

Original Research Paper

# Newcastle Traffic Classification Using Clustering Algorithms

<sup>1</sup>Hamad B Matar, <sup>1</sup>Talal Almutairi and <sup>2</sup>Nayef Z Al-Mutairi

<sup>1</sup>Department of Civil Engineering, College of Technological Studies (PAAET), Kuwait

<sup>2</sup>Department of Civil Engineering, College of Engineering & Petroleum, Kuwait University, Kuwait

## Article history

Received: 13-12-2019

Revised: 20-02-2020

Accepted: 02-04-2020

Corresponding Author:

Nayef Z Al-Mutairi

Department of Civil  
Engineering, College of  
Engineering and Petroleum,  
Kuwait University, Kuwait

Email: nayef.almutairiku.edu.kw

**Abstract:** The urban road traffic network evolution is complex and varies depend on road type, zoning types and social activities. Typical traffic pattern variation of road network could be examined by considering the daily human travel activities. Thus, factor and cluster analysis is carried out. This paper is a comparative analysis of various Data Mining clustering methods for the grouping of roads based on traffic profile. The analysis was carried out using data available from 45 Automatic Traffic Recorder (ATR) sites in Newcastle, UK. Factor and cluster analysis were applied on the road traffic data so that roads could be classified, allowing diurnal traffic profiles to be assigned a group to roads with similar attributes. These groups could be classify based on road traffic characteristics. Five road classifications were found.

**Keywords:** Data Mining, Traffic, Newcastle, Clustering

## Introduction

Due to the growth of vehicle ownership and the road traffic demand exceeding the road network capacity in urban area, the traffic congestion causes a longer traveling time, more pollution emissions and higher accident risk. Nair *et al.*, (2019) stated “the congestion index of different cities is strongly related to emissions, population density and GDP per capita”. The traditional development of road network in the cities, such as new roads or widening roads construction, is not sustainable solution due to limited resources. The road network can be manage effectively by using Intelligent Transportation Systems (ITSs), such as Advanced Traffic Management Systems (ATMSs), to reduce travel time of the road users (Lee *et al.*, 2011; Zhang *et al.*, 2011). The morning and evening traffic peak hours caused by traffic to and from workplace or school are recurrent traffic pattern due to repetitive daily traveling activities (Francesc, 2012), There are numerous monitoring systems available to conduct traffic measurement surveys. These are divided into intrusive and non-intrusive sensors. Intrusive sensors such as induction tubes are positioned in or on the road surface to monitor traffic flow (Heidemann *et al.*, 2008).

The non-intrusive sensors are positioned in a way which does not interfere with road traffic, either during operation or installation as a camera (Heidemann *et al.*, 2008). Road traffic monitoring is classified according to manual, semi-automatic and automatic methods. The

manual method uses sheets of paper, mechanical or electronic counting boards and have typical manual count intervals of 5, 10 or 15 min, which would be conducted over a period of less than 24 h (Smith and McIntyre, 2002). Vehicle type, turning movements, travel direction and vehicle occupancy are some of the types of data collected during a manual count (Smith and McIntyre, 2002). The semi-automatic method uses manual and pneumatic tubes for the traffic count. The pneumatic tube is used to count traffic automatically by recording the air pulse on the tube, due to vehicle axle passing over the detected tube (Mimbela and Klein, 2000). Then a manual traffic count is conducted for a short period of time, which focuses on the type of vehicle. The vehicle type and axle number ratio are calculated using the manual traffic count data. These ratios from the manual and axle count of the pneumatic tube will be used to extrapolate the traffic volume. The inductive loop system is one of the automatic methods. This system records the electrical current pulse on a detection wire set in the road and is caused by the electromagnetic field as the vehicle [metal] passes over it (Ritchie *et al.*, 2002; Lin *et al.*, 2004).

The general shapes for the daily traffic profiles have been analysed by a number of researchers (Festin, 1996; DfT, 2001; US DoT, 2001; Chrobok *et al.*, 2004) and mainly focused on different road and vehicle types. These researchers classified the days into working and weekend days. The daily traffic profile for working days typically illustrates morning and peak periods and

between peak periods. Based on the results from one of the UK studies, the typical value for the peak hour ranges from 8% to 12% of the total daily traffic volume (Taylor *et al.*, 1996). Traffic volumes on weekend days are typically lower than traffic volumes on working days. Researchers investigated the difference in traffic volume between week days and weekends and concluded that there was no significant difference between the days of the week during week days, although there was a significant difference between Saturdays and Sundays (Stathopoulos and Karlaftis, 2001).

The advance of road traffic sensing technologies led to collect traffic data from multiple sensors, such as video cameras, inductive loop detectors and mobile phones (Leduc, 2008). A data mining technical is applied to the traffic data extract useful information for urban traffic management (Elhenawy *et al.*, 2014). Weijermars and Van Berkum, (2005) illustrate the advantage of using traffic patterns from data mining to forecast traffic flow, imputation missing data and buildup traffic models. One of data mining tool is clustering techniques which used to traffic pattern profile including traffic volume prediction (Stutz and Runkler, 2002; Yang *et al.*, 2017), dynamical traffic flow identification (Jiang *et al.*, 2003), congestion patterns extraction of daily traffic (Wen *et al.*, 2014) and non-recurrent traffic congestion detection on urban road networks (Anbaroglu *et al.*, 2014). Nair *et al.*, (2019) compare traffic congestion conditions across multiple cities of the world by using a standard data from crowdsourced.

## Methodology

Traffic data was obtained from Traffic and Accident Data Unit [TADU] at Gateshead Council. A number of traffic counters were selected that are located within Gosforth Area. The traffic data, in vehicles per hour [veh/h], was obtained from 45 counter sites. The traffic data varies between daily records for the year or for a couple of weeks depending on whether the traffic count site is either temporary or a fixed counter. The traffic data was the only the traffic count that was available. The traffic composition and speed were not provided. Due to limitations in the availability of data with regards to the flow characteristics, the sites could not be classified into congested, busy or quiet roads. Therefore, cluster analysis was conducted on the data to classify daily traffic profiles into several groups.

Cluster analysis is one of the methods recommended by the Traffic Monitoring Guide published by the US Department of Transport for classifying roads (US DoT, 2001). This guide is used to assist the local authorities to apply appropriate extrapolation factors for annual average daily traffic [AADT] to short periods of counting traffic. The short period traffic

count calculates the traffic for specific times of the day or specific days. This guide provides a method to classify a road based on traffic and road characteristics such as average speed and traffic flow then appropriate AADT interpolation factors are applied to each road based on the classification. Chen *et al.* (2008) showed that cluster analysis is an adequate technique for road classification and it requires little or no knowledge of the physical layout of the road network.

The method of unsupervised classification of behaviors or patterns into groups or clusters is called clustering. The patterns are clusters based on input data, which means unsupervised classification. By using a number of features, a pattern will be defined mathematically during the cluster analysis. These features are defined by hourly traffic flow or peaks and off peak traffic flow, which are daily traffic flow profiles. In some studies first cluster analysis are used then the results are validated by factor analysis (Berlage and Terweduwe, 1988; Clark *et al.*, 2003). One of the cluster techniques used is the k-means algorithm, as it is easy to apply, simple and efficient (Anil, 2010). It has been documented that the k-means algorithm is applicable in many fields such as engineering, energy, medical, electrical and transport (Chen *et al.*, 2008; Docquier *et al.*, 2009; Mora-Flórez *et al.*, 2009; Pandit *et al.*, 2011; Yiakopoulos *et al.*, 2011). Cluster analysis uses classified data that is divided into meaningful subgroups (Fraley and Raftery, 1998). The main categories of the cluster analysis techniques are hierarchical and non-hierarchical (Fraley and Raftery, 1998; Grimm and Yarnold, 2000; Anderson, 2003).

Non-hierarchical techniques allow multiple passes of the data which is an advantage over hierarchical methods. Each data point is allowed to move among clusters with each subsequent iteration and aims to maximise intra-cluster similarity and inter-cluster dissimilarity given that it will not increase square error within each cluster (Aldenderfer and Blashfield, 1984). However, these techniques are required for prior information to be applicable (Ketchen and Shook, 1996). For example, traffic characteristics and fleet compositions were used to classify more than 3000 roads in Leicester by using Non-hierarchical techniques (Chen *et al.*, 2008). The K-means clustering algorithm was used, which specifies the number of clusters before conducting the analysis. The numbers of clusters allowed have to be flexible, in order that the analysis will be derived from the data and not by the subsequent assumption.

The K-means algorithm is one of most commonly used cluster analysis techniques due to its ease of use and simplicity. Starting from an initial partition, the objects move iteratively from one group or cluster centre to another, which is the procedure for Kmeans cluster analysis (Fraley and Raftery, 1998; Cheung, 2003; Chitta

and Murty, 2010). Like all non-hierarchical procedures, the number of clusters has to be specified before conducting K-means cluster analysis (Grimm and Yarnold, 2000). The cluster centre has to be selected randomly, then objects grouped into the nearest cluster (Ketchen and Shook, 1996). The square error has to be calculated within each cluster (Kalyani and Swarup, 2011). For each cluster, the square error is minimized by allowing objects to move between clusters (Jeffrey *et al.*, 2006). The cluster centre is recalculated and the process is restarted in an iterative method in case an object has moved to another cluster. The final groupings are determined once the clusters become stable (Hair *et al.*, 1992).

This research employed the cluster analysis to classify roads based on traffic counts and as a result, the traffic data would be driving the classification process. The 2010 traffic count data for 45 sites in Newcastle was obtained from TADU within Gosforth Area, Newcastle. The traffic count data from each site varied from daily records for the year to a couple of weeks. Excel and SPSS software packages were used to analyse the traffic flow profile for each site. Before conducting cluster analysis, days with missing traffic data were edited and the missing traffic data for a day

was removed. The zero traffic volume was interpolated by assuming a linear relationship between the traffic data before and after that within two traffic volumes greater than 100 vehicles.

The data was coded before conducting cluster analysis and was divided into three groups, weekdays, Saturdays and Sundays. Each group was analysed independently for each site. The means and standard deviations were calculated for each hour of the 24 hrs. traffic count. The data from one day was excluded in the case of having four hourly traffic volumes in sequence or six hourly traffic volumes, which exceeded the 95% confidence interval. Subsequently, the means and standard errors were calculated for each hour of the 24-hr traffic count for each site.

### Traffic Data Analysis

In this section, the description of the traffic data analysis is presented in detail. The chapter consists of several stages of traffic data analysis. The aim of collecting the traffic data was to select the sites based on the traffic data classification. Due to the data limitation the aim was changed into classifying the daily traffic profile into groups.

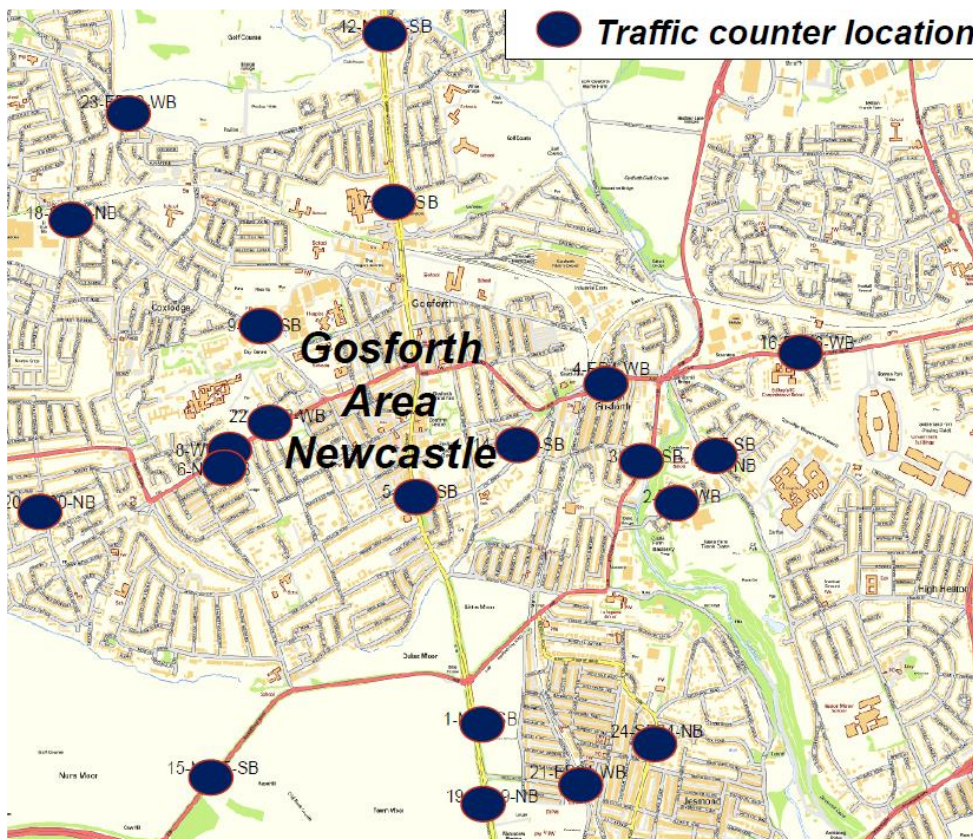


Fig. 1: Traffic counters location [Source: Edina]

### *Traffic Counters Location*

Traffic Counters Location Traffic data provided by TADU operating at Gateshead Council is used for traffic analysis. As shown in Fig. 1, the 2010 traffic count data for a number of sites in Newcastle were obtained from TADU. Sixty traffic count sites were identified that were located within Gosforth Area, Newcastle. Four traffic counters were not counting road traffic and eleven traffic counters had no traffic records. Furthermore, 45 counter sites had obtained traffic data. The traffic count data had the traffic flow per hour, which was saved in an Excel file. Depending on the traffic count site, traffic data varied between daily records for the year or for a couple of weeks. These record variations depended on whether the traffic count site had a fixed or a temporary counter. Indeed, the variation in the records could have been caused by traffic data being lost or corrupted. This data was analysed using Excel and SPSS software packages to carry out a traffic flow profile. A statistical method was used to identify traffic on a typical day. Next, a cluster analysis was conducted to group roads based on traffic data for a usual day.

### *Traffic Data Limitation*

The traffic count was the only traffic data that was available. The traffic composition and speed were not provided. Due to limitations in the available data with regards to the flow characteristics, the sites could not be classified into congested, busy or quiet roads. Therefore, the cluster analysis was conducted on the data to classify the daily traffic profile into several groups.

### *Cluster Analysis*

This research employed the cluster analysis as explained before to classify roads based on traffic counts, thus, the traffic data would be driving the classification process. The 2010 traffic count data for 45 sites in Newcastle was obtained from TADU within Gosforth Area, Newcastle. The traffic count data from each site varied from daily records for the year to a couple of weeks. Excel and SPSS software packages were used to analyse the traffic flow profile for each site.

Before conducting cluster analysis, the days with missing traffic data were edited. The missing data for the whole day was coded to be excluded from the analysis. The traffic volumes equal to zero within two traffic volumes greater than 100 vehicles were interpolated by assuming a linear relationship between the traffic data before and after the zero value. The day of the week and the month were coded to enable identification after conducting cluster analysis. Working days, weekends and holidays were the types of day coded to be identified

after conducting cluster analysis. The annual traffic flow was classified into three groups, specifically work days, Saturday and Sunday (Cheung, 2003). Four groups, Monday to Thursday, except a holiday, Friday or the day before a holiday and Saturday and Sunday are the other classification methods for the annual traffic flow (Chrobok *et al.*, 2004). Therefore, five clusters were specified for cluster analysis, attempting to capture the main clusters for each site. K-means cluster analysis was employed to classify hourly traffic data into five clusters for each traffic counter data. By using hourly traffic volumes for a day, a pattern was defined mathematically using K-means cluster analysis.

The traffic data were analysed to produce daily traffic profiles for each site. A flag was assigned to the missing data for the whole day, which was excluded from the analysis. The traffic data was divided into three groups, weekdays, Saturdays and Sundays. Each group was analysed independently for each site. The means and standard deviations were calculated for each hour of the 24-hr traffic count. The data from one day was excluded in case of having four hourly traffic volumes in sequence or six hourly traffic volumes exceeding a 95% confidence interval. Subsequently, the means and standard errors were calculated for each hour of the 24-hr traffic count for each site.

## **Results**

The results of the cluster analysis varied across the 45 traffic count sites. There were two to four clusters for each traffic counter site data. The first group classifies two clusters of weekday and weekend. Three clusters of weekday, Saturday and Sunday is the second group, the third and fourth groups include three to four clusters either two weekdays and weekend clusters or two weekdays, Saturday and Sunday clusters. Yang *et al.*, (2017) find five clusters of traffic profiles which is not different from the finding of four clusters in this paper despite of absent of traffic speed data.

Based on that the traffic count sites were grouped into four groups. The first group is consisted of 21 traffic counter sites which classified into weekday, Saturday and Sunday clusters. The second group was from ten traffic counter sites which classified into two clusters of weekday and weekend. The third group was consisted of data from another eleven traffic counter sites which classified into four clusters that can group the cluster into weekday 1, weekday 2, Saturday and Sunday. Weekday 1, weekday 2 and weekend were the classification of the fourth group clusters which is the data from three traffic counter sites.

K-means cluster analysis was a useful tool to classify traffic data based on the hourly flow traffic data, which could explain each cluster by day types. The results of the daily traffic profiles varied depending on the hourly volumes for the traffic and day type [weekday, Saturday and Sunday] Fig. 2 to 4. The profile trend for three day types followed similar patterns across the sites. The weekday profiles could be divided into four patterns based on the peaks, Fig. 2. One of the patterns had similar morning and evening peaks. Having either a morning or an evening peak was

the pattern for two weekday profiles. The last pattern did not have a morning or an evening peak; instead it had a steady flow of traffic from the morning to the evening. The traffic profiles for Saturdays and Sundays were divided into three traffic profile shapes based on the peaks, Fig. 3 and 4. Two small peaks occurred at noon and in the afternoon, which was one of the traffic profile patterns. A further traffic profile pattern was one peak during the noon. The third traffic pattern had one peak and had a high volume of traffic during the early hours of the morning.

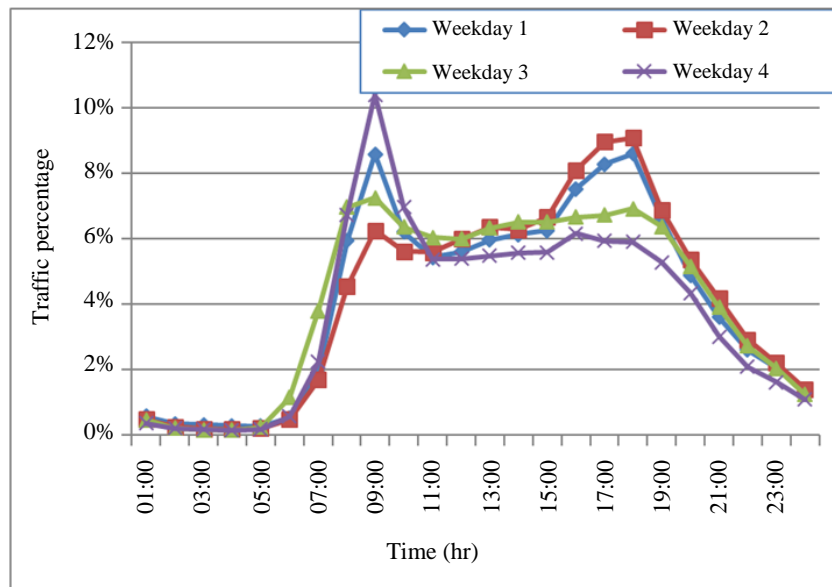


Fig. 2: Weekday traffic profiles

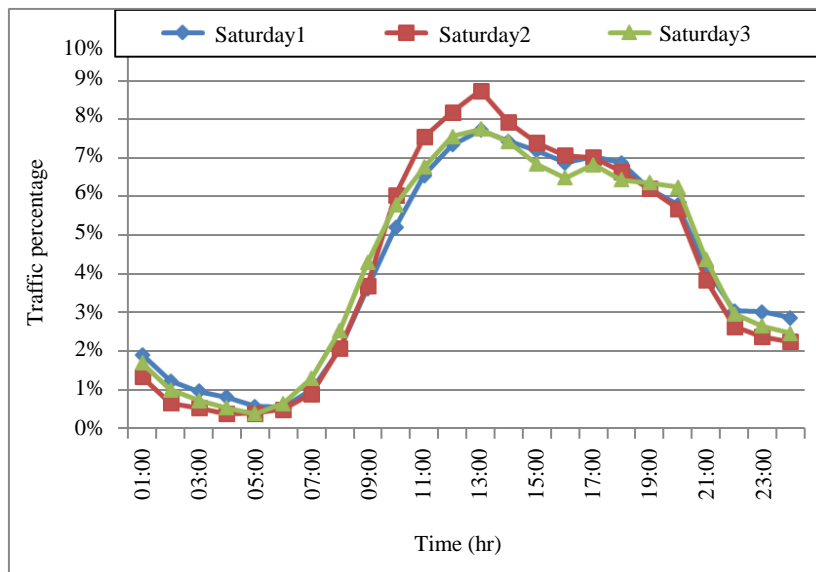


Fig. 3: Saturday traffic profiles

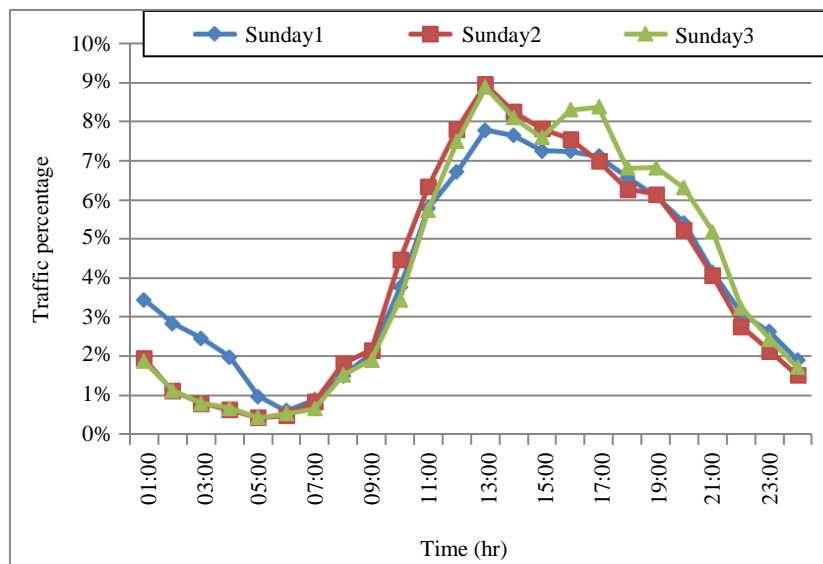


Fig. 4: Sunday traffic profiles

## Conclusion

This paper gave a description of traffic data analysis, the cluster analysis is employed to exam daily traffic data variation and produce traffic patterns profile for the road. A daily traffic profile was classified based on the traffic data caused by several limitations. Data from each traffic counter site was grouped into two to four clusters. The traffic data was classified Successfully by employing the K-means cluster analysis based on the hourly flow traffic data, which could explain each cluster by day types. The daily traffic profiles varied depending on the day type and the hourly volumes for the traffic. These results could be used prior to road traffic or air quality modelling. Road transport emissions can be calculated based on daily traffic profiles.

## Author's Contributions

**Hamad B Matar:** Writing the paper and analysis.

**Talal Almutairi:** Litteration review, data collection and preparation.

**Nayef Z Al-Mutairi:** Editing and preperere for publication.

## Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

## References

Aldenderfer, M.S. and R.K. Blashfield, 1984. Cluster analysis. Sage Publications, Beverly Hills.

Anbaroglu, B., B. Heydecker and T. Cheng, 2014. Spatio-temporal clustering for non-recurrent traffic congestion detection on urban road networks. *Transport. Res. C*, 48: 47-65.  
 DOI: 10.1016/j.trc.2014.08.002

Anderson, T.W., 2003. *An Introduction to Multivariate Statistical Analysis*. John Wiley and Sons.

Anil, K.J., 2010. Data clustering: 50 years beyond K-means. *Patt. Recogn. Lett.*, 31: 651-666.  
 DOI: 10.1016/j.patrec.2009.09.011

Berlage, L. and D. Terweduwe, 1988. The classification of countries by cluster and by factor analysis. *World Dev.*, 16: 1527-1545.  
 DOI: 10.1016/0305-750X(88)90225-2

Chen, H., A. Namdeo and M. Bell, 2008. Classification of road traffic and roadside pollution concentrations for assessment of personal exposure. *Environ. Modell. Software*, 23: 282-287.  
 DOI: 10.1016/j.envsoft.2007.04.006

Cheung, Y.M., 2003. k\*-Means: A new generalized k-means clustering algorithm. *Patt. Recogn. Lett.*, 24: 2883-2893.  
 DOI: 10.1016/S0167-8655(03)00146-6

Chitta, R. and N.M. Murty, 2010. Two-level k-means clustering algorithm for k- relationship establishment and linear-time classification. *Patt. Recogn.*, 43: 796-804.  
 DOI: 10.1016/j.patcog.2009.09.019

Chrobok, R., O. Kaumann, J. Wahle and M. Schreckenber, 2004. Different methods of traffic forecast based on real data. *Eur. J. Operat. Res.*, 155: 558-568.  
 DOI: 10.1016/j.ejor.2003.08.005

- Clark, W.C., J.P. Kuhl, M.L. Keohan, H. Knotkova and R.T. Winer *et al.*, 2003. Factor analysis validates the cluster structure of the dendrogram underlying the Multidimensional Affect and Pain Survey (MAPS) and challenges the a priori classification of the descriptors in the McGill Pain Questionnaire (MPQ). *Pain*, 106: 357-363.  
DOI: 10.1016/j.pain.2003.08.005
- DfT, 2001. Transport statistics bulletin, road traffic statistic. Statistics Report SB (02), Department for Transport, UK.
- Docquier, P.L., L. Paul, R. Menten, O. Cartiaux and B. Francq *et al.*, 2009. Measurement of bone cyst fluid volume using k-means clustering. *Magnetic Resonance Imag.*, 27: 1430-1439.  
DOI: 10.1016/j.mri.2009.05.017
- Elhenawy, M., H. Chen and H.A. Rakha, 2014. Dynamic travel time prediction using data clustering and genetic programming. *Transport. Res. C*, 42: 82-98.  
DOI: 10.1016/j.trc.2014.02.016
- Festin, S.M., 1996. Summary of national and regional travel trends, 1970-1995.
- Fraley, C. and A. Raftery, 1998. How many clusters? Which clustering method? Answers via model-based cluster analysis. *Comput. J.*, 41: 578- 588.  
DOI: 10.1093/comjnl/41.8.578
- Francesc, S., 2012. Deriving traffic flow patterns from historical data. *J. Transport. Eng.*, 138: 1430-1441.  
DOI: 10.1061/(ASCE)TE.1943-5436.0000456.
- Grimm, L.G. and P.R. Yarnold, 2000. Reading and understanding MORE multivariate statistics. American Psychological Association, Washington, DC, US.
- Hair, J.F., R.E. Anderson, R.L. Tatham and W.C. Black, 1992. *Multivariate data analysis: With readings.* Macmillan Publishing Company, New York USA.
- Heidemann, J., F. Silva, X. Wang, G. Giuliano and M. Hu, 2008. Sensors for unplanned roadway events--simulation and evaluation. Final Report, University of Southern California, Transportation Service Department, Los Angeles, CA.
- Jeffrey, E., A. Martin and M. Anirban, 2006. Traffic classification using clustering algorithms. Proceedings of the SIGCOMM Workshop on Mining Network Data, Sept. 11-15, ACM, Pisa, Italy, pp: 281-286. DOI: 10.1145/1162678.1162679
- Jiang, G., J. Wang, X. Zhang and L. Gang, 2003. The study on the application of fuzzy clustering analysis in the dynamic identification of road traffic state. Proceedings of the IEEE International Conference on Intelligent Transportation Systems, Oct. 12-15, IEEE Xplore Press, Shanghai, China, pp: 1149-1152. DOI: 10.1109/ITSC.2003.1251986
- Kalyani, S. and K.S. Swarup, 2011. Particle swarm optimization based K-means clustering approach for security assessment in power systems. *Expert Syst. Applic.*, 38:10839-10846.  
DOI: 10.1016/j.eswa.2011.02.086
- Ketchen, D.J. and C.L. Shook, 1996. The application of cluster analysis in strategic management research: An analysis and critique. *Strategic Manage. J.*, 17: 441-458. DOI: 10.1002/(SICI)1097-0266(199606)17:6<441::AID-SMJ819>3.0.CO;2-G
- Leduc, G., 2008. Road traffic data: Collection methods and applications. Working Papers on Energy, Transport and Climate Change, European Commission Joint Research Centre Institute for Prospective.
- Lee, W., S.S. Tseng, J.L. Shieh and H.H. Chen, 2011. Discovering traffic bottlenecks in an urban network by spatiotemporal data mining on location-based services. *IEEE Trans. Intell. Transport. Syst.*, 12: 1047-1056. DOI: 10.1109/TITS.2011.2144586
- Lin, W.H., J. Dahlgren and H. Huo, 2004. Enhancement of vehicle speed estimation with single loop detectors. *J. Transport. Res. Board*, 1870: 147-152.  
DOI: 10.3141/1870-19
- Mimbela, L.E.Y. and L.A. Klein, 2000. Summary of vehicle detection and surveillance technologies used in intelligent transportation systems.
- Mora-Flórez, J., J. Cormane-Angarita and G. Ordóñez-Plata, 2009. k-means algorithm and mixture distributions for locating faults in power systems. *Electric Power Syst. Res.*, 79: 714-721.  
DOI: 10.1016/j.epsr.2008.10.011
- Nair, D.J., F. Gilles, S. Chand, N. Saxena and V. Dixit, 2019. Characterizing multicity urban traffic conditions using crowdsourced data. *PLOS One*, 14: e0212845-e0212845. 10.1371/journal.pone.0212845
- Pandit, Y.P., Y.P. Badhe, B.K. Sharma, S.S. Tambe and B.D. Kulkarni, 2011. Classification of Indian power coals using K-means clustering and Self Organizing Map neural network. *Fuel*, 90: 339-347.  
DOI: 10.1016/j.fuel.2010.09.012
- Ritchie, S.G., S. Park, C. Oh and C. Sun, 2002. Field investigation of advanced vehicle reidentification techniques and detector technologies-Phase 1.
- Smith, D. and J. McIntyre, 2002. *Handbook of simplified practice for traffic studies.*
- Stathopoulos, A. and M. Karlaftis, 2001. Temporal and spatial variations of realtime traffic data in urban areas. *J. Transport. Res. Board*, 1768: 135-140.  
DOI: 10.3141/1768-16
- Stutz, C. and T.A. Runkler, 2002. Classification and prediction of road traffic using application-specific fuzzy clustering. *IEEE Trans. Fuzzy Syst.*, 10: 297-308. DOI: 10.1109/TFUZZ.2002.1006433

- Taylor, M.A.P., W. Young and P.W. Bonsall, 1996. Understanding traffic systems: Data analysis and presentation (0291398154).
- US DoT, 2001. Traffic monitoring guide, technical report. Federal Highway Administration, Washington, USA.
- Weijermars, W. and E. Van Berkum, 2005. Analyzing highway flow patterns using cluster analysis. Proceedings of the IEEE Intelligent Transportation Systems, Sept. 16-16, IEEE Xplore Press, Vienna, Austria, pp: 308-313.  
DOI: 10.1109/ITSC.2005.1520157
- Wen, H., J. Sun and X. Zhang, 2014. Study on traffic congestion patterns of large city in China taking Beijing as an example. Proc. Soc. Behav. Sci., 138: 482-491. DOI: 10.1016/j.sbspro.2014.07.227
- Yang, S., J. Wu, G. Qi and K. Tian, 2017. Analysis of traffic state variation patterns for urban road network based on spectral clustering. Adv. Mechanical Eng.
- Yiakopoulos, C.T., K.C. Gryllias and I.A. Antoniadis, 2011. Rolling element bearing fault detection in industrial environments based on a K-means clustering approach. Expert Syst. Applic., 38: 2888-2911. DOI: 10.1016/j.eswa.2010.08.083
- Zhang, J., F.Y. Wang, K. Wang, W.H. Lin and X. Xu *et al.* 2011. Data-driven intelligent transportation systems: A survey. IEEE Trans. Intell. Transport. Syst., 12: 1624-1639.  
DOI: 10.1109/TITS.2011.2158001