2011-2015 Food Trend Analysis Report

-Based on Facebook Post Data

MSBA Guanyu Zhang

1. Summary

This report aims at delivering methods of detecting emerging trends of certain food given time series Facebook post data. Generally, emerging trends is closely related to the frequency of words people post on Facebook. So the method is typically extracting the number of word frequency related to food from massive Facebook text data and plot the frequency changes to detect when the trends are emerging.

2. Method

When extracting word frequency of the Facebook, there are two major difficulty that should be overcame.

1. Extracting frequency of words only related to food.
2. Extracting phrases with multiple words that represent food.

In this case, the simple solution is to use the 'Ingredient.txt' as the food dictionary to apply when create the document term matrix, and use the both the unigram tokenizer and bigram tokenizer (since most phrases consist of two words).

However, there are usually the cases that some food or new food is not in the Ingredient dictionary thus will not be extract into the document term matrix. So we have to make some fine adjustment to the process. In order to intuitively logical and computationally user-friendly, in other words make sense and save computing power. Here I propose two ways

One way to do it is to implement the ingredient dictionary, for example, robustly add new food terms to the ingredient dictionary so that the document term matrix will include those terms.

Another way I do is to analyze the correlation of the frequency of food terms to infer there is some food or combination of food is trending.

In this report, I will do both.

3. Cauliflower Rice Trend

Firstly, I plot the frequency change of 'cauliflower' and 'rice' over time. Then I added the 'cauliflower rice' to the ingredient dictionary to enrich the document term matrix. Below are the time series plots based on the document term matrix.

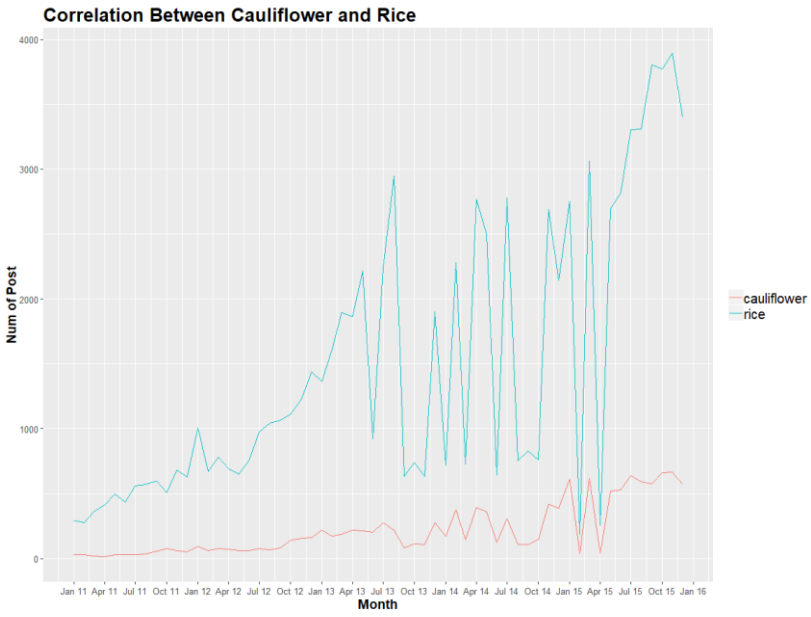

Fig.1

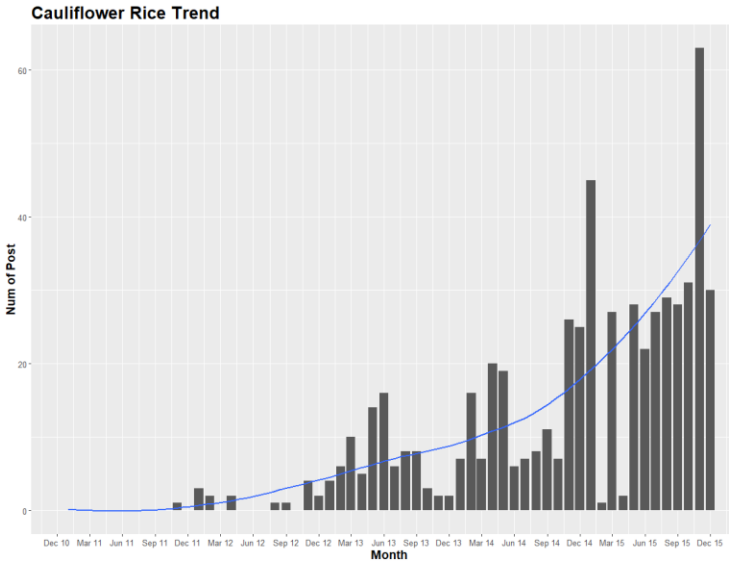

Fig.2

From Figure1 and Figure2, we can see a clear increase in mentioned post over time. **Around June 2013**, there is a small spike of both 'cauliflower' and 'rice', also the term 'cauliflower rice' start to appear more times. And **in December 2014 and November 2015,** a large spike of both 'cauliflower' and 'rice' is detected, and number of appearance of the term 'cauliflower rice ' increased dramatically.

To conclude, we can see the cauliflower rice first started to trend **around June 2013** and the trend grow bigger around **December 2014 and November 2015.**

4. Vegetable Noodle Trend

One major concern about detecting the vegetable noodle trend is that, vegetable noodle may be named under different terms, for example, 'zoodle' or 'zucchini noodle'. So my approach is to get the result from its major ingredients--'zucchini'.

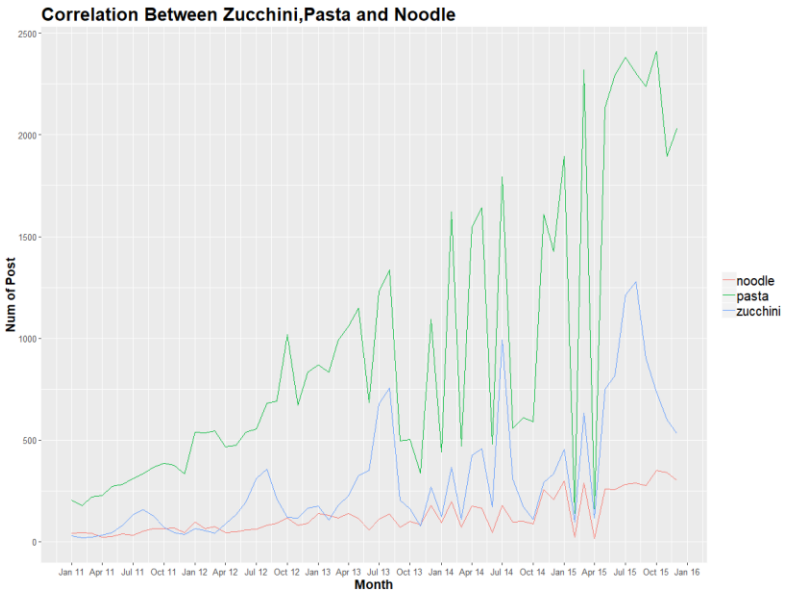

Fig.3

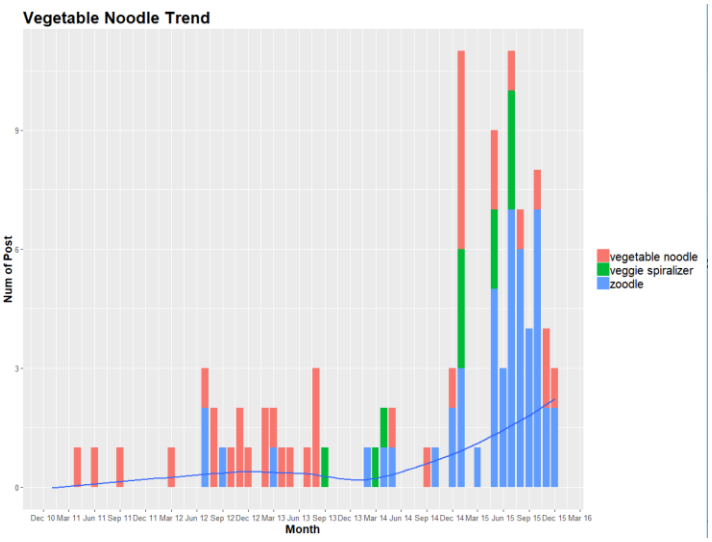

Fig.4

We can see from Figure.4 that the number of post regarding related terms is relatively small. So the trend is not so obvious to detect. While in Figure.3, we can see clearly a spike of 'zucchini', 'pasta' and 'noodle' emerging **around January 2015, a**nd a large spike **around July 2015** which aligns with the result of Figure.4.

Hence, the trend of vegetable noodles would probably emerge around **January 2015,** and grow into more popular trend around **July 2015**.

5. Validation

To validate the method, I used pumpkin pie as the validation object. It is widely known that pumpkin pie is in high demand around Thanksgiving every year (October to November). So applying our method to detect the seasonal effect of pumpkin pie would validate the effectiveness of our method.

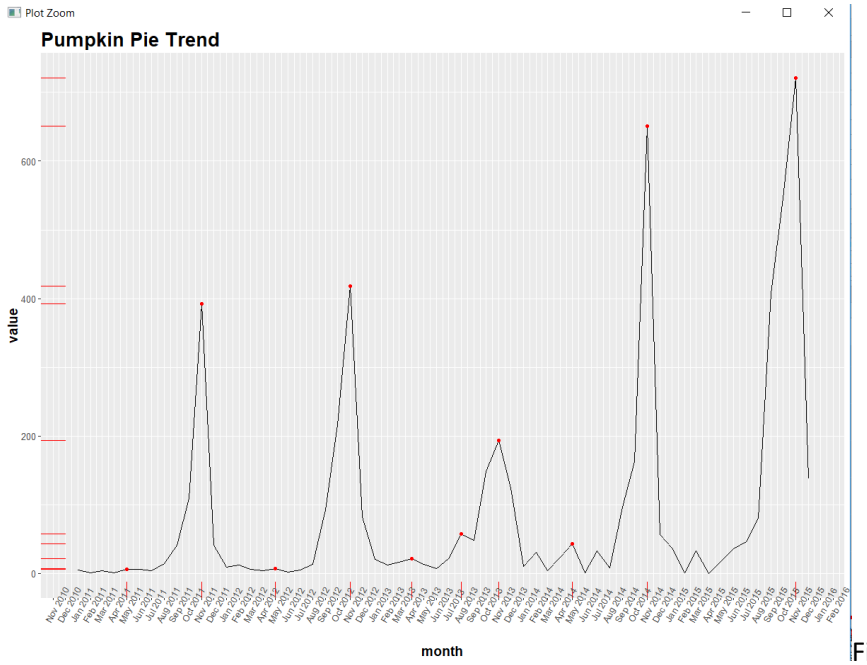

Fig.5

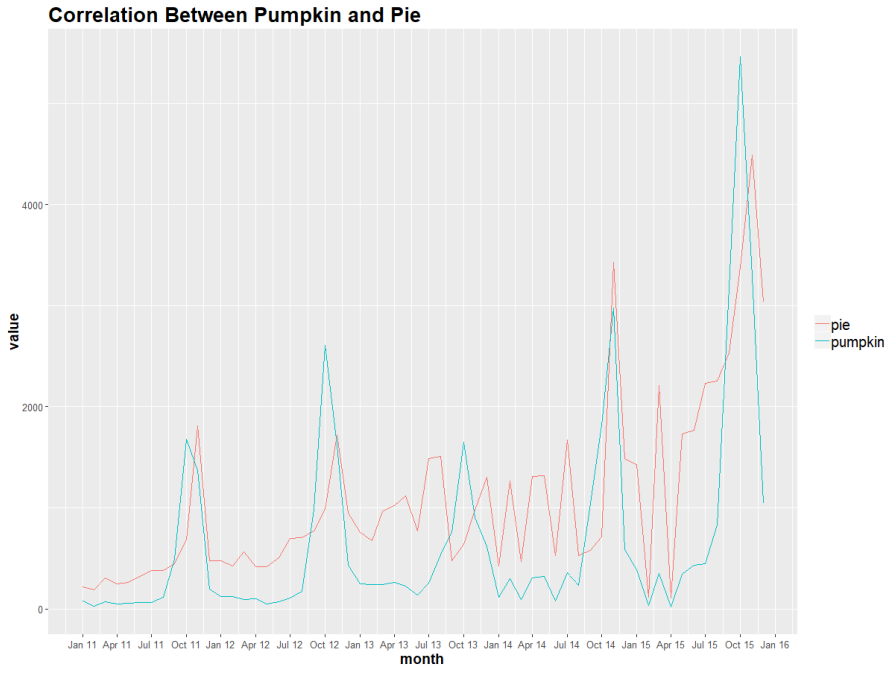

Fig.6

As can be seen from Figure.5 and Figure.6 that, obviously pumpkin pie has five huge spikes, one every year in November. It aligns with the ground truth that people would post more about the pumpkin pie.

6. Comment

Major usage of detect food trend is to help companies to produce favorable food one step ahead of its competitors. Thus earlier detection will generate a competitive advantage. By analyzing past time post of food terms 'cauliflower rice', we can see a trend line over time so in long term we can see if the food is becoming popular and whether the company should start to produce more products.

On the other hand, in short term by looking at the correlation of food terms, for example, 'zucchini' and 'pasta'. If they suddenly start to peak at the same time. We can infer that a certain combination of food is becoming popular. Thus, the company can consider introduce new food or combo products.

7. Future implementation

There are two major implementation I would propose to adjust the method.

First, use post ratio, as in the term posts percentage of all posts in the same period, instead of the absolute number of the post. So food trending would be more accurately detected, taken out the possibility that Facebook is more popularly used to post food.

Second, if we have to blind-guess a new food trend, the new food term may not be included in the ingredient list and we wouldn't know its name either before it starts to greatly trend. So we have to try a more complicated and time-consuming approach. We can PosTag all Facebook posts and extract all noun words in the posts (both unigram and bigram). By looking at which noun(s) as well as its related nouns are drastically appearing more times, we can infer the food related to the noun is starting to trend. Basically, this method is ideal but in practice we will more rely on intuitions about consumers' behaviors and nutrition to select food word list that is potentially emerging and its related food terms.