Final Report for Frontiers in GIS

Spatial and temporal co-location patterns of gas prices in Columbus, Ohio

Jialin Li

The Ohio State University

Spatial and temporal co-location patterns of gas prices in Columbus, Ohio

**Introduction**

Gas prices vary significantly among stations across space and through time. It can significantly influence people’s lives, because the role of automobiles is becoming more and more important in people’s daily life (Tanguay & Gingras, 2011). And the expenditure on gasoline has steadily increased to be one of the largest portions in household expenditures in most of the developed countries around the world (Yeimin Chung & Hojeong Park, 2014). Research about the gas price distribution can be useful for both customers and government. For the reason that, there may be lots of interesting spatial and temporal patterns and rules among the gas prices of stations. This essay is aimed at understanding the spatial and temporal patterns of gas prices and discovering useful rules of the gas price distribution around Columbus, Ohio, using data collected by a member-based website, through spatial co-location pattern mining method based on Apriori algorithm and density estimation.

Most of the current research about gas price focuses on investigating the influencing factors for gas prices and determining whether the clustering degree of gas prices is relatively high. Akio Kusakabe and Jinhwan Oh(2011) used Moran’s I to examine the spatial autocorrelation in Japan and apply spatial regression analyses to find out factors influencing gas prices. They found that gas prices throughout Japan exhibit spatially dependent patterns and that regional conditions, including weather, are the main determinants influencing gas prices in Japan. Kihm et al. (2014) utilize Quantile Regression method to identify what factors influence retail gasoline price based on crude oil price. And their analysis indicates that several factors, including the absence of nearby competitors and regional market concentration, play a significant role in mediating the oil price on the retail gas price. Ryan et al. (2015) investigated gas price structure in North Ireland by spatial statistics and regression methods. The results present significant spatial variations among the gas price at local level. And they also found that high prices tend to occur in remote rural areas and urban areas with high rates of gas heating. Prices are closely related to a complex set of interacting factors, such as market structures, supply costs and market competitiveness.

Perhaps, there are some spatial and temporal patterns and rules existing in gas price data of Columbus. For example, there may be a low-price gas station in the neighborhood of a high-price gas station. And there is no such research about gas price in Columbus to obtain the patterns and rules. This study will contribute to understand the spatial and temporal pattern and discover valuable rules of gas prices in Columbus, which will provide useful information for government to make gas policies and for customers to choose cheap gas stations. The paper will use the following key methods to study the gas price distribution:

In this paper, we intend to understand the spatial and temporal pattern and discover some useful rules of the gas prices in Columbus. However, there hasn’t been any relevant research for Columbus till now. In this paper, we first introduce the problems solved in our study, and then explain the methodology. Next, the experiment results are shown, and discussion for the results are conducted. At the end, we make the conclusion for this study.

**Methodology**

In this study, we first use Kernel Density Estimation (Peter Hall, Jeff Racine, & Qi Li, 2004) to evaluate the density of high and low price gas station in the past two years. Then, in order to discover spatial and temporal co-location patterns in the past two years high and low gas prices, Spatial co-location mining method based on Apriori algorithm is utilized in the study. Additionally, Fourier transform is used to detect periodicity in the change of spatial co-location pattern (Yan Huang, Shashi Shekhar, & Hui Xiong, 2004) of high and low gas prices.

Kernel Density Estimation

Kernel Density Estimation(KDE) is able to generate overall density of a variable, considering weighted local density at each observation point (D.W. Scott, 2008). In KDE, a weight which decreases from the target point is assigned to observation points, instead of giving equal weight to every observation points in the neighborhood. Kernel density estimator is like this:

where is the target point and is observation point. And h, which is used to control the neighborhood size of , is the smoothing parameter. The function K is called the Kernel and it can control the weight given to the observations in the neighborhood of . A kernel with subscript is called *scaled kernel* and defined as . In our study, Gaussian Kernel is used. The formula of Gaussian Kernel is represented below:

Where is the distance from target point to observation point.

The estimation result of Kernel Density Estimation is then used to draw a density map of high and low price gas stations respectively.

Spatial co-location pattern mining

Apriori algorithm is a classical method in data mining. Here, we use Apriori algorithm to detect the spatial co-location patterns in the high and low gas price data in the past two years. Spatial co-location pattern is the pattern that subsets of features frequently located together(Huang). There are two important outputting value in Apriori algorithm, Support degree and Confidence degree. In spatial co-location pattern mining, the frequency that there are at least one feature B in the neighborhood of feature A is represented by , then the support degree of is as follows:

Where represents the total number of features in dataset.

Another outputting value, confidence degree is represented below:

Where is the number of feature A. Confidence degree indicates the extent to which we can think that the corresponding spatial pattern is true.

In this paper, our gas price data is hourly data, including two categories, 15 highest gas prices and 15 lowest gas prices in a time slice. In this case, support degree is not that crucial for detecting spatial co-location patterns, because support degree is around half of the value of confidence degree. In our study, we aim at detecting the four possible co-location patterns, namely, high to high pattern, high to low pattern, low to low pattern and low to high pattern.

Fourier Transform

Fourier transform is a way to transfer the time-series data into frequency domain, in which periodicity of time-series can be detected in a straightforward way (Bracewell, 1986). Fourier transform has been commonly used in periodicity detection in time series (Elfeky & Aref, 2005). The fast Fourier transform(FFT) reduces the time complexity of the computation process and can be used for large sizes of databases (Knuth, 1981). Therefore, in this paper, FFT is conducted to test whether there is obvious periodicity in the time series of confidence degree for different spatial co-location patterns using fft function in numpy package of Python.

**Experiment and discussion**

In this study, we use data collected by a member-based website to understand the spatial and temporal patterns of gas prices around Columbus, Ohio. It includes hourly prices for the regular gasoline for the past two years, and at each hour the 15 stations with the highest prices and 15 with lowest are stored. The locations of these gas stations in our database is shown in Fig. 1 below.

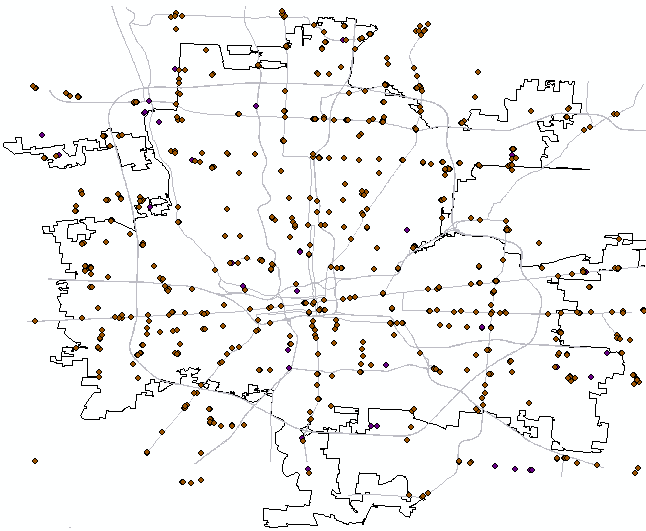
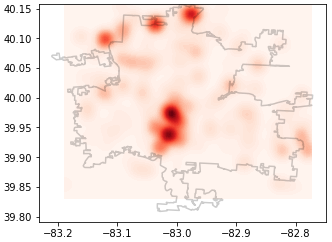


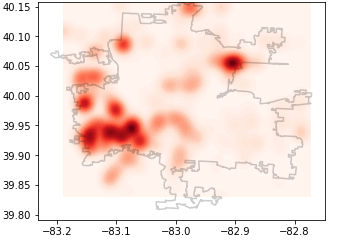
Fig 1. Gas station locations in gas price database

In our experiment, we first evaluate the density of overall area of Columbus using Kernel Density Estimation, and then draw the density maps for high gas prices and low gas prices respectively in different scales. Then the confidence degree of the four spatial co-location patterns in different scales during the past two years are calculated and then the time series of confidence degree are transformed into frequency domain using Fourier Transform to detect periodicity in the time series.

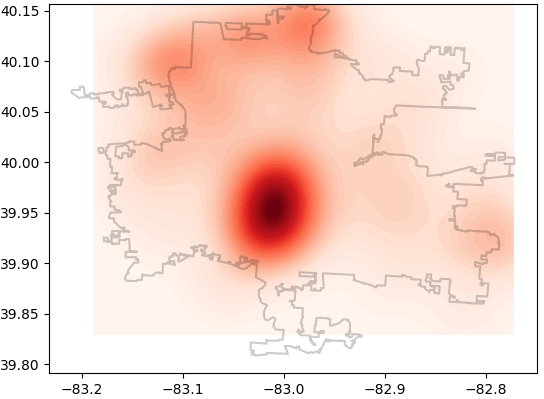
Density maps of high and low price gas stations are applied to show the experiment results of KDE for different neighborhood sizes in Fig. 1 below. These figure is from the result of Kernel Density Estimation based on the occurrences of high and low gas price in the past two years.



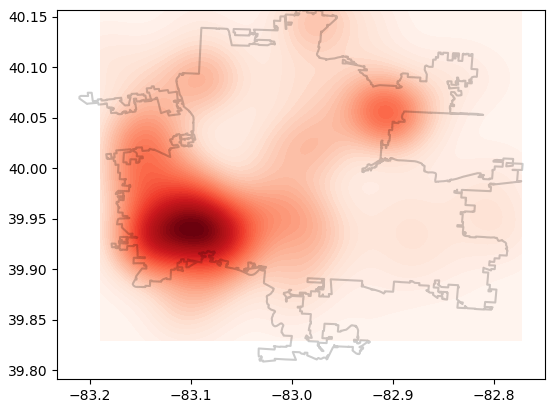
1. Density map of high price gas stations (d=0.01 degree)



1. Density map of low price gas stations (d=0.01 degree)



1. Density map of high price gas stations (d=0.025 degree)



1. Density map of low price gas stations (d=0.025 degree)

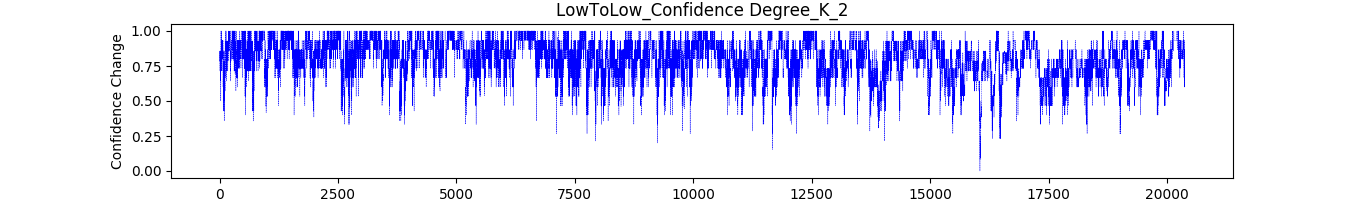
Fig. 1 Density maps of high and low price gas stations

In Fig. 1, sub-figure (a) and (c) shows the density map of high price gas stations in the last two years around Columbus area, and sub-figure (b) and (d) represents the density map of low price gas stations. Sub-figure (a) and (b) are the density map for the case that neighborhood size of target point is 0.01 degree, and sub-figure (c) and (d) are for the case that neighborhood size is 0.025 degree. The base map is the outline of Columbus area. In these figures, x axis is the longitude of places and y axis indicates the latitude. And the extent of red color on the maps means the density of high or low gas price stations. The darker the red color is, the denser the gas price stations are. From the density maps, we can discover the rules of high and low price gas stations’ distribution in the past two years.

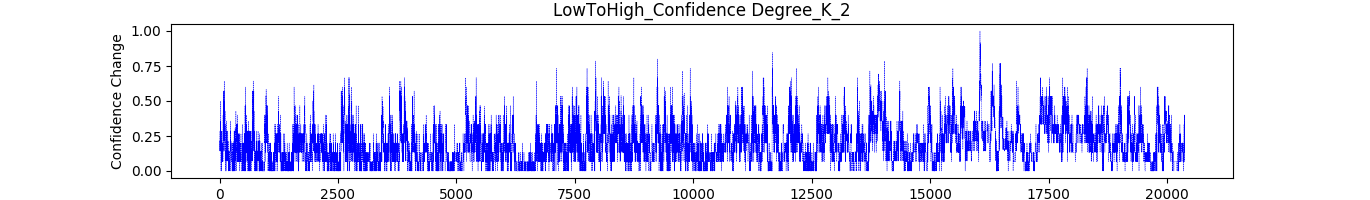
Density maps in different scales show similar density of high and low price gas stations. In general, from sub-figures (c) and (d), the center of Columbus is the most common area of high gas prices, while southwest of Columbus is the area where gas prices tend to be low.

In the experiments, the density maps in sub-figure (a) and (b) where neighborhood distance is 0.01 degree show the specific areas for high or low gas prices in the city. From sub-figure (a) in Fig. 2, we can find that the center and north of Columbus tend to have high price gas stations. The center of Columbus is the downtown area and north of Columbus is the Polaris Fashion Place, which is one of the biggest shopping malls in Columbus. These two places are both population dense areas, which are likely to have relatively expensive costs for products. This result is reasonable based on our common sense about the gas price distribution. And from sub-figure (b) in Fig.2, the places which are most likely to have low price gas stations include the southwest and northeast of Columbus.

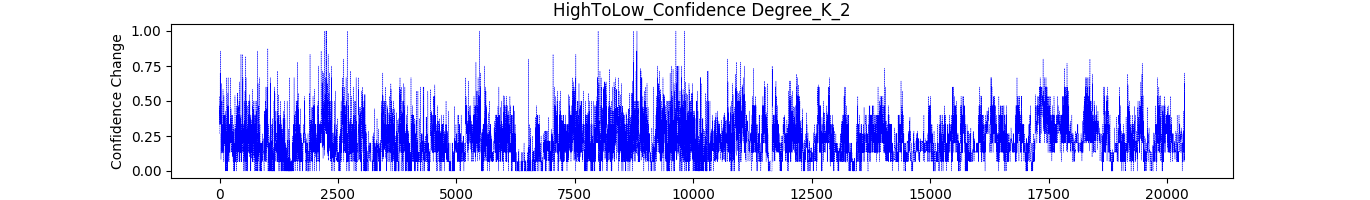
The figures in Fig. 2 represent the confidence degree of four spatial co-location patterns in different scales during the past two years. In each of the figures, x axis means the time slices in the past two years and y axis means the confidence degree of the co-location patterns in each time slices. And each sub-figure represents a specific spatial pattern. In sub-figure (a), low to low pattern means that there will be no less than a low-price gas stations in the neighborhood of a low-price gas station. Similarly, sub-figures (b)(c)(d) show the confidence degree for Low to High, High to Low and High to High patterns. The neighborhood size is defined as K-nearest neighbors, and K equals 1 in Fig. 2, which means that only the nearest gas station will be considered as the neighbor of a target gas station.



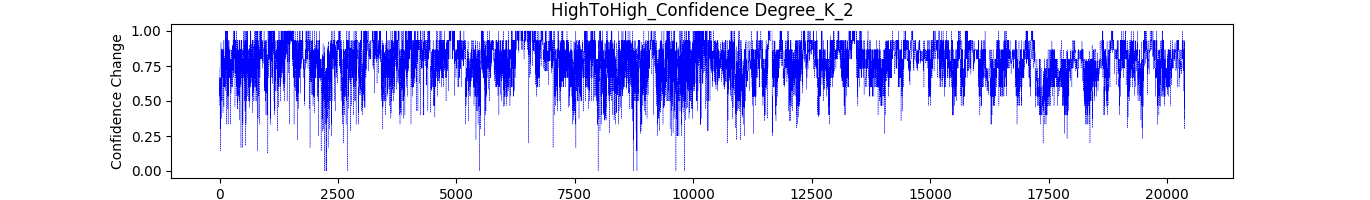
1. Low to low pattern confidence degree



1. Low to High pattern confidence degree



1. High to Low pattern confidence degree

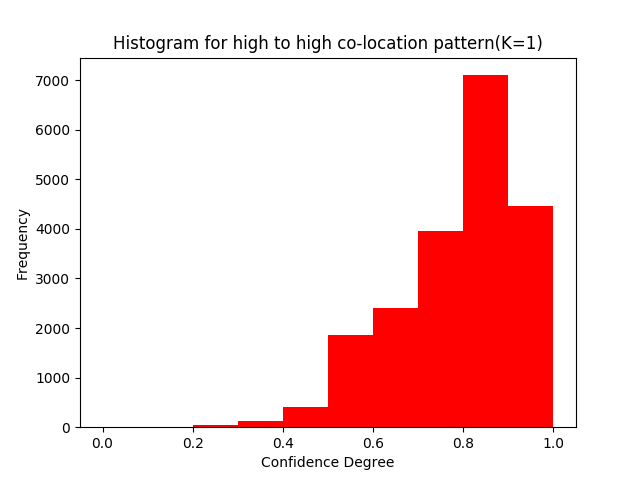


1. High to High pattern confidence degree

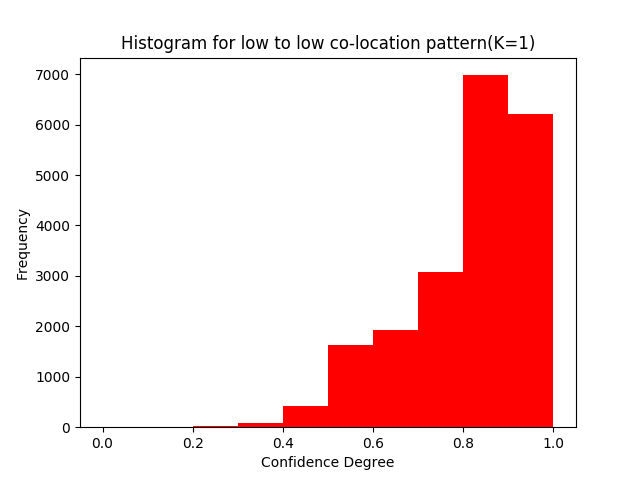
Fig. 2 Confidence degree of different spatial patterns in the past two years (K=1)

In our study, we classified the gas stations into two categories, namely, high-price gas stations and low-price gas stations. So, if the gas stations are distributed randomly, the confidence degree of each spatial patterns should be around 0.5 in the case where neighborhood of co-location only includes the nearest point. However, from Fig. 2, the confidence degrees of low to low pattern and high to high pattern in the past two years are more than 0.5 in most of the time slices, and those of low to high pattern and high to low pattern are less than 0.5 in most of the time slices.

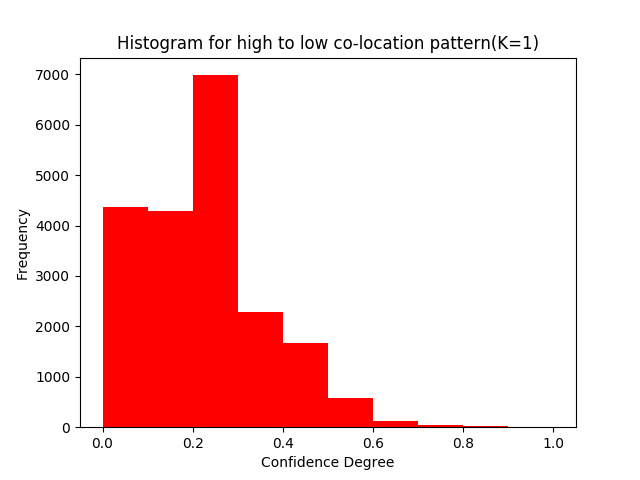
In order to compare the real confidence degrees of spatial co-location patterns with the expected ones in a more straightforward way, histograms of confidence degrees are drawn below.



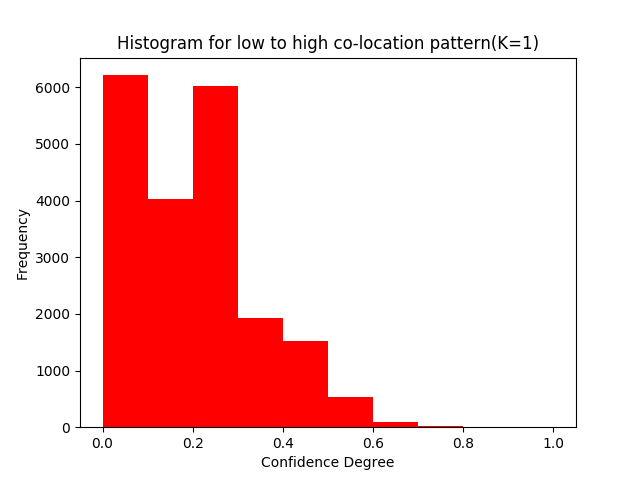
1. High to High co-location pattern



1. Low to Low co-location pattern



1. High to Low co-location pattern

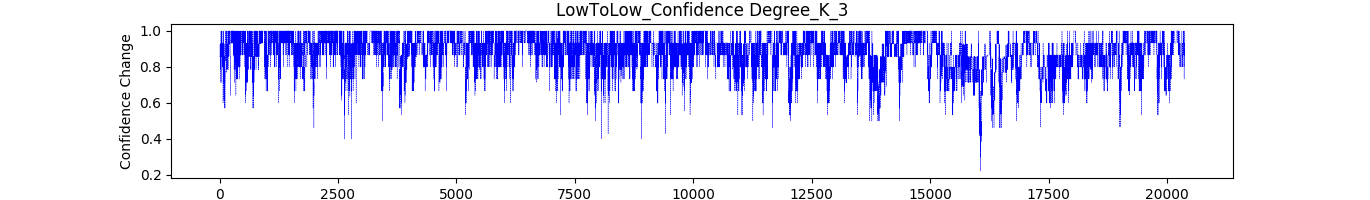


1. Low to High co-location pattern

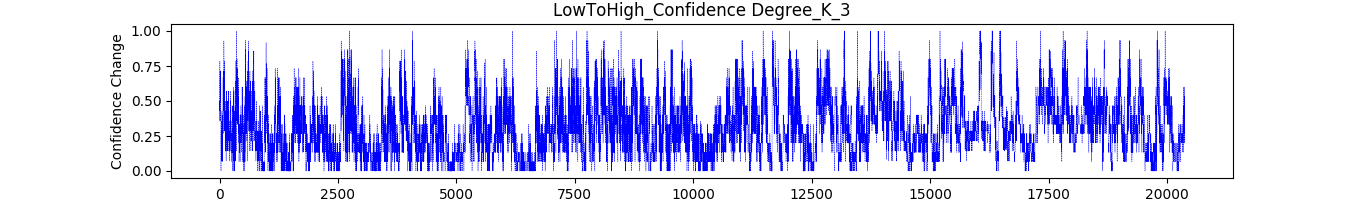
Fig. 3 Histograms for confidence degrees of the four co-location patterns (K=1)

In Fig. 3, we can find that for most of the time slices, High to High and Low to Low co-location patterns are significantly more than 0.5. On the contrary, High to Low and Low to High co-location patterns are much less than 0.5. This result means that low to low pattern and high to high patterns were more likely to happen in the past two years, while high to low and low to high patterns happened scarcely.

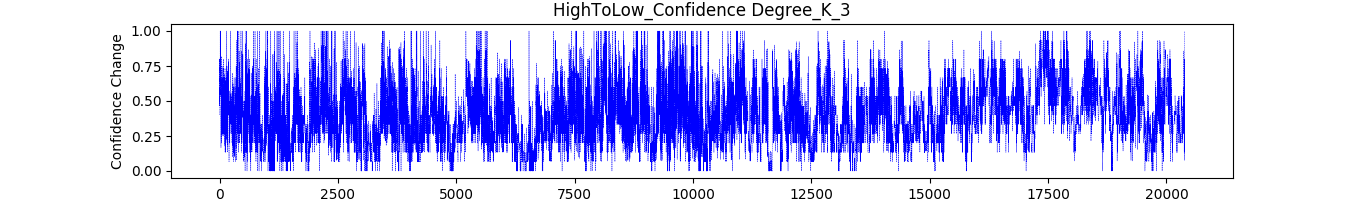
Moreover, in different scales of neighborhood to detect co-location patterns, the experiment results are similar with K=1. The confidence degrees of four spatial co-location patterns for K=2 and K=3 are shown in Fig. 4 and Fig. 5 respectively.



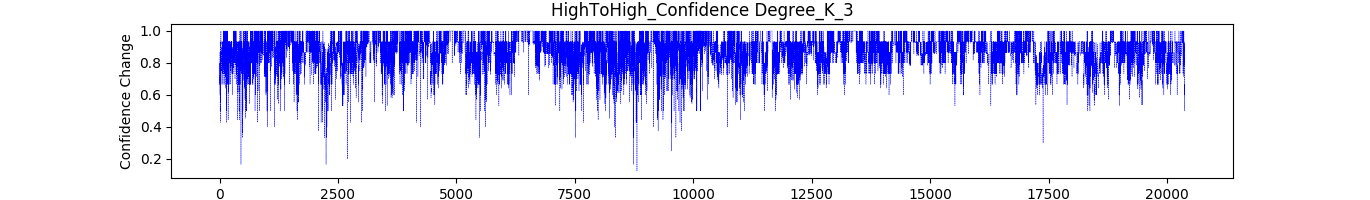
1. Low to low pattern confidence degree



1. Low to High pattern confidence degree

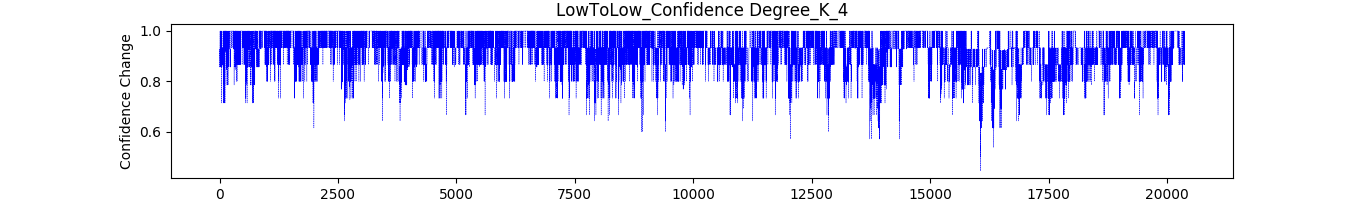


1. High to Low pattern confidence degree

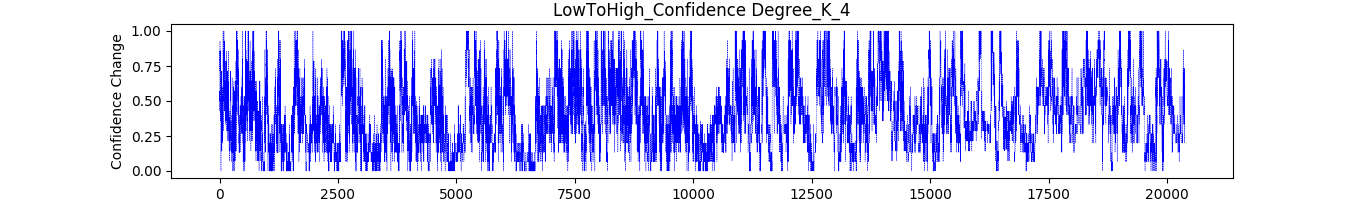


1. High to High pattern confidence degree

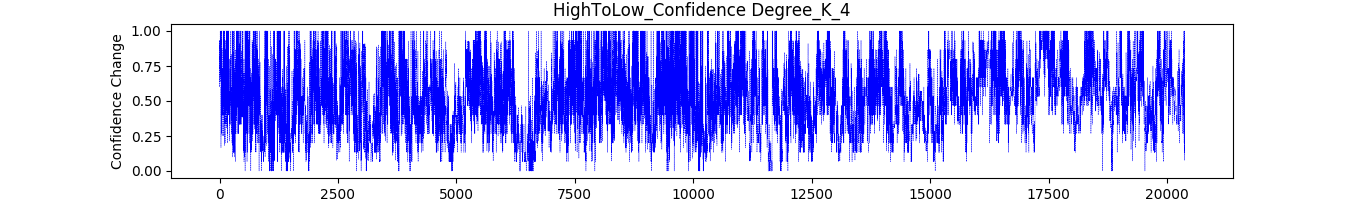
Fig. 4 Confidence degree of different spatial patterns in the past two years (K=2)



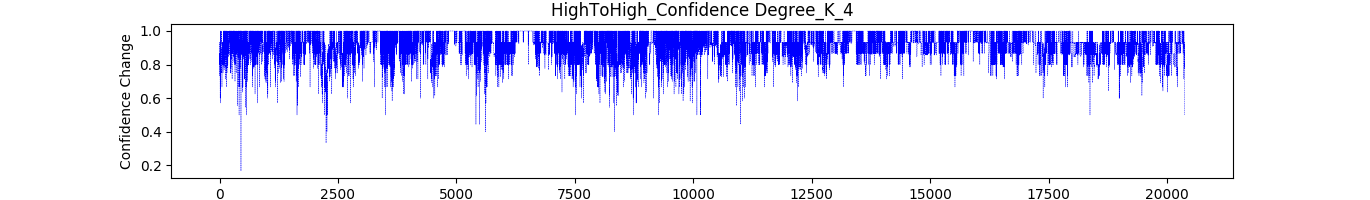
1. Low to low pattern confidence degree



1. Low to High pattern confidence degree



1. High to Low pattern confidence degree

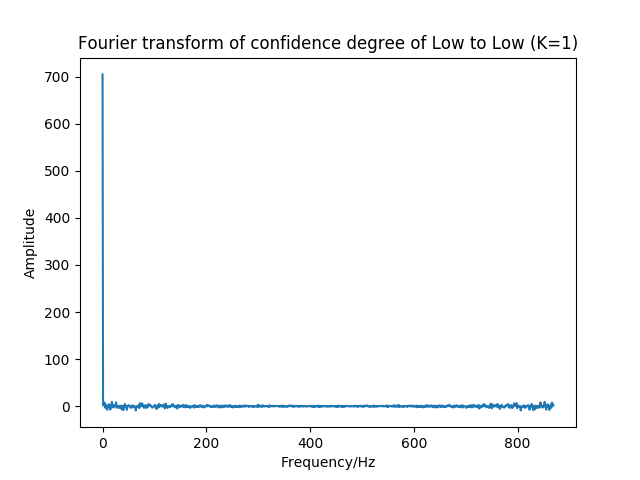


1. High to High pattern confidence degree

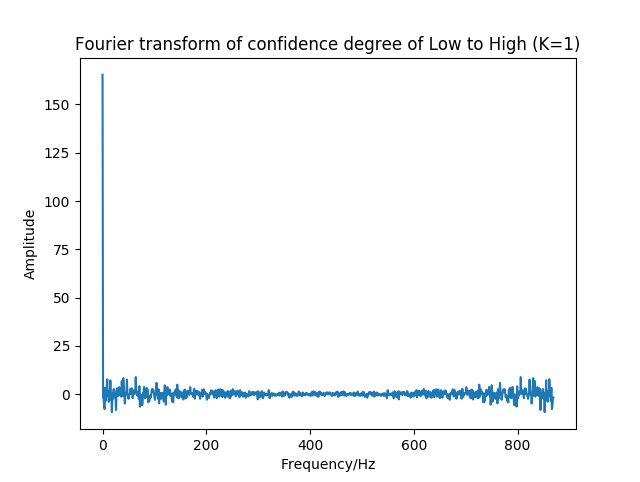
Fig. 5 Confidence degree of different spatial patterns in the past two years (K=3)

Therefore, the nearest gas station of a high or low price gas station tends to be the same category with it. That is to say, high-price gas stations are more likely to be near high price gas stations, and low-price gas stations are more likely to be closed to low price gas stations.

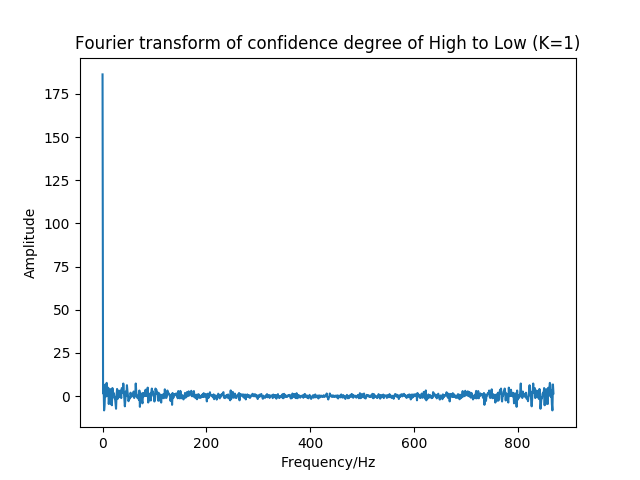
The experiment results of Fourier transform are represented in Fig. 6 below. These figures are for the neighborhood size of K=1.



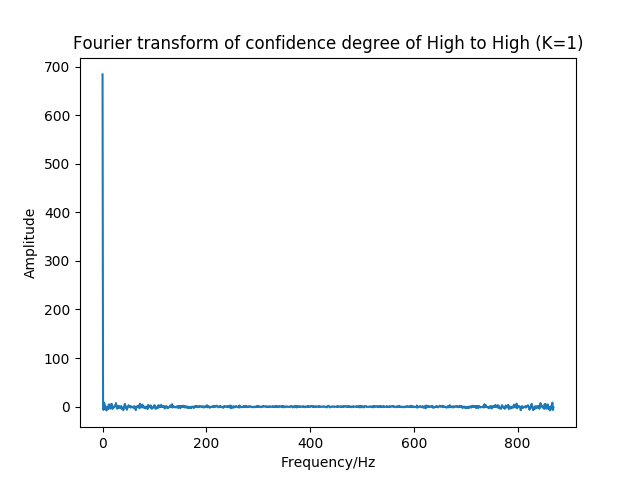
1. Confidence degree of low to low pattern



1. Confidence degree of low to high pattern



1. Confidence degree of high to low pattern



1. Confidence degree of high to high pattern

Fig. 6 Confidence degree of co-location patterns in frequency domain

In Fig.6, confidence degree of the four co-location patterns in frequency domain show similar figures. There is only significant pulse on 0 Hz in each figure. And the small pulses around Amplitude 0 are noises in the results. The pulse on 0 Hz means that the time series are mainly composed by a wave whose frequency is 0 Hz. So the period of the wave should be infinite large, which indicates that no obvious periodicity exists in the time series of confidence degree of co-location patterns.

Additionally, the results of Fourier transform for the cases where K equals 2 and 3 are similar with the results of K=1. Therefore, conclusion can be made that there is no significant periodicity in confidence degree of co-location patterns in the past two years.

**Conclusion**

In summary, based on the experiment result, we conclude that in the four different spatial co-location patterns, high to high pattern and low to low pattern are significantly frequent in the past two years, and high to low pattern and low to high pattern are significantly scarce. These patterns indicate that there will likely be a same type (high price or low price) of gas stations in the neighborhood of a gas station. The closest gas station of a high price gas station tends to be a high price gas station, while the nearest gas station of a low price gas station is likely to be a low price gas station. This means that, if a customer in a high price gas station wants to find a low price gas stations, he or she should drive relatively far away. Besides, there is no obvious periodicity in the time series of confidence degree of co-location patterns. Another implication of the data is that the center and north areas of Columbus tend to have high price gas stations, while the southwest and northeast areas of Columbus are likely to hold low price gas stations. This rule can be used by customers when they want to find low price gas stations and avoid high price gas stations.

Reference List

Agrawal, R., & Srikant, R. (1998). Fast algorithms for mining association

rules. *Readings in database systems (3rd ed.)*. Morgan Kaufmann Publishers Inc.

Bracewell, R. N., & Bracewell, R. N. (1986). *The Fourier transform and its*

*applications (Vol. 31999)*. New York: McGraw-Hill.

Chung, Y., & Park, H. (2014). Analysis of spatial interaction effect of retail

gasoline price in Seoul: a spatial econometric approach. *Korea and the World Economy,* 15(2),209-241.

Elfeky, M. G., Aref, W. G., & Elmagarmid, A. K. (2005). Periodicity detection in time

series databases. *IEEE Transactions on Knowledge and Data Engineering*, 17(7), 875-887.

Hall, P., Racine, J., & Li, Q. (2004). Cross-validation and the estimation of

conditional probability densities. *Journal of the American Statistical Association,* *99*(468), 1015-1026.

Huang, Y., Shekhar, S., & Xiong, H. (2004). Discovering colocation patterns from

spatial data sets: a general approach. *IEEE Transactions on Knowledge & Data Engineering,* *16*(12), 1472-1485.

Kihm, A., Ritter, N., & Vance, C. (2016). Is the German retail gas market

competitive? *Land Economics*, 92(4), 718-736.

Knuth, D.E. (1981). The art of computer programming. 2, seminumerical

algorithms. *Atmospheric Chemistry and Physics*, 2.

Kusakabe, A., & Oh, J. (2010). Spatial analysis of the Japanese gas industry. *Review*

*of Urban & Regional Development Studies*, *22*, 143–164.

Scott, D. W. (2008). *Multivariate Density Estimation: Theory, Practice, and*

*Visualization*.

Tanguay, G. A., & Gingras, I. (2011). Gas prices variations and urban sprawl: an

empirical analysis of the 12 largest canadian metropolitan areas. *Cirano Working Papers,* *44*(7), 1728-1743.

Walker, R., McKenzie, P., Liddell, C., & Morris, C. (2015). Spatial analysis of

residential fuel prices: Local variations in the price of heating oil in Northern Ireland. *Applied Geography*, 63(2015), 369-379.