New Model Distances and Uncertainty Measures for Multivalued Logic

Alexander Vikent'ev[1,2](✉) and Mikhail Avilov[2]

[1] Sobolev Institute of Mathematics of the Siberian Branch of the Russian Academy of Sciences, Acad. Koptyug Avenue 4, 630090 Novosibirsk, Russia
vikent@math.nsc.ru

[2] Novosibirsk State University, Pirogova Str. 2, 630090 Novosibirsk, Russia
mikhail8avilov@gmail.com

Abstract. There is an increasing importance of problems regarding the analysis of propositions of experts and clustering of information contained in databases. Propositions of experts can be presented as formulas of n-valued logic L_n. This paper is concerned with defining metrics and degrees of uncertainty on formulas of n-valued logic. After metrics and degrees of uncertainty (as well as their useful properties) have been established, they are used for the cluster analysis of the sets of n-valued formulas. Various clustering algorithms are performed and obtained results are analyzed. Established methods can be further employed for experts propositions analysis, clustering problems and pattern recognition.

Keywords: Distances on formulas of n-valued logic · Metrics · Uncertainty measures · Cluster analysis · Pattern recognition

1 Introduction

The problem of ranking the statements of experts according to their informativeness and inducing the metric on the space of statements was introduced by Lbov and Zagoruiko in the early 1980s [5,8,12]. Expert statements were written in the form of logical formulas. Thus, the task of comparing and ranking the statements of experts turned into a task of comparing and ranking the logical formulas. In order to do this, the distance between the formulas and the uncertainty measure of the formulas were introduced.

Lbov and Vikent'ev used the normalized symmetric difference of the formulas models as a distance for the case of two-valued logic [3]. Then the model-theoretic approach to the analysis of multivalued formulas was proposed [1,2,4]. Formulas belong to the n-valued logic L_n [10]. Using the theory of models for n-valued formulas, the different versions of the distances and uncertainty measures were introduced and the properties of those quantities were established [4]. Then, to rank the statements of experts the clustering analysis of the finite sets of formulas of n-valued logic was performed based on the introduced distances and uncertainty measures. Clustering was performed for small n, and the utilized distance had constant weights [9].

C. Dichev and G. Agre (Eds.): AIMSA 2016, LNAI 9883, pp. 89–98, 2016.
DOI: 10.1007/978-3-319-44748-3_9

In this paper we introduce a new distance, which generalizes the distances used for multivalued logical formulas previously. The metric properties of the new distance are proven and it is shown that there is a continuum of such distances. A new uncertainty measure of multivalued logical formulas is also introduced. Using those new quantities the clustering analysis of finite sets of the formulas of n-valued logic L_n is performed. The results of various clustering algorithms are obtained for different dimensions n of logic L_n.

2 Definitions and Notations

In this section we define n-valued logic L_n and some useful model-theoretic properties.

Definition 1. *Propositional language L consists of the following propositional symbols:*

1. x, y, z, ... – propositional variables;
2. $\neg$, $\rightarrow$ – propositional logical connectives;
3. (,) – auxiliary symbols.

Definition 2. *Formulas are the finite sequences of propositional symbols defined the following way:*

1. x, y, z, ... – elementary formulas;
2. If φ and ψ are formulas, then $\neg\varphi$, $\varphi \rightarrow \psi$ are formulas;
3. No other finite sequences of propositional symbols, except those mentioned in 1, 2, are formulas.

Now we can introduce n-valued logic L_n.

Definition 3. *N-valued logic L_n is defined the following way:*

$M_n = \langle V_n, \neg, \rightarrow, \{1\}\rangle$ *– n-valued Lukasiewicz matrix ($n \in N, n \geqslant 2$);*
$V_n = \left\{0, \frac{1}{n-1}, ..., \frac{n-2}{n-1}, 1\right\}$ *– set of truth values;*
$\neg : V_n \rightarrow V_n,$ *– unary negation operation;*
$\rightarrow: V_n \times V_n \rightarrow V_n$ *– binary implication operation;*
$\{1\}$ *– selected value of truth.*

Let us now introduce other logical connectives for the truth values of n-valued logic L_n.

Definition 4. *Logical connectives on V_n are defined from the input connectives:*

$\neg x = 1 - x$ *– negation;*
$x \rightarrow y = min\{1, 1 - x + y\}$ *– implication;*
$x \vee y = (x \rightarrow y) \rightarrow y = max\{x, y\}$ *– disjunction;*
$x \wedge y = \neg(\neg x \vee \neg y) = min\{x, y\}$ *– conjunction.*

We also formulate model-theoretic properties and notations that will be used further in this paper.

Definition 5. *Let Σ be the finite set of formulas of L_n. The set of propositional variables $S(\varphi)$ used for writing a formula φ of n-valued logic L_n is called the support of a formula ϕ.*

$S(\Sigma) = \cup_{\varphi \in \Sigma} S(\varphi)$ is called the support of the set Σ.

Definition 6. *A model M is a subset of attributed variables. $P(S(\Sigma))$ is the set of all models.*

We use the notation $\varphi_{\frac{k}{n-1}}$ if the formula φ has the value $\frac{k}{n-1}, k = 0, ..., n-1$ on a model M.

$M(\varphi_{\frac{k}{n-1}})$ is the number of models on which formula φ has the value $\frac{k}{n-1}$.

$M(\varphi_{\frac{k}{n-1}}, \psi_{\frac{l}{n-1}})$ is the number of models on which φ has the value $\frac{k}{n-1}$ and ψ has the value $\frac{l}{n-1}$.

Notes and Comments. The cardinality of the set of models $P(S(\Sigma))$ is

$$\mid P(S(\Sigma)) \mid = n^{|S(\Sigma)|}. \tag{1}$$

The proof of this fact as well as the definition of the logical connectives on models and other auxiliary statements are detailed in papers [1,4].

3 Model Distances and Uncertainty Measures

Let us show how to introduce a distance between formulas φ and ψ of n-valued logic L_n. It is natural to assume, that the less the absolute difference between the values of the formulas is, the closer those formulas are. So we will multiply the number of models with the same absolute difference values by the coefficient which considers the proximity of the values of the formulas. Those coefficients used to be precisely the truth values of L_n, so the model distance ρ_0 between formulas φ and ψ of n-valued logic L_n used to be defined as the normalized quantity $\tilde{\rho_0}$ [9]:

$$\rho_0(\varphi, \psi) = \frac{1}{n^{|S(\Sigma)|}} \cdot \sum_{k=0}^{n-1} \sum_{l=0}^{n-1} \frac{\mid k - l \mid}{n-1} \cdot M(\varphi_{\frac{k}{n-1}}, \psi_{\frac{l}{n-1}}). \tag{2}$$

Example 1. Let $n = 5$. Then $\rho_0(\varphi \wedge \psi, \varphi \vee \psi) = 0.4$, $\rho_0(\varphi \wedge \psi \wedge \chi \wedge \omega, \varphi \rightarrow \omega) = 0.2576$.

The particularity of the quantity (2) is that its coefficients (weights) $\frac{|k-l|}{n-1}$ are constant. This particularity does not allow to adjust the weight of the quantity $M(\varphi_{\frac{k}{n-1}}, \psi_{\frac{l}{n-1}})$ (the number of models with the same absolute difference values) to be properly included in the final distance ρ_0.

The uncertainty measure I_0 of formula φ of n-valued logic L_n used to be defined as follows [9]:

$$I_0(\varphi) = \rho_0(\varphi, 1) = \sum_{i=0}^{n-2} \frac{n-1-i}{n-1} \cdot \frac{M(\varphi_{\frac{i}{n-1}})}{n^{|S(\Sigma)|}}. \tag{3}$$

Example 2. Let $n = 5$. Then $I_0(\varphi \to \psi) = 0.2$, $I_0(\varphi \vee \psi \vee \chi \vee \omega) = 0.1416$.

The quantity (3) possesses the same particularity as the quantity (2): its weight $\frac{n-1-i}{n-1}$ is constant. This particularity also does not allow to adjust the weight of the quantity $M(\varphi_{\frac{i}{n-1}})$ to be properly included in the final uncertainty measure $I_0(\varphi)$.

Let us modify the quantities above to get rid of the particularities associated with the rigid structure of the weights. In order to do this, we substitute constant weights of the quantities ρ_0 and I_0 for arbitrary acceptable weights.

Definition 7. *The model distance between formulas φ and ψ of n-valued logic L_n, $S(\varphi) \cup S(\psi) \subseteq S(\Sigma)$ on $P(S(\Sigma))$ is called the quantity*

$$\rho(\varphi, \psi) = \frac{1}{n^{|S(\Sigma)|}} \cdot \sum_{k=0}^{n-1} \sum_{l=0}^{n-1} \lambda_{|k-l|} \cdot M(\varphi_{\frac{k}{n-1}}, \psi_{\frac{l}{n-1}}) \tag{4}$$

$$0 = \lambda_0 \leqslant \lambda_1 \leqslant ... \leqslant \lambda_{n-1} = 1;$$

$$n \geqslant 2.$$

Definition 8. *The uncertainty measure I of formula φ of n-valued logic L_n, $S(\varphi) \cup S(\psi) \subseteq S(\Sigma)$ on $P(S(\Sigma))$ is called the quantity*

$$I(\varphi) = \sum_{i=0}^{n-2} \alpha_i \cdot \frac{M(\varphi_{\frac{i}{n-1}})}{n^{|S(\Sigma)|}} \tag{5}$$

$$0 \leqslant \alpha_i \leqslant 1;$$

$$\alpha_k \geqslant \alpha_i \, \forall k \leqslant i;$$

$$\alpha_i + \alpha_{n-1-i} = 1 \, \forall i = 0, ..., \frac{n-1}{2}.$$

Notes and Comments. We got a new continuum of distances, possessing the properties of metrics.

In Definition 8 the coefficients α of the uncertainty measure I actually depend on the coefficients λ of the model distance ρ because the uncertainty measure is itself a distance. Moreover,

$$I(\varphi) = \rho(\varphi, 1). \tag{6}$$

So the given definition of the uncertainty measure for n-valued logic corresponds with the earlier ideas of Lbov and Bloshitsin for two-valued logic [8].

Let us now establish some of the properties of the model distance (4) and uncertainty measure (5). We start with the properties of the distance ρ introduced in Definition 7.

Theorem 1. *For any formulas φ, ψ, χ of Σ the following assertions hold:*

1. $0 \leqslant \rho(\varphi, \psi) \leqslant 1$;
2. $\rho(\varphi, \psi) = 0 \Leftrightarrow \varphi \equiv \psi$;
3. $\rho(\varphi, \psi) = \rho(\psi, \varphi)$;
4. $\rho(\varphi, \psi) \leqslant \rho(\varphi, \chi) + \rho(\chi, \psi)$;
5. $\varphi \equiv \varphi_1, \psi \equiv \psi_1 \Rightarrow \rho(\varphi, \psi) = \rho(\varphi_1, \psi_1)$.

Proof. Each of the assertions will be treated separately.

1. The distance calculation formula involves all models with coefficients from 0 to 1. $\rho(\varphi, \psi) = 0$ when $\varphi \equiv \psi$ and $\rho(\varphi, \psi) = 1$ when $\varphi \equiv \neg\psi$. φ and ψ take only the values 0 and 1 on the models involved. So $0 \leqslant \rho(\varphi, \psi) \leqslant 1$.

2. Necessity follows from the proof of the previous assertion. Sufficiency follows from that, given the definition of equivalence, if $\varphi \equiv \psi$ then the values of φ and ψ are the same on all models. Then for $k = l$ every $M(\varphi_{\frac{k}{n-1}}, \psi_{\frac{l}{n-1}})$ in the formula $\rho(\varphi, \psi)$ is multiplied by 0 hence $\rho(\varphi, \psi) = 0$.

3. Symmetrical pairs $M(\varphi_{\frac{k}{n-1}}, \psi_{\frac{l}{n-1}}) \neq M(\varphi_{\frac{l}{n-1}}, \psi_{\frac{k}{n-1}})$ are multiplied by the same coefficient. So $\rho(\varphi, \psi) = \rho(\psi, \varphi)$.

4. Follows from the model-theoretic properties given in Sect. 2 and the paper [9].

5. Follows from the definition of equivalence of the two formulas [11]. □

Notes and Comments. Assertions 2–4 are the properties of the metric. This means we can define a metric on the equivalence classes of the formulas of L_n.

Let us now establish the properties of the uncertainty measure I introduced in Definition 8.

Theorem 2. *For any formulas φ, ψ of Σ the following assertions hold:*

1. $0 \leqslant I(\varphi) \leqslant 1$;
2. $I(\varphi) + I(\neg\varphi) = 1$;
3. $I(\varphi \wedge \psi) \geqslant max\{I(\varphi), I(\psi)\}$;
4. $I(\varphi \vee \psi) \leqslant min\{I(\varphi), I(\psi)\}$;
5. $I(\varphi \wedge \psi) + I(\varphi \vee \psi) \geqslant I(\varphi) + I(\psi)$.

Proof. Each of the assertions will be treated separately.

1. $I(\varphi) = \rho(\varphi, 1)$ hence $0 \leqslant I(\varphi) \leqslant 1$.

2. $$I(\varphi) + I(\neg\varphi) = \alpha_0 \cdot \frac{M(\varphi_0)}{n^{|S(\Sigma)|}} + \alpha_{n-1} \cdot \frac{M(\varphi_1)}{n^{|S(\Sigma)|}} + \sum_{i=1}^{n-2} (\alpha_i + \alpha_{n-1-i}) \cdot \frac{M(\varphi_{\frac{i}{n-1}})}{n^{|S(\Sigma)|}} =$$
$$= \frac{|\ P(S(\Sigma))\ |}{n^{|S(\Sigma)|}} = 1.$$

3. $$I(\varphi \wedge \psi) = \sum_{i=0}^{n-2} \alpha_i \frac{M((\varphi \wedge \psi)_{\frac{i}{n-1}})}{n^{|S(\Sigma)|}} =$$
$$= \sum_{i=0}^{n-2} \alpha_i \left(\sum_{k=i}^{n-1} \left(\frac{M(\varphi_{\frac{k}{n-1}} \wedge \psi_{\frac{i}{n-1}})}{n^{|S(\Sigma)|}} + \frac{M(\varphi_{\frac{i}{n-1}} \wedge \psi_{\frac{k}{n-1}})}{n^{|S(\Sigma)|}} \right) \right) - \alpha_i \frac{M(\varphi_{\frac{i}{n-1}} \wedge \psi_{\frac{i}{n-1}})}{n^{|S(\Sigma)|}}.$$

$$I(\varphi)=\sum_{i=0}^{n-2}\alpha_i\sum_{k=0}^{n-1}\frac{M(\varphi_{\frac{i}{n-1}}\wedge\psi_{\frac{k}{n-1}})}{n^{|S(\Sigma)|}}$$
$$=\sum_{i=0}^{n-2}\alpha_i\left(\sum_{k=i}^{n-1}\frac{M(\varphi_{\frac{i}{n-1}}\wedge\psi_{\frac{k}{n-1}})}{n^{|S(\Sigma)|}}+\sum_{k=0}^{i}\frac{M(\varphi_{\frac{i}{n-1}}\wedge\psi_{\frac{k}{n-1}})}{n^{|S(\Sigma)|}}\right)-\alpha_i\frac{M(\varphi_{\frac{i}{n-1}}\wedge\psi_{\frac{i}{n-1}})}{n^{|S(\Sigma)|}}.$$
$$I(\varphi\wedge\psi)-I(\varphi)=\sum_{i=0}^{n-2}\sum_{k=0}^{i}(\alpha_k-\alpha_i)\frac{M(\varphi_{\frac{i}{n-1}}\wedge\psi_{\frac{k}{n-1}})}{n^{|S(\Sigma)|}}+\sum_{i=0}^{n-2}\sum_{k=i}^{n-1}\alpha_i\frac{M(\varphi_{\frac{i}{n-1}}\wedge\psi_{\frac{k}{n-1}})}{n^{|S(\Sigma)|}}\geqslant$$
$$\geqslant 0.$$

So $I(\varphi\wedge\psi)\geqslant I(\varphi)$. Similarly we have an evaluation for ψ: $I(\varphi\wedge\psi)\geqslant I(\psi)$. Hence $I(\varphi\wedge\psi)\geqslant max\{I(\varphi),I(\psi)\}$.

4. $I(\varphi\vee\psi)=\sum_{i=0}^{n-2}\alpha_i\frac{M((\varphi\wedge\psi)_{\frac{i}{n-1}})}{n^{|S(\Sigma)|}}=$
$$=\sum_{i=0}^{n-2}\alpha_i\left(\sum_{k=i}^{n-1}\left(\frac{M(\varphi_{\frac{k}{n-1}}\wedge\psi_{\frac{i}{n-1}})}{n^{|S(\Sigma)|}}+\frac{M(\varphi_{\frac{i}{n-1}}\wedge\psi_{\frac{k}{n-1}})}{n^{|S(\Sigma)|}}\right)\right)-\alpha_i\frac{M(\varphi_{\frac{i}{n-1}}\wedge\psi_{\frac{i}{n-1}})}{n^{|S(\Sigma)|}}.$$
$$I(\varphi)=\sum_{i=0}^{n-2}\alpha_i\sum_{k=0}^{n-1}\frac{M(\varphi_{\frac{i}{n-1}}\wedge\psi_{\frac{k}{n-1}})}{n^{|S(\Sigma)|}}=$$
$$=\sum_{i=0}^{n-2}\alpha_i\left(\sum_{k=i}^{n-1}\frac{M(\varphi_{\frac{i}{n-1}}\wedge\psi_{\frac{k}{n-1}})}{n^{|S(\Sigma)|}}+\sum_{k=0}^{i}\frac{M(\varphi_{\frac{i}{n-1}}\wedge\psi_{\frac{k}{n-1}})}{n^{|S(\Sigma)|}}\right)-\alpha_i\frac{M(\varphi_{\frac{i}{n-1}}\wedge\psi_{\frac{i}{n-1}})}{n^{|S(\Sigma)|}}.$$
$$I(\varphi)-I(\varphi\vee\psi)=\sum_{i=0}^{n-2}\sum_{k=0}^{i}\alpha_i\frac{M(\varphi_{\frac{i}{n-1}}\wedge\psi_{\frac{k}{n-1}})}{n^{|S(\Sigma)|}}-\sum_{i=0}^{n-2}\sum_{k=0}^{i}\alpha_i\frac{M(\varphi_{\frac{k}{n-1}}\wedge\psi_{\frac{i}{n-1}})}{n^{|S(\Sigma)|}}\geqslant$$
$$\geqslant\sum_{i=0}^{n-2}\sum_{k=0}^{i}\alpha_i\frac{M(\varphi_{\frac{i}{n-1}}\wedge\psi_{\frac{k}{n-1}})}{n^{|S(\Sigma)|}}\geqslant 0.$$

So $I(\varphi\vee\psi)\leqslant I(\varphi)$. Similarly we have an evaluation for ψ: $I(\varphi\vee\psi)\leqslant I(\psi)$. Hence $I(\varphi\vee\psi)\leqslant min\{I(\varphi),I(\psi)\}$.

5. Follows from the equalities used in the proof of assertions 3 and 4. □

4 Clustering the Formulas of N-valued Logic

Clustering analysis is quite important while working with databases, statements of experts or performing statistical modeling [6]. For the sets of statements we know only the distances between the formulas and uncertainty measures of formulas. So two algorithms based on the distances were chosen for the clustering analysis – the hierarchic algorithm and the k-means algorithm. Those algorithms were adapted to work with multivalued formulas [7]. The complexity of computing the distance is exponential.

For the experiments below there was created a knowledge base, consisting of 300 multivalued formulas. The finite subsets of those formulas were randomly picked up for the clustering analysis. After that the value n for the logic L_n was chosen, the clustering algorithm was picked up and the weights λ were entered. Then the clustering analysis was performed utilizing either hierarchic algorithm or k-means algorithm. Both algorithms are based on new model distances and uncertainty measures introduced in Definitions 7 and 8 respectively. The results of the performed clusterings are presented in tables.

4.1 Hierarchic Algorithm, $n = 5$.

Let $n = 5$. Consider the set *test* 1, which consists of 8 formulas of five-valued logic L_5: $\varphi_1 = x \to y$, $\varphi_2 = \neg(x \to y)$, $\varphi_3 = (x \vee z) \to y$, $\varphi_4 = \neg((x \wedge y) \vee z) \to w$, $\varphi_5 = y \to (x \wedge z)$, $\varphi_6 = (\neg y \vee (x \to z))$, $\varphi_7 = z \to (x \vee y)$, $\varphi_8 = \neg((z \wedge y) \to x$.

We perform clustering analysis of the set *test* 1 using hierarchic algorithm. We choose the following weights: $\lambda_0 = 0, \lambda_1 = \frac{1}{4}, \lambda_2 = \frac{2}{4}, \lambda_3 = \frac{3}{4}, \lambda_4 = 1$ (the standard weights). Based on the distances matrix the minimal distance is $\rho_{4,6} = 0,0510$, so the first cluster is $\varphi_{4,6}$. After six more iterations the results of the performed clustering are presented in Table 1.

Table 1. Hierarchic algorithm, *test* 1, $n = 5$

Iteration	Δ	Clusters
0	0,0000	$\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_6, \varphi_5, \varphi_7, \varphi_8$
1	0,0508	$\varphi_1, \varphi_2, \varphi_3, \varphi_{4,6}, \varphi_5, \varphi_7, \varphi_8$
2	0,1000	$\varphi_{1,3}, \varphi_2, \varphi_{4,6}, \varphi_5, \varphi_7, \varphi_8$
3	0,1376	$\varphi_{1,3}, \varphi_2, \varphi_{4,6,7}, \varphi_5, \varphi_8$
4	0,1376	$\varphi_{1,3}, \varphi_2, \varphi_{4,6}, \varphi_{5,8}$
5	0,2092	$\varphi_2, \varphi_{5,8}, \varphi_{1,3,4,6,7}$
6	0,2092	$\varphi_2, \varphi_{1,3,4,5,6,7,8}$

To stop our clustering algorithm we use the quantity Δ – the maximal value of uncertainty measure among all elements of every cluster. For instance, if we set $\Delta = 0,1500$, then the algorithm stops after fourth iteration and gives 4 clusters as a result: $\varphi_{13}, \varphi_2, \varphi_{46}, \varphi_{58}$. We set $\Delta = 0,2100$, so the algorithm stops after sixth iteration giving 2 clusters as a result: $\varphi_2, \varphi_{1,3,4,5,6,7,8}$. If we do not stop the algorithm, then after seven iterations all 8 formulas merge into a single cluster.

4.2 K-means Algorithm, $n = 5$.

Let $n = 5$. Consider the set *test* 1, which consists of 8 formulas of five-valued logic L_5: $\varphi_1 = x \to y$, $\varphi_2 = \neg(x \to y)$, $\varphi_3 = (x \vee z) \to y$, $\varphi_4 = \neg((x \wedge y) \vee z) \to w$, $\varphi_5 = y \to (x \wedge z)$, $\varphi_6 = (\neg y \vee (x \to z))$, $\varphi_7 = z \to (x \vee y)$, $\varphi_8 = \neg((z \wedge y) \to x$.

We perform clustering analysis of the set *test* 1 using k-means algorithm. Suppose we need 3 clusters as a result. We choose the following weights: $\lambda_0 = 0, \lambda_1 = \frac{1}{4}, \lambda_2 = \frac{2}{4}, \lambda_3 = \frac{3}{4}, \lambda_4 = 1$. Based on the matrix of distances we choose 3 centres: $\varphi_2, \varphi_4, \varphi_5$ (the centres are approximately equidistant from each other and the sum of distances between them is maximal). The remaining formulas are distributed into the source clusters according to those centres. This gives us the following clusters: $\varphi_2, \varphi_{1,3,4,6,7}, \varphi_{5,8}$. Then the algorithm calculates the centres of mass again and redistributes the formulas according to the renewed centres. After this we have the following clusters: $\varphi_2, \varphi_{1,3,4,6,7}, \varphi_{5,8}$. As we see, the clusters didn't change – this means the algorithm stops. The results of the performed clustering are presented in Table 2.

Table 2. K-means algorithm, *test* 1, $n = 5$

Iteration	Centres	Clusters
1	$\varphi_2, \varphi_4, \varphi_5$	$\varphi_2, \varphi_{1,3,4,6,7}, \varphi_{5,8}$
2	$\varphi_2, \varphi_3, \varphi_5$	$\varphi_2, \varphi_{1,3,4,6,7}, \varphi_{5,8}$

The algorithm stops after second iteration giving 3 clusters as a result: $\varphi_2, \varphi_{1,3,4,6,7}, \varphi_{5,8}$.

The results obtained after the clustering of the set *test* 1 utilizing the algorithms based on new distances and uncertainty measures are different from the results obtained after the clustering of the same set using the distances and uncertainty measures with rigid weights in [9]. This demonstrates the feasibility of using different distances in clustering algorithms.

4.3 Hierarchic Algorithm, $n = 9$.

Let $n = 9$. Consider the set *test* 2, which consists of 10 formulas of nine-valued logic L_9: $\varphi_1 = \neg(z \vee y)$, $\varphi_2 = (x \to y) \to w$, $\varphi_3 = \neg((x \to y) \wedge z)$, $\varphi_4 = (x \vee z) \wedge y$, $\varphi_5 = z \to (x \vee y)$, $\varphi_6 = (\neg y \vee (x \to z))$, $\varphi_7 = w \wedge (x \to z)$, $\varphi_8 = (y \vee z) \to (x \vee w)$, $\varphi_9 = z \to (x \wedge w)$, $\varphi_{10} = \neg(x \to y)$.

We perform clustering analysis of the set *test* 2 using hierarchic algorithm. We choose the following weights: $\lambda_0 = 0, \lambda_1 = \frac{1}{30}, \lambda_2 = \frac{1}{20}, \lambda_3 = \frac{1}{10}, \lambda_4 = \frac{1}{5}, \lambda_5 = \frac{3}{10}, \lambda_6 = \frac{2}{5}, \lambda_7 = \frac{3}{5}, \lambda_8 = 1$. The stopping criterion is $\Delta = 0,2000$. The results of the performed clustering are presented in Table 3.

Table 3. Hierarchic algorithm, *test* 2, $n = 9$

Iteration	Δ	Clusters
0	0,0000	$\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5, \varphi_6, \varphi_7, \varphi_8, \varphi_9, \varphi_{10}$
1	0,0073	$\varphi_1, \varphi_{2,3}, \varphi_4, \varphi_5, \varphi_6, \varphi_7, \varphi_8, \varphi_9, \varphi_{10}$
2	0,0173	$\varphi_{2,3}, \varphi_4, \varphi_5, \varphi_{1,6,7}, \varphi_8, \varphi_9, \varphi_{10}$
3	0,0952	$\varphi_{2,3}, \varphi_4, \varphi_5, \varphi_{1,6,7}, \varphi_{8,9}, \varphi_{10}$
4	0,0952	$\varphi_{2,3}, \varphi_5, \varphi_{1,4,6,7}, \varphi_{8,9}, \varphi_{10}$
5	0,1907	$\varphi_{2,3}, \varphi_5, \varphi_{1,4,6,7}, \varphi_{8,9,10}$
6	0,1907	$\varphi_{2,3,5}, \varphi_{1,4,6,7}, \varphi_{8,9,10}$

The algorithm stops after sixth iteration giving 3 clusters as a result: $\varphi_{2,3,5}, \varphi_{1,4,6,7}, \varphi_{8,9,10}$.

4.4 K-means Algorithm, $n = 7$.

Let $n = 7$. Consider the set *test* 3, which consists of 15 formulas of seven-valued logic L_7: $\varphi_1 = y \to (x \wedge z)$, $\varphi_2 = \neg z \to w(x \wedge y)$, $\varphi_3 = z \to (x \vee y)$, $\varphi_4 = \neg((x \wedge y) \vee z) \to w$, $\varphi_5 = \neg(x \wedge z) \to y$, $\varphi_6 = (\neg y \vee (x \to z))$, $\varphi_7 = z \to (x \vee y)$, $\varphi_8 = \neg((z \wedge y) \to x$, $\varphi_9 = \neg z \to x$, $\varphi_{10} = \neg((x \wedge y) \vee z) \to w$, $\varphi_{11} = y \to (x \wedge w) \to z$, $\varphi_{12} = y \vee (x \to z)$, $\varphi_{13} = z \wedge (x \to y)$, $\varphi_{14} = (x \wedge z) \to w$, $\varphi_{15} = (x \vee w) \to y$.

We perform clustering analysis of the set *test* 3 using k-means algorithm. Suppose we need 4 clusters as a result. We choose the following weights: $\lambda_0 = 0, \lambda_1 = \frac{1}{6}, \lambda_2 = \frac{2}{6}, \lambda_3 = \frac{3}{6}, \lambda_4 = \frac{4}{6}, \lambda_5 = \frac{5}{6}, \lambda_6 = 1$. We also select the future centres of the clusters: $\varphi_2, \varphi_5, \varphi_7, \varphi_9$. During every iteration of the algorithm the formulas are distributed into the source clusters (according to the centres), then the centres of mass are calculated, and the resulting clusters are updated. The results of the performed clustering are presented in Table 4.

Table 4. K-means algorithm, *test* 3, $n = 7$

Iteration	Centres	Clusters
1	$\varphi_2, \varphi_5, \varphi_7, \varphi_9$	$\varphi_{1,2,3,8,10}, \varphi_{5,14}, \varphi_{6,7}, \varphi_{4,9,11,12}$
2	$\varphi_8, \varphi_5, \varphi_7, \varphi_9$	$\varphi_{2,3,8,10}, \varphi_{5,14}, \varphi_{6,7}, \varphi_{1,4,9,11,12}$
3	$\varphi_8, \varphi_5, \varphi_7, \varphi_4$	$\varphi_{2,3,8,10}, \varphi_{5,14}, \varphi_{6,7}, \varphi_{1,4,9,11,12}$

The algorithm stops after third iteration giving 4 clusters as a result: $\varphi_{2,3,8,10}, \varphi_{5,14}, \varphi_{6,7}, \varphi_{1,4,9,11,12}$.

5 Conclusions

In this paper, a continuum of new distances and uncertainty measures is offered for logical multivalued formulas. The new quantities are generalizations of the quantities that were used previously. Theorems 1 and 2 which respectively establish the properties of the new model distance and uncertainty measure are proven. Those new quantities can be used to analyse knowledge bases, to create expert systems, or to construct logical decision functions for the problems of recognition.

Based on new quantities the clustering analysis of multivalued logical formulas is performed. The formulas belong to the n-valued logic L_n. The software package for clustering analysis of sets of logical formulas, using the hierarchical and k-means algorithms is developed. The number of formulas, the dimension of logic L_n and the values of weights were chosen differently. The results are obtained for $n = 5, n = 7, n = 9$ (and more). A comparison of the clustering results with the results of previous works for the case $n = 5$ is also performed.

In the future we plan to use the new quantities for the analysis of the large sets of statements of experts. For this purpose the coefficient matrix composed of weights $\lambda_{|k-l|}$ will be used. This approach will explore the relationship between the selection of the optimal clustering and the properties of the coefficient matrix and multivalued logic.

Acknowledgments. This work is supported by the Russian Foundation for Basic Research, project nos. 10-0100113a and 11-07-00345a.

References

1. Vikent'ev, A.A.: Distances and degrees of uncertainty in many-valued propositions of experts and application of these concepts in problems of pattern recognition and clustering. Pattern Recogn. Image Anal. **24**(4), 489–501 (2014)
2. Vikent'ev, A.A.: Uncertainty measure of expert statements, distances in many-valued logic and adaptation processes. In: XVI International Conference "Knowledge-Dialogue-Solution" KDS-2008, Varna, pp. 179–188 (2008). (in Russian)
3. Vikent'ev, A.A., Lbov, G.S.: Setting the metrics and informativeness on statements of experts. Pattern Recogn. Image Anal. **7**(2), 175–183 (1997)
4. Vikent'ev, A.A., Vikent'ev, R.A.: Distances and uncertainty measures on the statements of N-valued logic. In: Bulletin of the Novosibirsk State University, Serious of Mathematics Mechanics, Computer Science, Novosibirsk, vol. 11, no. 2, pp. 51–64 (2011). (in Russian)
5. Lbov, G.S., Startseva, N.G.: Logical Solving Functions and the Problem on Solutions Statistical Stability. Sobolev Institute of Mathematics, Novosibirsk (1999). (in Russian)
6. Berikov, V.B.: Grouping of objects in a space of heterogeneous variables with the use of taxonomic decision trees. Pattern Recogn. Image Anal. **21**(4), 591–598 (2011)
7. Avilov, M.S.: The software package for calculating distances, uncertainty measures and clustering sets of formulas of N-valued logic. In: ISSC-2015, Mathematics, Novosibirsk, p. 6 (2015). (in Russian)
8. Lbov, G.S., Bloshitsin, V.Y.: On informativeness measures of logical statements. In: Lectures of the Republican School-Seminar "Development Technology of Expert Systems", Chiinu, pp. 12–14 (1978). (in Russian)
9. Vikent'ev, A.A., Kabanova, E.S.: Distances between formulas of 5-valued lukasiewicz logic and uncertainty measure of expert statements for clustering knowledge bases. In: Bulletin of the Tomsk State University, Tomsk, vol. 2, no. 23, pp. 121–129 (2013). (in Russian)
10. Karpenko, A.S.: Lukasiewicz Logics and Prime Numbers. Nauka, Moscow (2000). (in Russian)
11. Ershov, Y.L., Palutin, E.A.: Mathematical Logics. Fizmatlit, Moscow (2011). (in Russian)
12. Zagoruiko, N.G., Bushuev, M.V.: Distance measures in knowledge space. In: Data Analysis in Expert Systems, 117: Computing Systems, Novosibirsk, pp. 24–35 (1986). (in Russian)