

Implementation of Fuzzy Cluster Analysis to Partition Cost Performance into Typical Groups Based Upon Project Characteristics

Phuong H.D. Nguyen, Ph.D. Student, Dan Tran, Ph.D., and Brian C. Lines, Ph.D.
University of Kansas
Lawrence, Kansas

Quantification of project cost performance plays an essential role in any decision-making process in any transportation infrastructure projects. Project characteristics, such as facility type, project type, and complexity are critical factors that certainly affect predictions and classifications of cost performance. Due to the fact that project complexity and other project attributes are often given in terms of subjective judgements and ratings of experts, it is difficult to quantify those qualitative data types. This research implemented the fuzzy set theory in the context of fuzzy cluster analysis to classify/partition input data of cost performance and project characteristics into meaningful groups. The fuzzy classification process was conducted with a dataset of 254 horizontal transportation projects collected by the Federal Highway Administration (FHWA) in 2012. As a result, this paper shows the applicability of fuzzy cluster analysis within the construction industry with common cluster validity indices. This research contributes to the construction body of knowledge and practitioners a new method to classify cost performance data and understand its underlying structures and behaviors. Future work of this study is to examine other project performance measurement metrics, such as schedule performance and quality.

Key words: fuzzy set theory, fuzzy c-means cluster analysis, cost performance, project complexity

Introduction

Cost performance is one of the key criteria for construction project owners and stakeholders to consider in any decision-making process. Determination of cost performance is varied on a project-to-project basis and depends on many factors, including project characteristics, internal and external conditions, and other features (Baccarini 2004). During feasibility, planning, and design stages, cost estimates are prepared in detail based upon project characteristics and take into account potential costs for associated project uncertainty and complexity (Creedy *et al.* 2010). Due to the fact that different levels of project complexity drive the estimated cost performance diversely, it is important that both industry practitioners and academic researchers need to generate empirical and sufficient methods/frameworks/models to quantify cost performance based on project characteristics.

Traditionally in construction, cost performance is measured and estimated based on probabilistically mathematical frameworks and simulation models (Creedy 2006, Dikmen *et al.* 2007). Among the factors that affect cost performance, levels of project complexity and uncertainty play an important role in establishing more deterministically predicting results (Hegazy and Ayed 1998). However, uncertainty and complexity are often given in terms of linguistic input variables. In other words, those factors are assessed based on experience of professionals and experts in the context of subjective judgements. One of the weaknesses of probability theory and simulation methods is not able to quantify qualitative input data (Hastie *et al.* 2009). Despite the fact that those methods are often considered to quantity project characteristics and other attributes, a more comprehensive method, which can incorporate quantifications of both quantitative and qualitative input data, is necessary. Hegazy and Ayed (1998) declared that fuzzy set theory is useful in the construction industry where realistic historical project data are limited.

In the field of engineering, fuzzy set theory has been used to capture qualitative domain professional judgements to generate theoretical decision-making models and widely applied to many areas, such as computer science, mechanical engineering, aerospace engineering, and chemical engineering (Elwood 2014, Kruse *et al.* 2007). As a promising method to mathematically take into account the subjective judgements and expressions of construction professionals and experts about project complexity, fuzzy set theory was applied in this study in terms of a soft clustering method introduced by Bezdek (1993). This method is also known as the fuzzy c-means (FCM).

Literature Review

In order to measure and predict project performance in construction, many studies have proposed various applications of fuzzy sets. Baccarini (2004) and Creedy (2006) showed that utilizing fuzzy sets is one of the essential methods of estimating construction project cost contingency in terms of conceptual models to reduce impacts of risks and uncertainty. Paek *et al.* (1993) proposed the use of fuzzy set application in a risk-pricing model to identify and quantify project risks to estimate project cost under uncertainty. A fuzzy decision-aid framework to take into account global risk factors that affect cost performance was established by Baloi and Price (2003), which indicated how well fuzzy set theory is viable for modelling and examining project risks. Dikmen *et al.* (2007) recommended the use of fuzzy sets in generating a fuzzy risk assessment method to estimate project cost-overruns based on risk ratings. However, those studies have mainly focused on fuzzy logic applications, such as fuzzy systems and models, to support decision-making processes. Another critical feature of fuzzy sets is to classify or partition data into meaningful clusters to recognize typical patterns of given datasets (Hoppner *et al.* 1999, Zadeh 1978). Practicing fuzzy classification enhances understanding, handling, and predicting data information and behavior, which is essential in investigation of construction project performance based on empirical data. Fuzzy classification is often demonstrated in the context of fuzzy cluster analysis.

Cluster analysis is one of unsupervised machine learning techniques used to classify data based on similarities in attributes, features, and other characteristics. It concentrates on grouping data to identify and study underlying data structures (Hoppner *et al.* 1999, Klir and Yuan 1995). There are two common types of cluster analysis: (1) hard cluster analysis, which is developed based on crisp sets, and (2) soft cluster analysis, which is formulated based on fuzzy sets or fuzzy set theory. The scope of this research was to enhance fuzzy set theory to deal with qualitative input data of complexity's ratings; therefore, a soft cluster analysis or fuzzy cluster analysis was conducted. Fuzzy cluster analysis is different from the crisp cluster analysis in terms of assigning membership values to the clustered data points instead of limiting a single data point to belong to only one cluster (Elwood 2014). This characteristic helps fuzzy cluster analysis can implement in any given real-world problem.

Research Objectives

The motivation of this research was to search for underlying structures of cost performance in empirical transportation project data based on project characteristics and classify these structures into meaningful clustering groups. Degrees of similarity and dissimilarity of empirically collected project data give a precise representation of cost performance within the domain of transportation infrastructures. This study has three main research objectives, which are based upon three fundamental problems of cluster analysis, as follows:

The first objective aims to investigate if the dataset is actually clusterable by assessing clustering tendency of the dataset. The assessment investigates if cluster analysis is appropriate for this type of datasets in consideration of establishing meaningful clustering groups of cost performance from individual empirical transportation projects. In other words, this question determines if utilizing cluster analysis provides valid generalizations of typical groups of cost performance based on typical transportation project characteristics, such as facility type, project type, and complexity.

The second objective aims to solve the problem of establishing the "best" number of partitioning groups. An appropriate cluster analysis algorithm should be selected in order to examine input data of quantitative project characteristics and qualitatively subjective judgements of project complexity. The fuzzy c-means method was selected because of its capacity of studying linguistic/fuzzy information and modeling common groups of cost performance. The key mechanism of this clustering algorithm is to evaluate heterogeneity, the overall diversity among data points in all of the clusters. Increasing in heterogeneity leads to merging of dissimilar clusters, which reduces opportunities to determine the most optimal number of clusters.

The third objective aims to validate the appropriateness of the identified numbers of clusters with multiple fuzzy clustering indices. Theses clustering indices assess membership levels of data points in each cluster to affirm if the data points actually belong to that cluster. Due to the fact that fuzzy cluster analysis belongs to an unsupervised learning environment, no validation is actually required. In addition, determination of the proper number of clusters is also intuitive and depends on experience of analysts. Therefore, it is challenging to explicitly declare the best number of clusters. However, using clustering indices certainly helps verify and confirm suitability of the identified cost performance's clusters.

Methodology

Data Collection

This research paper investigated an empirical dataset of 254 horizontal transportation projects within 50 states in the U.S., from a Federal Highway Administration (FHWA) survey in 2012. Collected projects consist various characteristics of facility type, project type, highway type, complexity rating, and project cost and schedule data. Fuzzy cluster analysis is critically influenced by outliers, so identified outliers of the dataset were removed. The sample size of this dataset is adequate to use with cluster analysis because no statistical inference power is considered within this domain. In other words, cluster analysis only provides the representativeness of the sample within the underlying structure. No normality, linearity, or homoscedasticity was assumed with this dataset. Prior to conducting cluster analysis, selection of major input features/variables was required. This study selected five common variables (project characteristics) in order to conduct cluster analysis as shown in Table 1.

Table 1

Descriptive statistics of selected project characteristics

Project Characteristic	N	Data Type	Mean	Min	Max
Facility Type – Road (%)	254	Continuous	45.56	0	100
Facility Type – Bridge (%)	254	Continuous	33.58	0	100
Facility Type – Drainage (%)	254	Continuous	6.71	0	100
Facility Type – ITS* (%)	254	Continuous	3.08	0	100
Facility Type – Other (%)	254	Continuous	11.10	0	100
Project Type – New Construction (%)	254	Continuous	45.47	0	100
Project Type – Reconstruction (%)	254	Continuous	39.46	0	100
Project Type – Other (%)	254	Continuous	15.07	0	100
Complexity	254	Ordinal	2	1	3
Cost Growth (%)	254	Continuous	3	-10	20

Note. *ITS - Intelligent Transportation System

The first variable is facility type, which consists of five sub-variables: road, bridge, drainage, intelligent transportation system (ITS), and others, based on approximate percentages of the total project cost. The range of each type is from 0 to 100%. For example, a transportation project might have 80% of road, 10% of bridge, and 10% of ITS; the total should be always 100%. Facility type contains continuous data and provides five input variables. The second variable is project type, which includes three sub-variables: new construction/expansion, reconstruction/rehabilitation/resurfacing, and others, based on approximate percentages of the total project cost. Similarly to facility type, the range of each type is from 0 to 100%, and it follows the continuous data type. For instance, a transportation project might have two-thirds of new construction and one-third of resurfacing. Project type provides three input variables. The third variable is project complexity, which is rated based on a 3-point Likert scale and follows the ordinal data type. First, the "most complex" projects include those that are new highways/major relocations, new interchanges, capacity adding/major widening, major reconstruction, require congestion management studies, and have complex environmental assessment or environmental impact statements. Second, the "moderately complex" projects include those that are minor roadway relocations, non-complex bridge replacements with minor roadway approach work, and non-complex environmental assessment required. Third, the "non-complex" projects include those that are maintenance betterment projects, overlay projects with simple widening, little or no utility coordination, non-complex enhancement projects without new bridges, and categorical exclusion. The final selected variable is cost performance. Several data points were collected to understand the cost performance of each project within the dataset, such as the engineer's estimate, contract award value, and final cost. The final cost is equal to the contract award plus costs of all change orders. To represent the cost performance variable, this study used cost growth, which is the overall performance at project completion, calculated from the project cost data with the following equation:

$$\text{Cost Growth (\%)} = (\text{Final Cost} - \text{Contract Award})*100/\ \text{Contract Award} \quad (1)$$

The calculated cost growths were divided into five linguistic groups ranging from -10% to 20%: “saving” from -10% to -1%, “none” from -1% to 1%, “low” from 1% to 5%, “medium” from 5% to 10%, and “high” from 10% to 20%. This study removed three projects which contain completely missing cost growth values and eighteen projects which have extreme cost growth values out of the selected range. The majority of the cost growth values fall within the range from -1% to 5%.

Data Analysis

The process of fuzzy cluster analysis in this study included four key steps: data standardization, data clustering tendency’s assessment, clustering c-means algorithm, and validation of clustering results. These analyses were conducted in the R programming environment with multiple clustering packages.

The dataset of selected variables were standardized prior to conducting fuzzy cluster analysis. Because selected variables were collected in different units, it is not meaningful to assess similarity/dissimilarity of the data points. In order to assess the similarity of two projects, cluster analysis calculates the Euclidean distance between them, which is a geometric measure of closeness between data points (Castellano *et al.* 2007). The value of this distance is closely related to the measuring scale of selected variables and influences the shape of the clusters (Chiu 1994). Thus, selected variables were standardized to a unified scale to avoid impacts of dissimilarity measures. Within the domain of fuzzy cluster analysis, ranging is one of the recommended methods where comparisons of data points are more proper, and its equation is as below (Elwood 2014, Hoppner *et al.* 1999). Accordingly, the standardized data ranged from 0 to 1.

This research evaluated data clustering tendency to confirm the feasibility of cluster analysis. Specifically, this process examined whether or not the dataset potentially produces meaningful clusters/groups. It is crucial to assess the tendency of the dataset because of a critical issue of cluster analysis where the algorithm will return a number of clusters even if the dataset does not consist of any meaningful clustering group (Kruse *et al.* 2007). There were two methods for assessing the clustering tendency used in this research: (1) statistical with Hopkins statistics and (2) visual assessment of cluster tendency (VAT) algorithm. The Hopkins statistics method measures the probability with which the given dataset is established by a uniform data distribution in order to examine the spatial randomness of the dataset which determines if meaningful clusters exist (D’Urso 2007). The null hypothesis of this method is that the dataset is uniformly distributed while the alternative hypothesis is that the dataset is not uniformly distributed. The VAT method provides the graphical determination of the clustering tendency. This step identifies applicability of cluster analysis to the dataset.

FCM algorithm, the most well-known clustering method for fuzzy cluster analysis according to (Bezdek *et al.* 1999), was implemented to partition the dataset in terms of cost growth. The main objective of this algorithm is to partition/classify all data points into homogeneous clustering groups where the data points achieve higher levels of similarity in the same clusters and contain lower levels of similarity in other clusters. The similarity of the data points is measured by geometric proximity or distance in an n-dimensional space. The result of the FCM algorithm provides a family of fuzzy sets where clustering groups are generated based on the membership values; the closer the data points within the cluster, the higher membership value (Kruse *et al.* 2007). The main attribute of FCM is that an individual data point shares membership with more than one cluster (Elwood 2014). In other words, each data point is assigned a degree of membership in each cluster, so a data point can belong to multiple clusters, which shows the overlapping characteristic of fuzzy clustering groups. In the context of fuzzy cluster analysis, no cluster is empty as well as no cluster can contain all of the data points, axiomatically. A particular restriction of FCM is that the sum of membership values of an individual data point across all clusters should be equal 1. The essential outcomes of FCM are a fuzzy c-partition matrix, where the degree of membership of a data point in a specific cluster is described, and a vector of cluster center coordinates, which provides underlying prototypes or prototypical representations of the clustered data points. After clustering, the partitioned data should be hardened to classify the data points into actual crisp clustering groups for subsequent humanistic judgements; this process is called defuzzification. The defuzzification method used in this research is the maximum membership values, which the most common method of reducing fuzzy information into a single crisp value in the field of fuzzy set theory.

To fuzzily partition project data, the FCM method considers an optimization function which attempts to simultaneously minimize the distance between data points within clusters and maximize the distance between clusters. This optimization function is used to support the determination of the optimal number of clustering groups, which is subjective and depends on measurements of similarity and clustering parameters. This research utilized two direct methods (or visualization): Elbow method and Silhouette method, and two statistical methods: gap statistics and NbClust method, in order to select the most appropriate number of clusters. The Elbow method selects a number of clusters from which adding another cluster does not increase the total within sum of squares, measured by the total distances between data points. The Silhouette method evaluates how well each data point lies within the associated cluster; in other words, this method measures the clustering quality (Ross 2010). The optimal number of clusters is the one that achieves the highest value of the average silhouette. The gap statistics method produces the optimal number of clusters by comparisons of the total within intra-cluster variation for different numbers of clusters with statistically anticipated values. The maximum gap statistic value provides the optimal number of clusters. The NbClust method examines thirty clustering indices to select the optimal number of clusters based upon the majority rule. This step determines the number of meaningful clusters.

To evaluate the goodness of FCM algorithm clustering outcomes, five clustering indices, partition entropy, partition coefficient, fuzzy silhouette index, the Dunn index, and the silhouette width, were used. Using cluster validity helps avoid randomness in identifying clusters to provide better recognition of underlying structures within the dataset. This step addresses the problem of validating the clustering results.

Results

Evaluation of Data Clustering Tendency

Investigation of tendency assessment with two methods, Hopkins statistic and VAT, confirmed that the FHWA dataset contained meaningful clustering groups with five selected variables: facility type, project type, complexity, and cost growth. In the first method, if the Hopkins statistic is close to zero, the null hypothesis of the dataset of uniformly distributed is rejected. As a result, the hypothesis testing provided a Hopkins statistic of 0.274 which is far below the threshold of 0.5 and concluded that the FHWA dataset is significantly clusterable. In the second method, the clustering tendency is visually assessed by counting the amount of dark squares along the diagonal of a dissimilarity matrix in the VAT image. This method was also in line with the Hopkins method to confirm that there was a clustering structure in the FHWA dataset.

Identification of Optimal Numbers of Clusters

Selection of an optimal number of clusters is subjective and depends on methods of similarity measurement and clustering parameters. This study utilized two visualization-based techniques, elbow and silhouette, and two statistics-based approaches, gap statistics and Nbclust, to determine the most optimal number of clusters. The elbow method visualizes the total within sum of square (WSS), which is calculated based on distances between data points in a cluster, to identify the minimum value of WSS or the elbow of the WSS' graph. This method recommended a range of the optimal numbers in which seven was the most optimal option. The silhouette technique computed the average silhouette width, which measures how well the clustering result is; the higher the average silhouette width, the better the clustering result. This method recommended six optimal options (two, three, four, six, seven, and eight) where seven was observed with the highest silhouette width. The gap statistics method selects the optimal number of clusters by comparing the WSS of different numbers of clusters with statistically anticipated values. The option producing the maximum value of the gap statistic was selected, which was the seven-cluster option. The last method, NbClust, which is a R programming-based function using thirty clustering indices to identify the most optimal number of clusters, proposed the seven-cluster option should be selected. As a result of the four methods, the range of the potential numbers of clusters was from two to ten clusters, and the most selected optimal number of clusters was seven. Accordingly, seven clusters were pre-defined as the input for the number of cluster centers to the FCM algorithm.

With the recommendation of seven clusters and N = 254 projects (data points), the FCM algorithm first identified seven cluster centers, and then assigned data points to the appropriate clusters based on the closeness of the data points to the cluster centers by calculating the distances between them. Essentially, this algorithm concurrently minimized the distance between data points within a cluster and maximized the distance between seven clusters based on similarity and dissimilarity in five selected features: facility type, project type, complexity, and cost growth. A detailed process of using FCM to minimize the within-cluster distances and maximize the between-cluster distances is provided in Ross (2010) and Klir and Yuan (1995).

Table 2 summarizes characteristics of clustering groups of cost growth based on the majority rule of fuzzy cluster analysis which proposes that data points belong to the cluster obtaining the maximum value of memberships. Group 1 represents transportation projects with saving to low cost growth. Group 2 represents transportation projects with low cost growth. Group 3 represents transportation projects with medium to high cost growth. Group 4 represents transportation projects with saving to low cost growth. Group 5 represents transportation projects with low cost growth. Group 6 represents transportation projects with none cost growth. Group 7 represents transportation projects with none to low cost growth.

Table 2

Characteristics of clustering groups of cost growth

Characteristic of Cost Growth	**Group 1**	**Group 2**	**Group 3**	**Group 4**	**Group 5**	**Group 6**	**Group 7**
Clustering Range (%)	[-10, 5]	[1, 5]	[5, 20]	[-10, 5]	[1, 5]	[-1, 1]	[-1, 5]
Linguistic Description	Saving to low	Low	Medium to high	Saving to low	Low	None	None to low

Discussion

Although conducting cluster analysis does not require any comprehensive validation process due to the fact that it is an unsupervised machine learning technique, the goodness of determination of the optimal number of meaningful clusters is critical for reliability of clustered results. This research employed five common cluster validity indices in the domain of fuzzy cluster analysis. The partition entropy has the decision criteria ranged from 0 to 1: 0 is excellent, 0 to 0.5 is good, 0.5 to 1 is fair, and 1 is bad; accordingly, the partition entropy index should be minimized (Říhová and Makhalova 2017). The clustering result showed that the partition entropy obtained a value of 0.46, which is good. The other indices (partition coefficient, fuzzy silhouette, silhouette width, and Dunn index) have the decision criteria ranged from 0 to 1: 0 is bad, 0 to 0.5 is fair, 0.5 to 1 is good, and 1 is excellent., which means these indices should be maximized to have more accurate clustering results (Pal *et al.* 1995, Wu and Yang 2005). The clustering result represented a fair partition coefficient (0.11), a fair fuzzy silhouette (0.44), a good silhouette width (0.58), and a fair Dunn index (0.11). According to the selected five common cluster validity indices, the result of seven clusters are valid and reliable in the field of fuzzy cluster analysis.

Conclusion

Identifying meaningful groups of cost performance based on given project characteristics, including facility type, project type, and project complexity, is an important task in construction in order to optimize decision-making processes. This research paper quantified cost performance associated with project characteristics in transportation construction based on fuzzy set theory in terms of an unsupervised learning technique, fuzzy cluster analysis. As a result, seven clustering groups of cost performance based on facility type, project type, and complexity were recognized and divided into five major clustering groups based on cost performance. The first clustering group was recognized with saving to low cost growth transportation projects. The second group was recognized with none cost growth transportation projects. The third group was recognized with none to low cost growth transportation projects. The fourth group was recognized with low cost growth transportation projects. The fifth group was recognized with medium to high cost growth transportation projects. The validation section shows that the identified groups of cost growth were valid, which indicates the positive applicability of fuzzy cluster analysis in construction. This study contributes the five typical cost performance groups to support decision makers and agencies in the feasibility and planning phases of the transportation project. A limitation of this study was that only transportation projects were considered. This research plans to implement fuzzy set theory to other project performance measurements, such as schedule performance and quality. In addition, more project attributes, including project size, project delivery method, procurement method, and payment method should be incorporated.

References

Baccarini, D. (2004). "Accuracy in estimating project cost construction contingencies— a statistical analysis." *Proc., COBRA 2004*, Leeds Metropolitan Univ.: RICS Foundation, U.K.

Baloi, D., & Price, A. (2003). "Modelling global risk factors affecting construction cost performance." *International Journal of Project Management*, *21*(4), 261-269.

Bezdek, J. C. (1993). "Fuzzy models—what are they, and why?" *IEEE Trans. Fuzzy Syst.,* 1(1), 1–6.

Castellano, G., Fanelli, A. M., & Mencar, C. (2007). "Mining diagnostic rules using fuzzy clustering." *Advances in fuzzy clustering and its applications*, J. V. de Oliveira and W. Pedrycz, eds., Wiley, West Sussex, England, 211–228.

Chiu, S. L. (1994). "Fuzzy model identification based on cluster estimation." *J. Intell. Fuzzy Syst.*, 2(4), 267–278.

Creedy, G. (2006). *Risk Factors leading to cost overrun in the delivery of highway construction projects* (Ph.D). Queensland University of Technology.

Creedy, G., Skitmore, M., & Wong, J. (2010). "Evaluation of Risk Factors Leading to Cost Overrun in Delivery of Highway Construction Projects." *Journal of Construction Engineering and Management*, 136(5), 528-537.

Dikmen, I., Birgonul, M., & Han, S. (2007). "Using fuzzy risk assessment to rate cost overrun risk in international construction projects." *International Journal of Project Management*, 25(5), 494-505.

D'Urso, P. (2007). "Fuzzy Clustering of Fuzzy Data." *Advances in Fuzzy Clustering and its Applications*. Eds. J. V. de Oliveira and W. Pedrycz. West Sussex, England: Wiley, 155 – 192.

Elwood, E. D. (2014). *Fuzzy classification and fuzzy pattern recognition of seismic damage to concrete structures*. ⟨http://www/proquest.com/products-services/dissertations/⟩ (Dec. 31, 2014).

Hastie, T., Tibshirani, R., & Friedman, J. (2009*). The elements of statistical learning: Data mining, inference, and prediction,* 2nd Ed., Springer, New York.

Hegazy, T. & Ayed, A. (1998). "Neural network model for parametric cost estimation of highway projects." *Journal of Construction Engineering and Management*, 210 –218.

Hoppner, F., Klawonn, F., Kruse, R., & Runkler, T. (1999). *Fuzzy cluster analysis methods for classification, data analysis and image recognition*, Wiley, New York.

Klir, G. J., & Yuan, B. (1995). *Fuzzy sets and fuzzy logic theory and applications*, Prentice Hall, Upper Saddle River, NJ.

Kruse, R., Doring, C., & Lesot, M.-J. (2007). "Fundamentals of fuzzy clustering." *Advances in fuzzy clustering and its applications,* J. V. de Oliveira and W. Pedrycz, eds., Wiley, West Sussex, U.K., 3–30.

Paek, J. H., Lee, Y. W., & Ock, J. H. (1993). "Pricing construction risk: Fuzzy set application." *J. Constr. Eng. Manage*., 119(4), 743–756.

Pal, N. R., & Bezdek, J. C. (1995). "On cluster validity for the fuzzy c-means model." *IEEE Trans. Fuzzy Syst.*, 3(3), 370–379.

Říhová, E., & Makhalova, T. (2017). "On Evaluating of Fuzzy Clustering Results." *In The 11th International Days of Statistics and Economics*. Prague, Czech Republic.

Ross, T. J. (2010). *Fuzzy logic with engineering applications*, 3rd Ed., Wiley, West Sussex, U.K.

Wu, K., & Yang, M. (2005). "A cluster validity index for fuzzy clustering." *Pattern Recognition Letters*, 26(9).

Zadeh, L. A. (1978). "Fuzzy sets as a basis for a theory of possibility." *Fuzzy Sets Syst.,* 1(1), 3–28.