Clustering Data and Vague Concepts Using Prototype Theory Interpreted Label Semantics

Hanqing Zhao[1,2] and Zengchang Qin[1](✉)

[1] Intelligent Computing and Machine Learning Lab, School of ASEE, Beihang University, Beijing, P.R. China
{hzhao,zcqin}@buaa.edu.cn

[2] École d'Ingénieur Généraliste, École Centrale de Pékin, Beihang University, Beijing, P.R. China

Abstract. Clustering analysis is well-used in data mining to group a set of observations into clusters according to their similarity, thus, the (dis)similarity measure between observations becomes a key feature for clustering analysis. However, classical clustering analysis algorithms cannot deal with observation contains both data and vague concepts by using traditional distance measures. In this paper, we proposed a novel (dis)similarity measure based on a prototype theory interpreted knowledge representation framework named label semantics. The new proposed measure is used to extend classical K-means algorithm for clustering data instances and the vague concepts represented by logical expressions of linguistic labels. The effectiveness of proposed measure is verified by experimental results on an image clustering problem, this measure can also be extended to cluster data and vague concepts represented by other granularities.

Keywords: Clustering · Label semantics · Prototype theory · K-means · Distance measure

1 Introduction

Clustering analysis (or clustering) is a main task of exploratory data mining, and a common technique for statistical data analysis [11]. Cluster analysis groups a set of observations (data) into several "clusters", and observations in the same "cluster" are considered as "similar" observations and they are "dissimilar" to those belong to other clusters. Besides, conceptual clustering is another type of clustering analysis for unsupervised classification, in which the observations are grouped according to their fitness to descriptive concepts, but not simple similarity measures. To our knowledge, these two types of clustering are rarely studied together though there are actual needs for grouping data and concepts [4].

Clustering algorithms are widely used in several fields, including machine learning, pattern recognition, image analysis, bioinformatics and so on. There are many successful classical cluster algorithms, such as K-means, fuzzy C-means

V.-N. Huynh et al. (Eds.): IUKM 2015, LNAI 9376, pp. 236–246, 2015.
DOI: 10.1007/978-3-319-25135-6_23

and rough C-means [1] for similarity based clustering, and hierarchical clustering algorithms for connectivity based clustering. Yet these classical clustering algorithms using classical distance measures (e.g. Euclidian and Mahalanobis distance) cannot cluster *vague concepts*, and their clustering results are heavily depend on the distance measure between observations. Thus, in past decade, many clustering algorithms using customized distance measure are proposed in literature, for example, belief K-modes method (BKM) proposed by Hariz et al. [2] and possibilistic K-modes method (PKM) proposed by Ammar and Elouediare [3] are effective methods for clustering numerical data described by categorical attributes (labels), where the distance measure between objects is defined by the total mismatches of the corresponding attribute.

However, these clustering algorithms are restricted to numerical or discrete data. However, in order to simulate the knowledge generation process, we hope to deal with clustering some high-level knowledge, vague concepts or linguistic expressions. For example, we have two sets of observations, including a set of data of human heights in meters

$$hight = \{1.0, 1.3, 1.4, 1.6, 1.7, 1.9, 2.0\}$$

and a set of descriptive vague concepts $concepts = \{short, medium, tall\}$ in which elements are defined by a set of prototypical elements. Given the numbers of cluster centers $k = 3$, these observations can be clustered into three following clusters: $\{short, 1.0, 1.3, 1.4\}, \{medium, 1.6, 1.7\}, \{tall, 1.9, 2.0\}$. In order to accomplish the above purpose, we need a suitable distance measure for measuring the dissimilarity between numerical data and descriptive vague concepts. Label semantics [5] can be used to construct distance measure between numerical data and descriptive concepts, where descriptive concepts are represented by a set of linguistic labels, Zhang and Qin [4] proposed such a distance measure based on fuzzy set interpreted label semantics, where linguistic labels are represented by fuzzy membership functions defined on a universe of discourse containing data to be described. The prototype theory based interpretation of label semantics is proposed by Lawry and Tang [6], where linguistic labels are represented by a set of prototypical data. Based on this interpretation, in this paper, we proposed a novel distance measure which makes it possible to cluster a set of observations including numerical data, descriptive concepts and linguistic expressions, and it's effectiveness is verified by applying it to the classical K-means clustering algorithm.

This paper is structured as the following. Section 2 gives a general introduction of label semantics. In Section 3, we propose the new distance measure based on prototype theory. Section 4 gives the extended K-means based on the new measure and Section 5 gives the experimental results and compared to previous research. Section 6 gives the final conclusion and discussions.

2 Label Semantics Framework

Label semantics [5] is a random set framework for modeling with words, which encodes the semantic meaning of linguistic labels according to how they are used by a population of individuals to convey information. Otherwise, it can also be regarded as a simulation of knowledge generation process, in order to acquire knowledge, an intelligent has to identify which label or logical expression is appropriate to describe a value or an observation, thus, the appropriateness of using a subset of labels to describe a certain object is named *appropriateness degrees* in label semantics framework.

Definition 1. *(Label expression). Given a finite set of labels $LA = \{L_1, ..., L_n\}$, the set of label expression LE is generated by logical expression of labels in LA as below:*

- *For* $\forall L \in LA, \quad$ *we have* $L \in LE$
- *For* $\forall(\theta, \varphi) \in LA^2,$ *we have* $(\theta \vee \varphi, \theta \wedge \varphi, \neg\theta) \in LE$

For set of labels $S \subseteq LA$, an observation in the universe of discourse $x \in \Omega$ when an individual in a population $I \in V$ makes an assertion of the form "x *is* θ", which provides information about "what label is appropriate for describing observation x", this information is named label description of x, it is a random set from a population V to the power set of LA, denoted by D_x, the associated distribution of D_x is referred to mass assignment, denoted by m_x as follow:

Definition 2. *(Mass Assignment). Mass assignment is agent's subjective belief in a population V that the subset S contains all and only appropriate label(s) for describing object x:*

$$\forall S \subseteq LA,\ m_x(S) = P(I \in V:\ D_x^I = S) \tag{1}$$

Thus, the mass assignment m_x can be also regarded as a mass function defined as $m_x : P(LA) \rightarrow [0, 1]$ where $P(LA)$ is the power set of LA and

$$\sum_{S \subseteq LA} m_x(S) = 1$$

Furthermore, to evaluate the how appropriate a single label $L \in LA$ is for describing a certain observation $x \in \Omega$, the appropriateness degree is defined as follows:

Definition 3. *(Appropriateness Degree). Appropriateness degree is a function defined as $\mu : LA \times \Omega \rightarrow [0, 1]$ satisfying:*

$$\forall x \in \Omega,\ \forall L \in LA,\ \mu_L(x) = \sum_{S \subseteq LA : L \in S} m_x(S) \tag{2}$$

Example 1. Given a finite set of labels for human age description: $LA_{Age} = \{young, middle\text{-}aged, old\}$ and a population of 10 individuals, Suppose 4 of 10 individuals consider that "young" is appropriated label for describing age 42, and other 6 support that both "young" and "middle-aged" are appropriate labels, according to Definition 2, the mass assignment for age 42 is:

$$m_{42} = \{middle - aged\} : 0.4, \{young, middle - aged\} : 0.6$$

Based on Definition 3, appropriateness degrees of each label for describing age 42 are:

$$\mu_{young}(42) = 0.6 \quad \mu_{middle-aged}(42) = 0.4 + 0.6 = 1$$

After defining the appropriateness degree evaluation method of single label, we may also interest in evaluating the appropriateness degree of a logical expression $\theta \in LE$, for this propose, it is necessary to identify what information is provided by a logical expression θ regarding the appropriateness of labels, thus, the λ function is defined to transform the information provided by a logical expression as below:

Definition 4. *(λ-Function). λ-function is a mapping from linguistic expression to the power set of labels: $\lambda : LE \rightarrow P(LA)$, which is defined as follow, for $\forall(\theta, \varphi) \in LE^2$:*

- $\forall L_i \in LA, \quad \lambda(L_i) = \{F \subseteq LA : \ L_i \in F\}$
- $\lambda(\theta \wedge \varphi) = \lambda(\theta) \cap \lambda(\theta)$
- $\lambda(\theta \vee \varphi) = \lambda(\theta) \cup \lambda(\theta)$
- $\lambda(\neg\theta) = \overline{\lambda(\theta)}$

Label semantics theory is a powerful tool for modeling with words, which has been well applied in machine learning and data mining, further details on using label semantics for data mining are available in [6].

3 Distance Measure Based on Logical Expressions

3.1 Prototype Theory Interpretation of Label Semantics

The proposed distance measure deals with labels and linguistic expressions interpreted by the prototype theory interpretation of label semantics framework. The label semantics framework is a random set framework for modeling with vagueness, where a set of labels is used by individuals vary across a population, such a theory cannot result in a truth-functional calculation [6]. In order to generate a functional calculus for appropriateness degrees, Lawry and Tang [6] have proposed an interpretation based on prototype theory. In this interpretation, each label $L_i \in LA$ is represented by a set of prototypical elements $P_i \in \Omega$, given a classical distance function $d(\cdot)$ define on the universe of discourse: $d : \Omega^2 \rightarrow [0, \infty)$, and δ is a probability density function which is defined on $[0, \infty)$, in our experiment, we consider d as the Euclidean distance. The appropriateness degree $\mu_{L_i}(x)$ of describing a data $x \in \Omega$ by using a certain label $L_i \in LA$ can be calculated as below:

$$\forall L_i \in LA, \ \forall x \in \Omega, \ \mu_{L_i}(x) = \int_{d(x,P_i)}^{\infty} \delta(t)dt \tag{3}$$

where $d(x, P_i) = min\{d(x, y) : \ \forall y \in P_i\}$. More details on the prototype theory interpretation of label semantics can be found in [6,9].

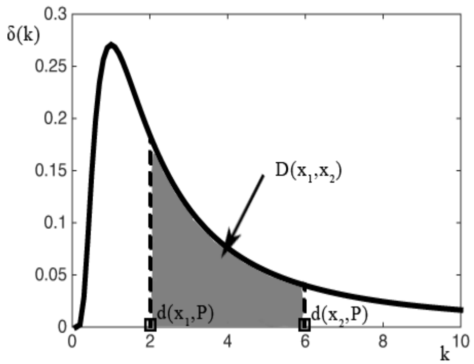

Fig. 1. Illustration of distance between two data points.

3.2 Distance between Vague Concepts

During the decision making process, an individual has to identify which label or logical expression can actually be used to describe an observation. The prototype theory interpretation of label semantics generates a functional calculus of appropriateness degree, thus, we can propose a measure based on appropriateness degrees to calculate dissimilarities between two observations of labels and logical expressions.

Definition 5. *(Distance Between Data Points). Given two observations in a universe Ω, and N labels $L_i \in LA$, $i \in [1, |LA|]$ and $i \in \mathbb{Z}$, for each label $L_i \in LA$, let there is a set $P_i \subseteq \Omega$ corresponding to prototypical elements for which L_i is certainly an appropriate description [6]. The distance between two observations (data points) is defined as a function define as $D(x_1, x_2) : \Omega^2 \to [0, \infty)$:*

$$D(x_1, x_2) = \sum_{i=1}^{N} |\mu_{L_i}(x_1) - \mu_{L_i}(x_2)| \tag{4}$$

where $\mu_{L_i}(x_j)$ is the appropriateness degree of describing data point x_j using label L_i, as defined in Definition 3. Given a single label L which is represented by a set of prototypical elements $P \in \Omega$, an illustration of distance between two data points is shown in Fig. 1, where the distance is defined as the integral of the density function δ from $d(x_1, L)$ to $d(x_2, L)$. Further more, the above distance measure has these following properties:

Theorem 1. *(Symmetric). Given $(x_1, x_2) \in \Omega^2$, the distance between two data points is symmetric*

$$D(x_1, x_2) = D(x_2, x_1)$$

Proof. According to Definition 1, for each $(x_1, x_2) \in \Omega^2$ we have

$$D(x_1, x_2) = \sum_{i=1}^{N} |\int_{d(x_1,P_i)}^{d(x_2,P_i)} \delta(t)dt| = \sum_{i=1}^{N} |\int_{d(x_2,P_i)}^{d(x_1,P_i)} \delta(t)dt| = D(x_2, x_1)$$

The proof is completed. The proof of this theorem is very intuitive as the distance is defined by the area between a range, it is symmetric as the area keeps the same from either the left to the right or from the right to the left.

Theorem 2. *(Triangular inequality). Given $(x_1, x_2, x_3) \in \Omega^3$, we have*

$$D(x_1, x_3) \leq D(x_1, x_2) + D(x_2, x_3)$$

Proof. According to Definition 1, for each $(x_1, x_2, x_3) \in \Omega^3$ we have

$$D(x_1, x_3) = \sum_{i=1}^{N} |\int_{d(x_1,P_i)}^{d(x_3,P_i)} \delta(t)dt| = \sum_{i=1}^{N} |\int_{d(x_1,P_i)}^{d(x_2,P_i)} \delta(t)dt + \int_{d(x_2,P_i)}^{d(x_3,P_i)} \delta(t)dt|$$
$$D(x_1, x_2) + D(x_1, x_2) = \sum_{i=1}^{N} |\int_{d(x_1,P_i)}^{d(x_2,P_i)} \delta(t)dt| + |\int_{d(x_2,P_i)}^{d(x_3,P_i)} \delta(t)dt|$$

According to the triangular inequality in the real number space where $\forall(a, b) \in \mathbb{R}^2$, $|a + b| \leq |a| + |b|$, as a result, for $\forall(x_1, x_2, x_3) \in \Omega^3$ and $\forall P_i \subseteq \Omega$ we have:

$$|\int_{d(x_1,P_i)}^{d(x_2,P_i)} \delta(t)dt + \int_{d(x_2,P_i)}^{d(x_3,P_i)} \delta(t)dt| \leq |\int_{d(x_1,P_i)}^{d(x_2,P_i)} \delta(t)dt| + |\int_{d(x_2,P_i)}^{d(x_3,P_i)} \delta(t)dt|$$

As a result, for $\forall(x_1, x_2, x_3) \in \Omega^3$:

$$D(x_1, x_3) \leq D(x_1, x_2) + D(x_2, x_3)$$

In conclusion, the distance between data points follows the triangular inequality.

Above definitions construct a functional calculus for measuring dissimilarity between two data points referring to labels which are represented by sets of prototypes defining on the universe of discourse. One step further, we consider how can we measure the dissimilarity between a certain label and a data point in the same universe.

Definition 6. *(Distance between point and label). Given a data point $x \in \Omega$ and a certain label $L_i \in LA$ represented by a set of prototypical elements $P_i \subseteq \Omega$, the distance between point and label is defined as below:*

$$D(x, L_i) = min\{D(x, y),\ \forall y \in P_i\} \tag{5}$$

where $D(x, y)$ is the distance between points as defined in Definition 1.

Specifically, when there is only one label $L \in LA$, $|LA| = 1$ which can be used to describe elements in Ω,we have:

$$D(x, L) = 1 - \mu_L(x) \tag{6}$$

Where $\mu_L(x)$ is the appropriateness degree of describing data point x using label L, thus, the distance $D(x, L)$ can be interpreted as the probability of "label L can not be used to describe data x". Furthermore, the distance between two sets of labels can be defined by:

Definition 7. *(Distance between set of labels). Given two sets of labels $(S_1, S_2) \in LA^2$, each label $L_i \in LA$ can be represented by a set of prototypical elements $P_i \in \Omega$, we have:*

$$D(S_1, S_2) = \frac{\sum_{L_i \in S_1} \sum_{L_j \in S_2} min\{D(x, y),\ \forall (x, y) \in P_i \times P_j\}}{|S_1| \cdot |S_2|} \tag{7}$$

where $|S_1|$ and $|S_2|$ are cardinalities of sets S_1 and S_2. $D(x, y)$ is the distance between points as defined by Definition 1. Based on the properties of distance between data points, it is obviously that the distance between set of labels is also symmetric and satisfies the triangular inequality.

The above distance measure is one dimensional, for an object with more than one feature to be described by labels. The distance measure between set of labels can be extended into multi-dimensional as shown in Definition 8.

Definition 8. *(Distance between multi-dimensional set of labels). The set of n-dimensional labels $MLA^{(n)}$ is a combination of descriptive labels of n different features $MLA(n) = LA_1 \times LA_2 \times ... \times LA_n$, where LA_i is the set of descriptive labels for describing the i^{th} feature. For two multi-dimensional labels $(ML_1, ML_2) \in MLA(n)^2$ where:*

- $ML_1 = (L_{11}, L_{12}, ..., L_{1n}),\ L_{1i} \in LA_i$
- $ML_2 = (L_{21}, L_{22}, ..., L_{2n}),\ L_{2i} \in LA_i$

we have:

$$D(ML_1, ML_2) = \sqrt{\sum_{i=1}^{n} D(L_{1i}, L_{2i})^2} \tag{8}$$

In Definition 4, the λ-function provides an application from logical expressions to set of labels, utilizing this function and distance measure between set of labels, we can define the distance between logical expressions intuitively:

Definition 9. *(Distance between logical expressions). Given two logical expressions $(\theta, \varphi) \in LE$, the distance between θ and φ is:*

$$D(\theta, \varphi) = D(\mathbb{S}^{\theta \wedge \neg \varphi}, \mathbb{S}^{\varphi}) + D(\mathbb{S}^{\varphi \wedge \neg \theta}, \mathbb{S}^{\theta}) \tag{9}$$

$D(\theta, \lambda)$ is the distance between label sets as defined in Definition 7, where label sets $\mathbb{S}^{\theta \wedge \neg \varphi}, \mathbb{S}^{\varphi}, \mathbb{S}^{\varphi \wedge \neg \theta}, \mathbb{S}^{\theta}$ are defined as follow:

- $\mathbb{S}^{\theta} = \{S | S \in \lambda(\theta)\}$
- $\mathbb{S}^{\varphi} = \{S | S \in \lambda(\varphi)\}$
- $\mathbb{S}^{\theta \wedge \neg \varphi} = \{S | S \in \lambda(\theta) \cap \overline{\lambda(\varphi)})\}$
- $\mathbb{S}^{\varphi \wedge \neg \theta} = \{S | S \in \lambda(\varphi) \cap \overline{\lambda(\theta)})\}$

specifically, when $|(\mathbb{S}^{\theta \wedge \neg \varphi})| = 0$,

$$D(\theta, \varphi) = D(\mathbb{S}^{\varphi \wedge \neg \theta}, \mathbb{S}^{\theta}) \tag{10}$$

and when $|(\mathbb{S}^{\varphi \wedge \neg \theta})| = 0$,

$$D(\theta, \varphi) = D(\mathbb{S}^{\theta \wedge \neg \varphi}, \mathbb{S}^{\varphi}) \tag{11}$$

4 Clustering Mixed Objects

First proposed by MacQueen [7] in 1967, the K-means is regarded as the simplest yet effective technique for clustering analysis. The classical K-means algorithm using Euclidean distance cannot cluster vague concepts (e.g. linguistic descriptions). Based on the above distance measure, classical K-means algorithm can be extended for clustering mixed objects, including data points, labels which are represented by sets of prototypical elements, as defined in Section 3.1 and linguistic expressions.

The main objective of K-means clustering is to minimize the sum of squared distance between objects in each cluster and their mean, given objects

$$(x_1, x_2 ... x_N) \in \Omega^N$$

and k clusters, and let m_j as the mean of objects in cluster j, we define $x \in j$ if $m_j = \{m| \ min||x, m_j||, \ \forall j \in [1, k], \ j \in \mathbb{N}\}$, which is also equivalent to minimizing the following objective function:

$$S = \sum_{j=1}^{k} \sum_{x \in k} ||x, m_j||^2 \tag{12}$$

With the same objective, given an unlabeled data set of mixed objects $(obj_1, ..., obj_N) \in (\Omega \bigcup LE)^N$ the extended K-means algorithm for clustering mixed objects can be described as pseudo-codes in Table 1.

Table 1. Pseudo-code of extended K-means algorithm for clustering mixed objects.

Given a finite set of mixed objects $\mathbb{S} = \{obj_1 ... obj_N\}$ and a number of cluster k,
A set of randomly initialized centers $\mathbb{K} = \{c_1, ..., c_k\}$, a threshold $\varepsilon > 0$, and counter p

While $||c^{(p)} - c^{(p-1)}|| > \varepsilon$
p++

Step1. For each object $obj_i \in \mathbb{S}$, determine the cluster $obj_i \leftarrow c_i^{(p-1)}$ of each object, if:
$D(obj_i, c_i^{(p-1)}) = min\{D(obj_i, c_t^{(p-1)}) : t = 1, ..., k\}$

Step2. Calculate new clusters $c_t^{(p)}$, $x = 1, ..., k$, which satisfy:
$\sum_{obj \in cluster \ t} D(c_x^{(p)}, obj) = min\{\sum_{obj \in cluster \ t} D(x, obj)\}$

The new proposed distance measure is used both in Step 1 and 2 to calculate distances between each two objects, thus, this algorithm can be used to cluster mixed objects including numerical data and linguistic labels, necessarily, the cluster centers of each cluster should be numerical data in implement.

5 Experimental Studies

5.1 Distance Variation

Given a continue universe of discourse $[1, 20]$ of numerical data points defined on $\mathbb{R}$, and three labels L_1, L_2, L_3 which can be represented respectively by three sets of prototypical elements, $P_1 = \{1\}$, $P_2 = \{5, 5.5, 7\}$ and $P_3 = \{8, 8.5\}$. Given fixed $x_1 = 7.5$, when x_2 varies from 0 to 20, the variation of distance $D(x_1, x_2)$ is shown illustratively in Fig. 2.

This illustration indicates that the distance $D(x_1, x_2)$ varies rapidly when the data point x_2 is close to prototypes, in contrast, the variation becomes more and more slowly when the data point x_2 moves away from the prototypes. This phoneme can be interpreted as when data points are close to a linguistic concept, we can determine their dissimilarity according to the appropriateness of describing these objects using this concept more precisely than these data points are far away from this concept.

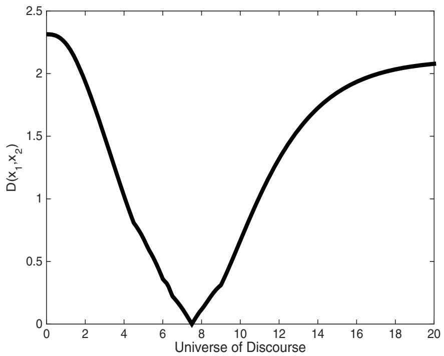

Fig. 2. Illustration of variation of distance between two data points: x_1 is fixed to 7.5 and x_2 varies from 0 to 20.

5.2 Clustering Images and Labels

In order to validate the performance of the novel distance measure for clustering images and vague concepts. We apply this measure in an extended K-means algorithm as introduced in Section 4 for clustering images and linguistic labels. We select 100 images from the Corel image data set[8] in 4 categories and 25 images in each category, each image is resized into 192 × 128, we chose 4 descriptive linguistic labels to describe images, including "sunset", "beach", "garden" and "polar bear". In our experiment, each image is represented as a 3-dimensional numerical data point according to its average HSV (Hue, Saturation, Value) [10] feature, besides, each label is represented as a set of 5 images which are randomly selected from the same category.

After designing labels and extracting image features, the data set of mixed objects including images represented as numerical data points and labels (as defined in Definition 8) represented as sets of 5 prototypical elements (images) for which the label is certainly an appropriate description. In our experiment, each label is represented by 5 images in same category, thus, the set which we have to cluster is constructed by 80 images (20 images in each category) and 4 labels, we cluster this set of mixed objects into 4 clusters. In our experiment, the 4 cluster centers are 4 data points which are randomly selected from the components of the 4 mutually different labels. Furthermore, a 20% cross-validation is used to evaluate the performance of proposed algorithm, each time we randomly change the 20 images (5 images of each label) which are considered as prototypical elements for representing labels, during the cross-validation process, each image is used as a component of label only once. The average accuracy of five times of experiment is regarded as the final experiment result, the comparison of accuracy between this method and the existing method proposed by Zhang and Qin [4], and their execution time to build the 4 clusters under the same hardware condition are shown in Table 2.

Further more, the illustration of above result and its variation is shown in Fig. 3.

Table 2. Performance of clustering mixed objects in terms of classification accuracy.

	Sunset	Beach	Garden	Polar-bear	Execution time
Our Model	81%	94%	67%	87%	18.2s
Zhang and Qin [4]	72%	60%	64%	96%	6318.1s

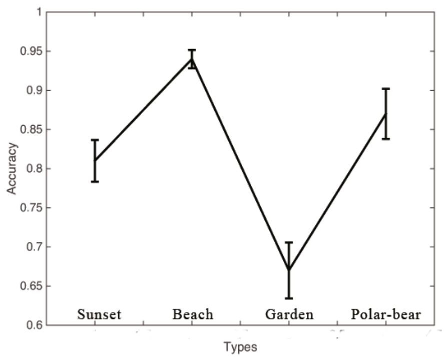

Fig. 3. Illustration of experimental result and its variation.

6 Conclusion

In this paper we proposed a novel distance measure based on prototype theory interpreted label semantics framework. This distance measure differs from the other distance measure by focusing on the difference of logical meanings which conveyed by the object. The new proposed distance measure is applicated to extend classical K-means algorithm for clustering numerical data and vague concepts which are in the form of linguistic labels. Experimental studies on a image clustering problem validated the effectiveness of our new proposed measure.

With a similar idea of measuring the dissimilarities according to the logical meaning which the object conveys, the proposed measure is extendable to measuring distances between any granularities, In future work, This measure can be applied to other applications and clustering vague concepts represented by other granularities.

Acknowledgments. This work is supported by the National Science Foundation of China No. 61305047.

References

1. Momin, B., Yelmar, P.: Modifications in K-Means Clustering Algorithm. International Journal of Soft Computing and Engineering (IJSCE) **2**(3) (2012)
2. Ben Hariz, S., Elouedi, Z., Mellouli, K.: Clustering Approach using belief function theory. In: Euzenat, J., Domingue, J. (eds.) AIMSA 2006. LNCS (LNAI), vol. 4183, pp. 162–171. Springer, Heidelberg (2006)
3. Ammar, A., Elouedi, Z.: A new possibilistic clustering method: the possibilistic K-modes. In: Pirrone, R., Sorbello, F. (eds.) AI*IA 2011. LNCS, vol. 6934, pp. 413–419. Springer, Heidelberg (2011)
4. Zhang, W., Qin, Z.: Clustering data and imprecise concepts. In: 2011 IEEE International Conference on Fuzzy Systems (FUZZ), pp. 603–608 (2011)
5. Lawry, J.: A framework for linguistic modelling. Artificial Intelligence **155**(1), 1–39 (2004)
6. Lawry, J., Tang, Y.: Uncertainty modelling for vague concepts: A prototype theory approach. Artificial Intelligence **173**(18), 1539–1558 (2009)
7. MacQueen, J.: Some methods for classification and analysis of multivariate observations. In: Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 281–297. University of California Press (1967)
8. He, X., Zemel, R., Carreira-Perpindn, M.: Multiscale conditional random fields for image labeling. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, p. 695 (2004)
9. Qin, Z., Tang, Y.: Uncertainty Modeling for Data Mining: A Label Semantics Approach. Springer (2014)
10. Smith, A.R.: Color gamut transform pairs. Computer Graphics **12**(3), 12–19 (1978)
11. Wikipedia. Cluster analysis. http://en.wikipedia.org/wiki/Cluster_analysis