Trend Detection

1 Fast Trend

Download the file *spritzer15.zip* from Blackboard. Uncompress the file to retrieve 12,895 text files, each containing a tweet randomly selected from one particular hour of the past.

Explore the hidden topics from the documents to identify an international event of major[1] political consequences.

[1] Although the meaning of "major" is somewhat subjective, it should be clear once you detect it. If you are not sure, chance is that you haven't detected it.

2 Slow Trend

You are a data scientist working for a large supermarket chain. Lately, the vice president of customer insights has been discussing with the manager who handles merchandising support about what the supermarket can learn from social media. In particular, they want to know whether it is possible to detect emerging trends of food consumption from social media data so that the company can have an edge over its competitors in launching new products, and if so, how the company can do it at scale. You are tasked with this mission and need to deliver a report in one week.

In the project report, you need to first discuss several potential approaches to this problem before arguing for the particular method you will be using. Then, you should present your results. Along with this report, you also need to submit a Python code file that can replicate the results in your report. Your Python code file should contain comments to explain the role of each code block, and your Python code should be able to run with minimal human intervention.

Suppose your first name is *John* and your last name is *Smith*. You should name the project report as *john_smith.pdf*. The project report must be in PDF format. Similarly, your Python code should be named *john_smith.py*.

2.1 Project Background

Early detection of emerging food trends can translate into great business opportunities. Today, a lot of food-related discussions occur on social media platforms such as Twitter and Facebook. Thus, such social media content presents a potentially valuable and real-time source of intelligence that can be leveraged by retailers to better serve its customers. The purpose of this project is to explore this possibility using techniques discussed in the social media analytics course. Ultimately, we wish to help retailers *see* the rise and fall of certain categories of food *before* competitors do.

Despite its potentials, we also need to keep in mind the limitation of detecting emerging food trends using social media analytics. Many successful new product launches are attributed to identifying an unserved or under-served market segment. For example, the key to the success of Breyers Gelato Indulgence[2] is targeting the specific moment in which there is a married mom,

[2]Gelato Indulgences is a super-premium frozen treats priced 70% higher

kids have been put to bed after an active day, and then there's this moment to unwind, connect with her spouse and enjoy a little reward. Identifying the opportunity of enriching people's end-of-day experience is clearly crucial in this case. In fact, Breyers learnt about this from many consumer interviews they conducted. However, even if those consumers do post their own end-of-day stories on social media, it would require a highly intelligent machine to understand the unique moments from such posts and further suggest a non-existent new product that can better serve this niche market. In this case, the gap from information to innovation is simply too large for a machine of our age.

On the other hand, some successful new product launches can be inspired by trends that might be detected through social media analytics. For example, the story of Dole Chopped Salad Kits[3] is really about the insight that the trend on chopped salads is changing. Chopped salads have been around since the 1930s without receiving much attention from consumers or inspiration from chefs. In the early 2000s, in the world of restaurants, recipes and ingredient variety of chopped salads soared. Food companies and retailers can clearly benefit from learning about this trend before competitors do. If there are enough signals buried in the social media content about people's changing habits or new preferences of consuming salads, then it should be possible to detect this emerging trend through careful social media analytics. Indeed, according to Dole's vice president of marketing, CarrieAnn Arias, "Not only were popular mainstream restaurants such as California Pizza Kitchen and Cheesecake Factory putting more creativity into their salads, but consumers were gushing about the experiences. One of the things that struck us in our conversations with consumers was how much they loved chopped salads."

The two cases above might not be typical. Most likely, a detected emerging trend, whether from social media or from other information sources, should be reviewed by human experts in order to gain an early-mover advantage, not necessarily first-mover advantage. In other words, social media

than mainstream ice cream. It was launched in 2014 by Breyers. The brand won a 2016 Nielsen Breakthrough Innovation Award for this product launch. See `http://www.chicagonow.com/marketing-strategist/2016/12/beyond-mainstream-pricing-breyers-goes-super-premium-with-gelato-indulgences/` for a media report, and `https://www.youtube.com/watch?v=R1Q-T_raX8Y` for its commercial.

[3]Dole Chopped Salad Kits is also highlighted as a winner of the 2016 Nielsen Breakthrough Innovation Award.

analytics may be able to provide us with the dots, but it will ultimately be up to human to finally connect the dots. For example, if the supermarket can detect the rising popularity of a new ingredient because of certain health benefits, it might still take human ingenuity to design food or drink using such ingredient. Nevertheless, machine algorithms with the capability of constantly generating candidate emerging food trends can greatly inspire and augment the creativity of human experts in a more efficient way.

2.2 Data

Apparently, you need social media data related to food consumption for this task. Thanks to one of your TA, we have access[4] to over 4 million Facebook posts from 2011 to 2015, which will be the dataset for you to work on.

I have organized the data into multiple files with each containing the textual content of all posts in one month. For example, the file *fpost-2011-4.txt* contains all the posts from the month April 2011. Within each file, each line corresponds to one post.

Figure 1 plots the histogram of all the posts by month.

2.3 Method

You need to develop a scalable method to extract signals from the large corpus of Facebook posts so that early signs of food trends could emerge from your analysis. You are also welcome to collect additional data for your method, as long as you share the additional data for our evaluation of your method and you explain clearly how the additional data is collected in your report.

At a high level, the task of food trend detection using social media analytics can be broken down into two components: constructing time series of potential food trends, and detecting (abrupt) changes in those time series. In this project, we will only work on the first component and simply use visual inspection for the second component. More specifically, you may first create a monthly index for a hypothetical trend. Then, you can plot how such an

[4]We used a loosely constructed lexicon for terms that might be related to food to search related Facebook pages. Over seven thousand relevant Facebook pages are identified, from which we obtained the Facebook posts. Clearly, this is a very crude data collection process. Nevertheless, the resulting dataset contains a lot of posts that discuss various aspects of food preparation or consumption.

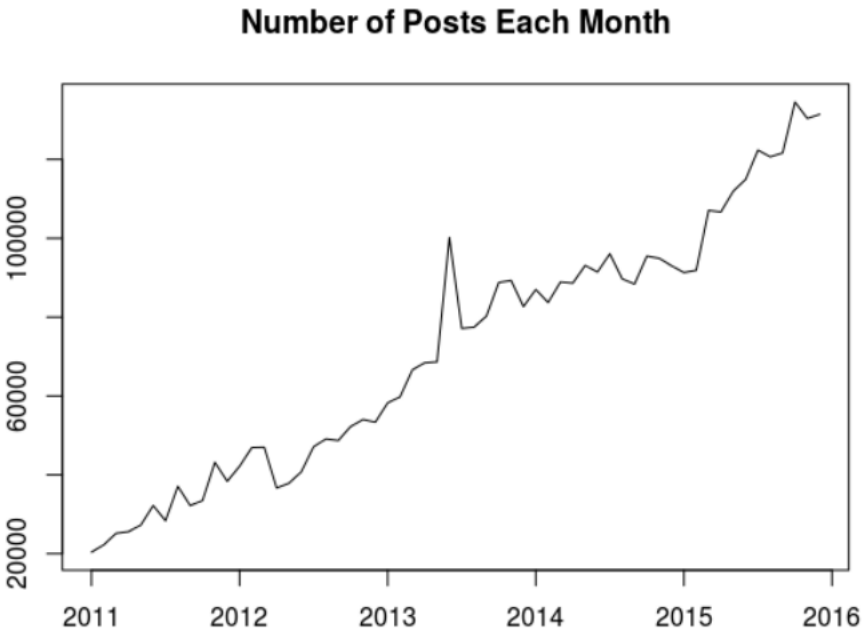

Figure 1: Number of food-related Facebook posts by month

index evolves over time. To facilitate the grading of this part, please report such plots in the PDF file (i.e., your project report). Of course, we should be able to reproduce these plots by running your Python code file.

2.4 Validation

To test the effectiveness of a proposed method, we first need some ground truth. In other words, when we evaluate a trend detection method, how do we know a food trend detected by this method actually reflects a true food trend in reality, and a non-existent food trend will not be falsely detected by the method?

One idea is to use seasonally consumed food to validate a method. For example, if a method cannot detect the spike of "pumpkin pie" around Thanksgiving, it is clearly failing the task.

By talking to people working in the food industry, I have identified some recent food trends that can serve as the ground truth to validate a method. I will provide the background of two of them: cauliflower rice, and vegetable noodle. **Your grade will mainly depend on whether, and how early**

your method can detect these food trends without too many false positive (i.e., non-existent trend incorrectly detected by your method).

There is also a practical question of how early a detection has to be for it to be useful. In other words, how early is early enough? But we will leave this question to the vice president.

Cauliflower Rice

Cauliflower rice is not actual rice. It is a grainy substance made by pulsing cauliflower florets in a food processor until they've broken down into tiny granules, and then lightly cooking the pieces in oil. Demand for cauliflower rice has been growing steadily over the past few years largely due to carbo-adverse consumers' desire for a healthy alternative to white rice and gluten-filled grains. Cauliflower rice is so popular at Trader Joe's that the grocery chain recently[5] began enforcing a two-bag limit per customer, rationing it to keep the item on shelf. This is also reflected in the spike of sales in cauliflower. In 2016, U.S. farmers sold $390 million worth of cauliflower[6], a big jump from the $239 million sales in 2012. According to Green Giant VP and general manager Jordan Greenberg, immediately after the company was acquired by manufacturer B&G in 2015, they expanded the cauliflower line to include three types of cauliflower rice, increasing the weekly amount of the vegetable harvested from 5 acres to 35 acres.

The popularity of cauliflower rice can partly be traced to recent trends of low-carb dietary. With less than one-eighth the calories of white or brown rice and about one-ninth the carbs, cauliflower rice, on the other hand, is rich in vitamins C, K, B6 and folate. One small head of cauliflower has over 125 mg of vitamin C, nearly twice as much as a medium orange. The rise of cauliflower rice fits into the recent trend of old-fashioned vegetables gaining new traction as more people gravitate toward plant-based foods.

Vegetable Noodle

Vegetable noodle, or veggie noodle, is a catch-all name for various spiralized vegetables that resemble the shape of noodle or pasta. The list of veggies one can spiralize is long, but zucchini, squash and cucumber are often used.

[5]This was reported in July, 2017. For more details, see `http://www.foodandwine.com/news/trader-joes-cauliflower-rice-rationing`.

[6]`http://time.com/4845148/cauliflower-rice-menu/`

Figure 2: Cauliflower Rice (left); Veggie Spiralizer (right)

Zucchini is probably the most widely used ingredient to make veggie noodles, probably because of its noodle-like texture once cooked. It's so popular that there is even a name for zucchini noodle: *zoodle*. Vegetable noodles pack lots of healthy benefits. For example, zucchini is extremely low in calories, is chock-full of antioxidants, and is also a great source of potassium.

Cutting vegetables up into tiny little strips was first mentioned in print as early as in the 18th century possibly in a French culinary book. At the time, this procedure was called *julienning*. According to Simone Baroke, an analyst at Euromonitor International specializing in the global health and wellness and fresh food markets, the vegetable spiralizing mania was brewing for some time before it reached critical mass and hit the global mainstream in early 2015, when cookery sections of major publications suddenly started raving about what a nifty little idea it was. In January 2015, *Vogue* (American edition) featured an article entitled "why you need a spiralizer" in its Arts and Lifestyle section, while in the UK, "Spiralizing: How to get the best results" appeared on the BBC's Good Food blog around the same time[7].

[7] `http://blog.euromonitor.com/2015/08/spiralized-vegetables-succeed-as-ersatz-pasta.html`