text effects. Font style transfer problems have been explored in various aspects, ranging from Atarsaikhan et al. (2016) where Gatys et al. was directly applied, to Azadi et al. (2018) where a conditional GAN model for glyph shape prediction and an ornamentation network for colour and texture prediction are trained in an end-to-end manner to achieve multi-content few-shot font style transfer. Text effects transfer was introduced by Yang et al. (2017), in which a texture-synthesis-based non-parametric method was proposed. Yang et al. (2018) synthesize artistic style and target texts in an unsupervised manner that blends seamlessly into the background image. Texture Effects Transfer GAN (TET-GAN) jointly train two parallel subnetworks for text style transfer and removal to provide more seamless integration of text effects into background images. Yang et al. (2019) control the stylistic degree of the glyph with a bidirectional shape-matching framework.

images/.png

Figure 13: A Map from the Map Set.

4.4 Scene Text Detection and Recognition

images/.png

Figure 14: A Map from the Map Set.

images/.png

Figure 15: A Map from the Map Set.

images/.png

Figure 16: A Map from the Map Set.

images/.png

Figure 17: A Map from the Map Set.

images/.png

Figure 18: Conditional GAN.

images/.png

Figure 19:

images/.png

Figure 20: NST.

images/.png

Figure 21: Deep Photo Style Transfer.

5

SECTION

World of Wine

Figure 22: Multi-content Few-shot Font Style Transfer.

images/.png

Figure 23: Text Effect Style Transfer.

5.1 Geolocalization with Image Retrieval

Image retrieval-based methods for geolocalization have been explored extensively where a query image is matched to the most similar reference images in a large image database of tens of thousands of images for a single city. Various low-level image features have been exploited to perform matching: BOW [Schindler et al., 2007, Gronat et al., 2013], color or texton histograms [Hays and Efros, 2008, Kalogerakis et al., 2009, Hays and Efros, 2015], descriptors like SIFT, SURF [Hakeem et al., 2006, Zamir and Shah, 2010, Gronat et al., 2013, Arandjelović and Zisserman, 2014], point and line features [Ramalingam et al., 2011, Li et al., 2014], building patterns [Torii et al., 2013], keypoints [Chen et al., 2011], GIST [Zamir and Shah, 2014, Zemene et al., 2018], among others. Various matching methods were proposed based on feature and geometric correspondence [Li et al., 2010, Bansal and Daniilidis, 2014, Gopalan, 2015, Zemene et al., 2018], representation learning [Gopalan, 2015, Zemene et al., 2018, Liu et al., 2019], segmentation [Ramalingam et al., 2010, Baatz et al., 2012], discriminative learning [Cao and Snavely, 2013], feature voting [Liu et al., 2020], feature reweighting [Kim et al., 2017], pose estimation [Ramalingam et al., 2011], etc. The extensive coverage of reference images required limits the applicability of such methods, and are mitigated by research efforts on cross-view image geolocalization [Lin et al., 2013, Tian et al., 2017] and weak supervision [Ge et al., 2020, Arandjelovic et al., 2016], etc. Our work complements this stream of research that circumvents the necessity of a large image database with greater transferrability and flexibility without additional visual datasets such as aerial [Lin et al., 2013, Tian et al., 2017, Hu et al., 2018] or unlabeled [Arandjelovic et al., 2016] ground-view images.

Classification-based image geolocalization provides a more memory and disk efficient alternative to retrieval-based solutions, by treating the task as a classification problem that divides the map into multiple discrete classes. gronat2013learning leverages geotags to train classifiers per location. hongsuck2018cplanet extends [Weyand et al., 2016] by enhancing the resolution of geoclasses into which convolutional neural networks classify with combinatorial partitioning. Such methods, despite quantization [Jegou et al., 2010] efforts, still require a large amount of image data for training and the performances degrade when less high-quality visual data is available. Our methods relax such constraints, and introduces greater modeling flexibility with widely available city-scale textual or numerical data sources.

Methodologically, our work is closest to visual landmark identification [Chen et al., 2011] — retrieving images from reference landmark databases — where descriptors of extracted image features are quantized to visual words, and methods from text search applied for image retrieval. matsuo2017twitter combines both textual and visual features from tweets for geolocalization by querying a large multimodal database. Unlike others that rely on a reference visual dataset, our methods seek alternative data sources that are comparatively disk efficient (storing metadata instead of images) and combine them in a search algorithm for efficient inference.

5.2 Fine-grained Visual Classification

6

SECTION

Grape Varieties

6.1 Contextual Embedding

6.2 Fine-grained Visual Classification

7

SECTION

Craft Cocktails

7.1 Semantic Networks

7.2 Beauty in Averageness

7.3 Recipe Generation

8

SECTION

Wine Lists

8.1 Information Retrieval

8.2 Text Generation

9

SECTION

Terrior

9.1 Causal Inference

9.1.1 Regression Discontinuity

9.1.2 Natural Experiments

SECTION 10

Trust and Ethics

10.1 Deception Detection

10.2 Information Concealment Detection

11

SECTION

Wine Auction

11.1 Game Theory

11.2 Mechanism Design

12

SECTION

Social Inequity

12.1 AI Ethics

12.2 Algorithmic Biases

SECTION 13

References

[Arandjelovic et al., 2016] Arandjelovic, R., Gronat, P., Torii, A., Pajdla, T., and Sivic, J. (2016). Netvlad: Cnn architecture for weakly supervised place recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5297–5307.

[Arandjelović and Zisserman, 2014] Arandjelović, R. and Zisserman, A. (2014). Dislocation: Scalable descriptor distinctiveness for location recognition. In *Asian Conference on Computer Vision*, pages 188–204. Springer.

[Baatz et al., 2012] Baatz, G., Saurer, O., Köser, K., and Pollefeys, M. (2012). Large scale visual geo-localization of images in mountainous terrain. In *European conference on computer vision*, pages 517–530. Springer.

[Bansal and Daniilidis, 2014] Bansal, M. and Daniilidis, K. (2014). Geometric urban geo-localization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3978–3985.

[Cao and Snavely, 2013] Cao, S. and Snavely, N. (2013). Graph-based discriminative learning for location recognition. In *Proceedings of the ieee conference on computer vision and pattern recognition*, pages 700–707.

[Chen et al., 2011] Chen, D. M., Baatz, G., Köser, K., Tsai, S. S., Vedantham, R., Pylvänäinen, T., Roimela, K., Chen, X., Bach, J., Pollefeys, M., et al. (2011). City-scale landmark identification on mobile devices. In *CVPR 2011*, pages 737–744. IEEE.

[Ge et al., 2020] Ge, Y., Wang, H., Zhu, F., Zhao, R., and Li, H. (2020). Self-supervising fine-grained region similarities for large-scale image localization. *arXiv preprint arXiv:2006.03926*.

[Gopalan, 2015] Gopalan, R. (2015). Hierarchical sparse coding with geometric prior for visual geo-location. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2432–2439.

[Gronat et al., 2013] Gronat, P., Obozinski, G., Sivic, J., and Pajdla, T. (2013). Learning and calibrating per-location classifiers for visual place recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 907–914.

[Hakeem et al., 2006] Hakeem, A., Vezzani, R., Shah, M., and Cucchiara, R. (2006). Estimating geospatial trajectory of a moving camera. In *18th International Conference on Pattern Recognition (ICPR'06)*, volume 2, pages 82–87. IEEE.

[Hays and Efros, 2008] Hays, J. and Efros, A. A. (2008). Im2gps: estimating geographic information from a single image. In *2008 ieee conference on computer vision and pattern recognition*, pages 1–8. IEEE.

[Hays and Efros, 2015] Hays, J. and Efros, A. A. (2015). Large-scale image geolocalization. In *Multimodal location estimation of videos and images*, pages 41–62. Springer.

[Hu et al., 2018] Hu, S., Feng, M., Nguyen, R. M., and Hee Lee, G. (2018). Cvm-net: Cross-view matching network for image-based ground-to-aerial geo-localization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7258–7267.

[Jegou et al., 2010] Jegou, H., Douze, M., and Schmid, C. (2010). Product quantization for nearest neighbor search. *IEEE transactions on pattern analysis and machine intelligence*, 33(1):117–128.

[Kalogerakis et al., 2009] Kalogerakis, E., Vesselova, O., Hays, J., Efros, A. A., and Hertzmann, A. (2009). Image sequence geolocation with human travel priors. In *2009 IEEE 12th international conference on computer vision*, pages 253–260. IEEE.

[Kim et al., 2017] Kim, H. J., Dunn, E., and Frahm, J.-M. (2017). Learned contextual feature reweighting for image geo-localization. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3251–3260. IEEE.

[Li et al., 2014] Li, A., Morariu, V. I., and Davis, L. S. (2014). Planar structure matching under projective uncertainty for geolocation. In *European Conference on Computer Vision*, pages 265–280. Springer.

[Li et al., 2010] Li, Y., Snavely, N., and Huttenlocher, D. P. (2010). Location recognition using prioritized feature matching. In *European conference on computer vision*, pages 791–804. Springer.

[Lin et al., 2013] Lin, T.-Y., Belongie, S., and Hays, J. (2013). Cross-view image geolocalization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 891–898.

[Liu et al., 2020] Liu, D., Cui, Y., Guo, X., Ding, W., Yang, B., and Chen, Y. (2020). Visual localization for autonomous driving: Mapping the accurate location in the city maze. *arXiv preprint arXiv:2008.05678*.

[Liu et al., 2019] Liu, L., Li, H., and Dai, Y. (2019). Stochastic attraction-repulsion embedding for large scale image localization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2570–2579.

[Ramalingam et al., 2011] Ramalingam, S., Bouaziz, S., and Sturm, P. (2011). Pose estimation using both points and lines for geo-localization. In *2011 IEEE International Conference on Robotics and Automation*, pages 4716–4723. IEEE.

[Ramalingam et al., 2010] Ramalingam, S., Bouaziz, S., Sturm, P., and Brand, M. (2010). Skyline2gps: Localization in urban canyons using omni-skylines. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3816–3823. IEEE.

[Schindler et al., 2007] Schindler, G., Brown, M., and Szeliski, R. (2007). City-scale location recognition. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–7. IEEE.

[Shi et al., 2020] Shi, Z. R., Wang, C., and Fang, F. (2020). Artificial intelligence for social good: A survey. *arXiv preprint arXiv:2001.01818*.

[Tian et al., 2017] Tian, Y., Chen, C., and Shah, M. (2017). Cross-view image matching for geo-localization in urban environments. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3608–3616.

[Torii et al., 2013] Torii, A., Sivic, J., Pajdla, T., and Okutomi, M. (2013). Visual place recognition with repetitive structures. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 883–890.

[Weyand et al., 2016] Weyand, T., Kostrikov, I., and Philbin, J. (2016). Planet-photo geolocation with convolutional neural networks. In *European Conference on Computer Vision*, pages 37–55. Springer.

[Zamir and Shah, 2010] Zamir, A. R. and Shah, M. (2010). Accurate image localization based on google maps street view. In *European Conference on Computer Vision*, pages 255–268. Springer.

[Zamir and Shah, 2014] Zamir, A. R. and Shah, M. (2014). Image geo-localization based on multiplenearest neighbor feature matching using generalized graphs. *IEEE transactions on pattern analysis and machine intelligence*, 36(8):1546–1558.

[Zemene et al., 2018] Zemene, E., Tesfaye, Y. T., Idrees, H., Prati, A., Pelillo, M., and Shah, M. (2018). Large-scale image geo-localization using dominant sets. *IEEE transactions on pattern analysis and machine intelligence*, 41(1):148–161.