ΑΞΙΟΛΟΓΗΣΗ ΤΗΣ ΑΝΙΟΥΣΑΣ ΙΕΡΑΡΧΙΚΗΣ ΤΑΞΙΝΟΜΗΣΗΣ ΜΕΘΟΔΟΣ BENKAR

Επιστήμη είναι η συστηματική κατάταξη της πείρας

Λιούϊς

6.0 Γενικά

Μία ανιούσα ιεραρχική ταξινόμηση των «υποκειμένων» ενός συνόλου I με πληθάριθμο card(I)=n, είναι μία διαδικασία που παράγει μια ακολουθία διαμελισμών του αρχικού συνόλου σε υποσύνολα μη κενά και ξένα ανά δύο μεταξύ τους, τις λεγόμενες **κλάσεις**, τη μία μέσα στην άλλη, συνενώνοντας κάθε φορά δύο μόνο κλάσεις οι οποίες βάσει κάποιας μετρικής παρουσιάζουν σε κάθε βήμα ομαδοποίησης την μικρότερη απόσταση.

Απ' ότι γίνεται αντιληπτό στόχος της ανιούσας ιεραρχικής ταξινόμησης είναι να ομαδοποιήσει το σύνολο των στατιστικών μονάδων ενός πληθυσμού σ' ένα περιορισμένο πλήθος ομοιογενών κλάσεων, λαμβάνοντας υπόψη το σύνολο των μεταβλητών, ώστε κάθε μία να διαφέρει από τις άλλες όσο το δυνατόν περισσότερο.

Οι κλάσεις δημιουργούνται βάσει ενός αντικειμενικού αλγορίθμου, πέρα από τις υποκειμενικές μεθόδους που μπορεί να αναπτύξει κάθε ερευνητής. Λέμε αντικειμενικό αλγόριθμο γιατί η ομαδοποίηση των στατιστικών μονάδων γίνεται χωρίς καμιά a priori υπόθεση στον αρχικό πίνακα δεδομένων και βάσει μιας συγκεκριμένης μετρικής.

Βέβαια ένας πίνακας λ.χ που περιέχει βαθμολογίες διαφόρων κριτηρίων μπορεί να περιλαμβάνει και ακραίες τιμές, οι οποίες υποχρεωτικά θα ληφθούν υπόψη στη διαδικασία ταξινόμησης των «αντικειμένων», όταν υπολογίζεται σε κάθε βήμα της ταξινόμησης η μείωση της εξωταξικής αδράνειας.

Αποτέλεσμα αυτού το γεγονότος είναι οι τιμές των «υποκειμένων» στα διάφορα κριτήρια, που περιλαμβάνονται σε μια συγκεκριμένη κλάση, να διαφέρουν μεταξύ τους, κάτι που αναιρεί σε κάποιο βαθμό την ομοιογένεια των απαντήσεων που περιλαμβάνονται στη κλάση.

Ακριβώς αυτή την ομοιογένεια των κλάσεων μιας Ανιούσας Ιεραρχικής Ταξινόμησης επιδιώκει να αξιολογήσει η προτεινόμενη μέθοδος.

6.1 Δημιουργία της Ανιούσας Ιεραρχικής Ταξινόμησης

Έστω ο πίνακας T(nxp) με n γραμμές και p στήλες. Σε κάθε ταξινόμηση ανεξάρτητα με ποια μετρική πραγματοποιείται, δημιουργούνται κλάσεις που κάθε μια περιέχει ένα συγκεκριμένο πλήθος «υποκειμένων», το οποίο παριστάνουν οι n γραμμές του πίνακα δεδομένων. Έτσι λ.χ σ' ένα ερωτηματολόγιο έρευνας αγοράς σε κάθε γραμμή του πίνακα αντιστοιχούν οι απαντήσεις του κάθε ερωτηθέντα στο σύνολο των p ερωτημάτων που του ετέθησαν.

Ως γνωστόν σε μία ταξινόμηση με την μέθοδο VACOR, χρησιμοποιείται ο αλγόριθμος του Ward,οπότε όταν περνάμε από ένα διαμελισμό με λ+1 κλάσεις σ' ένα άλλο διαμελισμό που έχει λ κλάσεις, συγχωνεύοντας δύο κλάσεις σε μία με κριτήριο την μείωση της διαταξικής αδράνειας, προβλέπει την εφαρμογή του θεωρήματος του Huggens. Επί πλέον θεωρείται ότι όλα τα στοιχεία συγκεντρώνονται στο **κέντρο βάρους** της κλάσης, το οποίο **σταθμίζεται** με το βάρος των στοιχείων της k κλάσης, το οποίο αποτελεί το βαρύκεντρο της κλάσης

Η συνένωση δύο παρατηρήσεων ή δύο κλάσεων σε μία κλάση δημιουργεί αυτό που ονομάζεται **κόμβος** της ιεραρχίας. Ο κάθε κόμβος της ιεραρχίας συμβολίζει το κέντρο βάρους των «υποκειμένων» που συμμετέχουν σ' αυτόν, η δε περιγραφή της ταξινόμησης γίνεται με το δενδρόγραμμα, του οποίου οι κόμβοι συμβολίζουν τον διαμελισμό του πληθυσμού σε n-1 κλάσεις, όπου n το πλήθος των «υποκειμένων» προς ταξινόμηση.

Αν δεν ενδιαφερόμαστε για την συνολική ιεραρχία των n «υποκειμένων», αλλά μόνο για ένα περιορισμένο αριθμό k κλάσεων, δεν έχουμε παρά να πάρουμε μία «**τομή**» του δενδρογράμματος στο επίπεδο ε_1, δηλαδή να **"κόψουμε"** το δενδρόγραμμα με μία ευθεία γραμμή, στο σημείο όπου οι κλάδοι που απομένουν να ικανοποιούν τον αριθμό των k κλάσεων που επιθυμούμε να διατηρήσουμε, οπότε δημιουργείται η «**τυπολογία**» της ταξινόμησης σε k κλάσεις.

6.2 Η μέθοδος BENKAR

Ως γνωστό μέχρι σήμερα μέθοδοι αξιολόγησης μιας ταξινόμησης μπορούν να πραγματοποιηθούν με διαδικασίες που προβλέπουν είτε την χρήση των νευρωνικών δικτύων, είτε χρησιμοποιώντας ταξινομητές μηχανικής μάθησης, οι οποίοι δεν αποδίδουν πιθανότητες αλλά μόνο εκτίμηση της επίδοσης μάθησης για το αποτέλεσμα που προκύπτει.

Με την προτεινόμενη μέθοδο, χρησιμοποιούνται εκτός από τις βασικές αρχές της Παραγοντικής Ανάλυσης των Αντιστοιχιών και των ιδιοτήτων του Ευκλείδειου διανυσματικού χώρου R^n, αποδίδεται η κατανομή πιθανοτήτων των αντικειμένων να ανήκουν σε συγκεκριμένες κλάσεις της ιεραρχίας.

Ειδικότερα οι k κόμβοι (δηλαδή τα k κέντρα των κλάσεων) μιας συγκεκριμένης τυπολογίας της ιεραρχίας, δημιουργούνται αφού πρώτα αθροίσουμε για κάθε στήλη τις τιμές των γραμμών του πίνακα T(n,p) που ανήκουν σε κάθε κλάση και στη συνέχεια τις k κλάσεις τις θεωρήσουμε ως νέες γραμμές, δημιουργώντας ένα επαυξημένο πίνακα T(n+k,p) τον οποίο αναλύουμε με την Παραγοντική Ανάλυση των Αντιστοιχιών, οπότε υπολογίζουμε τις συντεταγμένες τους πάνω στους p-1 παραγοντικούς άξονες.

Χρησιμοποιώντας τις συντεταγμένες του συνόλου των σημείων του νέφος N(I) των γραμμών, του νέφους N(J) των στηλών και των k κόμβων της ταξινόμησης πάνω στους p-1 παραγοντικούς άξονες, που προκύπτουν από την ανάλυση του πίνακα T(n+k,p), δημιουργείται μία ορθοκανονική βάση στον χώρο $R^{(p-1)}$, όπου τοποθετούνται οι p μεταβλητές, οι n στατιστικές μονάδες και τα κέντρα των k κλάσεων, στις πραγματικές τους θέσεις, απ' όπου αντλείται το σύνολο της πληροφόρησης που παρέχει ο πίνακας δεδομένων.

Ακολούθως με την χρήση της Ευκλείδειας μετρικής μπορούμε να υπολογίσουμε τις αποστάσεις της κάθε στατιστικής μονάδας από τα k κέντρα των κλάσεων, όπως προτείνει και ο αλγόριθμος του Ward. Τις k αποστάσεις της κάθε στατιστικής μονάδας τις μετατρέπουμε σε k πιθανότητες, όπου η πιο μικρή από τις k αποστάσεις, αντιστοιχεί στη μεγαλύτερη πιθανότητα που έχει η στατιστική μονάδα να είναι πλησίον του κέντρου της κλάσης με την μικρότερη απόσταση, όπου a priori στη μικρότερη αυτή απόσταση δεν αντιστοιχεί πάντοτε η κλάση που προσδιόρισε η διαδικασία της ταξινόμησης βάσει του αλγορίθμου του Ward.

Τούτου διότι, όταν σε μία κλάση περιλαμβάνονται στατιστικές μονάδες με ακραίες τιμές η **ομοιογένεια** της κλάσης αυτής, σχετικά με τις τιμές των άλλων στατιστικών μονάδων που περιλαμβάνει αλλοιώνεται, καθόσον το κέντρο της κάθε

κλάσης, όπως προαναφέρθηκε, σταθμίζεται με το βάρος των στοιχείων της k κλάσης σε κάθε βήμα της δημιουργίας ενός κόμβου της ιεραρχίας.

Γι' αυτό κρίνεται σκόπιμο να αξιολογηθεί η ορθή τοποθέτηση των στατιστικών μονάδων ως προς ένα καθορισμένο πλήθος κλάσεων της ταξινόμησης χρησιμοποιώντας τις μέγιστες πιθανότητες που προκύπτουν από την μετατροπή των ελάχιστων αποστάσεων κάθε στατιστικής μονάδος από τα κέντρα των κλάσεων, βάσει της Ευκλείδειας μετρικής.

Η μετατροπή, των p-1 αποστάσεων των n σημείων-γραμμών από τα κέντρα των k κλάσεων σε αντίστοιχες πιθανότητες, προκύπτει από την σχέση:

$$P(i, j) = tdis(i, j) / \sum_{j=1}^{k} tdis(i, j) \text{ για κάθε } i \in n \text{ και } j \in k$$

$$\text{με } tdis(i, j) = 1/[dis(i, j)]^2 \text{ για κάθε } i \in n \text{ και } j \in k$$

όπου $dis(i, j)$ η απόσταση κάθε $i \in n$ από κάθε κέντρο κλάσης $j \in k$

Ακολούθως σχηματίζεται η κατανομή των n μέγιστων πιθανοτήτων σε m ίσες τάξεις, απ' όπου προκύπτει η ζητούμενη αξιολόγηση της αρχικής ταξινόμησης. Μελετώντας την κατανομή των μέγιστων πιθανοτήτων, αν η αθροιστική συχνότητα (η οποία μετατρέπεται σε ποσοστό) των δύο τελευταίων τάξεων, που προσδιορίζει το πλήθος των στατιστικών μονάδων τα οποία κατατάχθηκαν με τις δύο μεθόδους στις ίδιες κλάσεις είναι σχετικά μικρό και αν ακόμη το ποσοστό που προσδιορίζει το εύρος των δύο τελευταίων τάξεων είναι αρκούντως ικανοποιητικό, η ταξινόμηση σε k ομοιογενείς κλάσεις θεωρείται ότι δεν είναι ικανοποιητική, επειδή η ασυμφωνία στην κατάταξη των αντικειμένων στις k κλάσεις υποδηλώνει ότι οι k κλάσεις περιέχουν ανομοιογενείς στατιστικές μονάδες ως προς τις τιμές των p κριτηρίων.

Αν βέβαια το ποσοστό των αντικειμένων των δύο τελευταίων τάξεων είναι ικανοποιητικό, σε συνάρτηση με το ποσοστό που προσδιορίζει το εύρος των δύο τελευταίων τάξεων της κατανομής πιθανοτήτων, καθορίζουν την αξιολόγηση της Ανιούσας Ιεραρχικής Ταξινόμησης με την μέθοδο VACOR.

6.3 Εφαρμογή της μεθόδου BENKAR

Για την εφαρμογή της προτεινόμενης μεθόδου θα χρησιμοποιηθεί ένα συγκεκριμένο ερωτηματολόγιο ποιοτικών μεταβλητών (για την μέτρηση των οποίων χρησιμοποιήθηκε η 5βάθμια κλίμα Likert, όπου το 5 αφορούσε την άριστη εντύπωση), στο οποίο απάντησαν 1721 άτομα. Ένα τμήμα του ερωτηματολογίου αφορούσε έξι ερωτήσεις σχετικά με το πώς βαθμολογούν οι ξένοι επισκέπτες α) τα αξιοθέατα της πόλης της Θεσσαλονίκης β) την Ελληνική κουζίνα γ) την νυχτερινή ζωή της πόλης δ) το αρχιτεκτονικό της στυλ ε) την ασφάλειά της και στ) την φιλικότητα των ντόπιων.

Στα δεδομένα που προέκυψαν εφαρμόστηκε η Ανιούσα Ιεραρχική Ταξινόμηση (-CAH-) με την διαδικασία VACOR. Για την αξιολόγηση της ταξινόμησης και με κριτήριο διαμελισμού το λ_r προκρίθηκε η τομή του δενδρογράμματος σε πέντε κλάσεις.

Τα στοιχεία αφορούν ξένους επισκέπτες της Θεσσαλονίκης και τα δεδομένα περιέχονται στην έρευνα που διεξήχθη στα πλαίσια του προγράμματος ΑΡΧΙΜΗΔΗΣ ΙΙΙ με τίτλο «Τεχνολογίες Ανάλυσης Δεδομένων και Διαχείρισης Γνώσης στο σχεδιασμό τουριστικών προϊόντων»

Οι έξι μεταβλητές παρίστανται αντιστοίχως ως εξής : Δ4,Δ5,Δ6,Δ7,Δ8,Δ9. Με δεδομένη την ταξινόμηση των 1721 ατόμων με την διαδικασία VACOR ο πίνακας 6.1 παρουσιάζει τις απαντήσεις τους και τις πέντε κλάσεις στις οποίες ανήκουν οι ερωτώμενοι.

Πίνακας 6.1: Τιμές των έξι μεταβλητών και οι πέντε κλάσεις στις οποίες ανήκουν οι ερωτώμενοι μετά την ταξινόμηση με την διαδικασία VACOR

IND	Δ4	Δ5	Δ6	Δ7	Δ8	Δ9	Class VACOR
I1	4	4	0	4	5	5	3
I2	5	4	5	5	4	5	4
.							.
1690	4	3	2	2	2	2	5
‘							.
1721	5	5	4	3	4	5	5

Ο πίνακας 6.2 παρουσιάζει τον επαυξημένο πίνακα που δημιουργείται στο βήμα 3.

Πίνακας 6.2 : Τμήμα του επαυξημένου πίνακα T(n+k,p)

IND	Δ4	Δ5	Δ6	Δ7	Δ8	Δ9
I1	4	4	0	4	5	5
I2	5	4	5	5	4	5
‘						
I1721	5	5	4	3	4	5
K1	398	212	224	387	378	421
K2	727	750	810	592	391	620
K3	548	545	35	482	430	521
K4	3679	3612	3192	3600	3403	3437
K5	1580	1883	1907	1548	1788	1963

Ακολουθώντας τα βήματα 4 και 5 δημιουργείται, ο πίνακας T3(n,2k) ο οποίος παρουσιάζει για κάθε σημείο-γραμμή τις αποστάσεις K_i (i=1,…5) των σημείων από τα κέντρα των πέντε κλάσεων και τις αντίστοιχες πιθανότητες P_i (i=1,…5)

Πίνακας 6.3:Τμήμα του πίνακα αποστάσεων K_i των γραμμών με τις αντίστοιχες πιθανότητες

IND	K1	P1	K2	P2	K3	P3	K4	P4	K5	P5
I1	12,769	0,077	41,496	0,007	3,816	0,865	20,147	0,031	25,590	0,019
I2	5,770	0,032	3,719	0,078	21,824	0,002	1,264	0,672	2,229	0,216
I3	26,919	0,024	5,980	0,485	29,984	0,019	10,966	0,144	7,280	0,327
.	.	.	.	.	.	.	.	.	.	.
I1690	11,800	0,051	4,899	0,298	12,839	0,043	3,795	0,497	8,065	0,110
.	.	.	.	.	.	.	.	.	.	.
I1719	25,445	0,019	4,003	0,755	35,039	0,010	12,399	0,079	9,378	0,138
I1720	9,281	0,003	5,912	0,006	21,472	0,001	1,164	0,162	0,514	0,829
I1721	9,253	0,015	4,365	0,069	15,999	0,005	2,087	0,302	1,470	0,609

Στη συνέχεια στο βήμα 6 προκύπτει ο πίνακας T4(n,3) ο οποίος παρουσιάζει για κάθε σημείο-γραμμή στη $1^η$ στήλη την ταξινόμησή του με την διαδικασία VACOR, στη $2^η$ στήλη την νέα ταξινόμησή του με την διαδικασία της ελάχιστης

απόστασης (DIS) και στη 3η στήλη την μέγιστη πιθανότητα να ανήκει στην κλάση που αντιστοιχεί στην ελάχιστη απόσταση.

Αυτό συμβαίνει διότι η ταξινόμηση με την ΕΥΚΛΕΊΔΕΙΟ μετρική (DIS) περιγράφει καλύτερα τις ομαδοποιήσεις των «υποκειμένων» σε 5 κλάσεις απ' ότι η μέθοδος VACOR, όπως διαπιστώθηκε στο κεφάλαιο V. Επομένως είναι προφανές να εντοπίζουμε με την μετρική DIS, σε ποια από τις 5 κλάσεις ανήκει το κάθε «υποκείμενο», καθώς και την μέγιστη πιθανότητα να ανήκει σε κάποια από αυτές, με τελική επιδίωξη να διαπιστωθεί πόσα «υποκείμενα» τακτοποιήθηκαν σε ίδιες κλάσεις των δύο διαφορετικών ταξινομήσεων.

Πίνακας 6.4: Ταξινόμηση των «υποκειμένων» με την διαδικασία VACOR και την Ευκλείδεια μετρική (DIS) στο χώρο R^5

IND	VACOR	DIS	maxProb
I1	3	3	0,8651
I2	4	4	0,6717
I3	2	2	0,4851
.	.	.	.
I1690	5	4	0,497
.	.	.	.
I1719	2	2	0,7552
I1720	5	5	0,829
I1721	5	5	0,6086

Από το λογισμικό MAD δίδονται τα ακόλουθα αποτελέσματα:

Τακτοποιήθηκαν στις ΙΔΙΕΣ κλάσεις :1416 άτομα

Τακτοποιήθηκαν σε ΔΙΑΦΟΡΕΤΙΚΕΣ κλάσεις : 305 άτομα

Ποσοστό καλής προσαρμογής : 82,28%

Συνεχίζοντας με το βήμα 7 έχουμε τις τρεις κατανομές σε πέντε κλάσεις που προήλθαν μετά από ταξινόμηση των 1721 ατόμων

α) με την Ανιούσα Ιεραρχική Ταξινόμηση (-CAH-) με τη μέθοδο VACOR

β) με την Ευκλείδειο μετρική (DIS) στο χώρο R^5 που δημιούργησαν οι πέντε παραγοντικοί άξονες μετά την εφαρμογή της Παραγοντικής Ανάλυσης των Αντιστοιχιών στον πίνακα δεδομένων T(1721,6)

γ) την κατανομή των μέγιστων πιθανοτήτων να ανήκουν τα 1416 κοινά ταξινομημένα άτομα των δύο μεθόδων από τα 1721 στις πέντε διαφορετικές κλάσεις με την μετρική DIS.

Πίνακας 6.5: Οι τρεις κατανομές του 7ου βήματος

VACOR	n_i	f_i	DIS	n_i	f_i	**Κατανομή πιθανοτήτων**	ni
K1	99	0,0575	**K1**	136	0,079	**T1: 0,2305 - 0,3855**	48
K2	181	0,1051	**K2**	225	0,1307	**T2: 0,3855 - 0,5408**	230
K3	129	0,0749	**K3**	142	0,0825	**T3: 0,5408 -0, 6962**	315
K4	848	0,4927	**K4**	764	0,4439	**T4: 0,6962 - 0,8515**	335
K5	464	0,2696	**K5**	454	0,2638	**T5: 0,8515 - 1,0000**	488
	1721	1		1721	1		1416

Παρατήρηση: Η ερμηνεία της κλάσης Τ5 είναι η εξής: 488 άτομα από τα 1416 που τακτοποιήθηκαν στις ίδιες κλάσεις, δηλαδή ποσοστό 34,5% έχει πιθανότητα από 0,8515 και άνω να ανήκει σε ΜΙΑ από τις πέντε κλάσεις της ταξινόμησης CAH με την μέθοδο VACOR.

Προχωρώντας στο βήμα 8 η αξιολόγηση της συγκεκριμένης ταξινόμησης των 1721 επισκεπτών με την μέθοδο BENKAR, χρησιμοποιώντας τις δύο τελευταίες τάξεις της κατανομής πιθανοτήτων είναι η εξής: 823 (=488+335) άτομα από τα 1721, δηλαδή το 47,82% του συνόλου των ερωτηθέντων έχει πιθανότητα από 0,6962 και άνω να ανήκει σε ΜΙΑ από τις πέντε κλάσεις της ταξινόμησης CAH με την μέθοδο VACOR.

Επειδή το ποσοστό 47,82% του συνόλου των ερωτηθέντων δεν είναι ικανοποιητικό, παρά το ότι το ποσοστό 82,28% της προσαρμογής των «αντικειμένων» στις πέντε κλάσεις των δύο ταξινομήσεων είναι αρκετά υψηλό, τα δεδομένα του πίνακα Τ(1721,6) εντός των κλάσεων φαίνεται ότι δεν έχουν την απαραίτητη επιθυμητή ομοιογένεια, σε σχέση με τις τιμές των έξι μεταβλητών.

Εφαρμόζοντας στη συνέχεια την μέθοδο BENKAR για το πώς αξιολογούν οι 1721 επισκέπτες της Θεσσαλονίκης τα τρία πρώτα ερωτήματα της έρευνας Δ1=την καθαριότητα της πόλης, Δ2=τις φυσικές ομορφιές και Δ3=τις τιμές των προϊόντων και υπηρεσιών, προέκυψαν τα εξής αποτελέσματα:

Πίνακας 6.6: Τιμές των τριών μεταβλητών και οι κλάσεις στις οποίες ανήκουν οι ερωτώμενοι μετά την ταξινόμηση με την διαδικασία VACOR

IND	Δ1	Δ2	Δ3	Class VACOR
I1	3	5	4	5
I2	2	5	3	5
I3	1	3	4	5
.				
1690	2	3	5	5
.				
1719	1	2	3	5
1720	1	5	2	3
1721	2	5	3	5

Εφαρμόζοντας στα δεδομένα του πίνακα 6.6 την τακτοποίηση των «υποκειμένων» σε 5 κλάσεις και με την Ευκλείδειο μετρική έχουμε

Τακτοποιήθηκαν στις ΙΔΙΕΣ κλάσεις :1638 άτομα

Τακτοποιήθηκαν σε ΔΙΑΦΟΡΕΤΙΚΕΣ κλάσεις : 83 άτομα

Ποσοστό καλής προσαρμογής : 95,18%

Πίνακας 6.7: Οι τρεις κατανομές του 7^{ου} βήματος

VACOR	n_i	f_i	DIS	n_i	f_i	Κατανομή πιθανοτήτων	ni
K1	119	0,0691	**K1**	135	0,0784	**T1: 0,3700 -0, 4970**	68
K2	294	0,1708	**K2**	348	0,2022	**T2: 0,4970 - 0,6241**	201
K3	221	0,1284	**K3**	219	0,1272	**T3: 0,6241 - 0,7512**	161
K4	691	0,4015	**K4**	681	0,3957	**T4: 0,7512 - 0,8784**	254
K5	396	0,23	**K5**	338	0,1963	**T5: 0,8784 - 1,0000**	954
	1721	1		1721	1		1638

Από τον πίνακα 6.7 προκύπτει η εξής αξιολόγηση: 1208 άτομα (954+254) από τα 1721, δηλαδή το 70,18% του συνόλου των ερωτηθέντων έχει πιθανότητα από 0,7512 και άνω να ανήκει σε ΜΙΑ από τις πέντε κλάσεις της ταξινόμησης.

Επειδή το ποσοστό 70,18% του συνόλου των ερωτηθέντων είναι ικανοποιητικό, ενώ και το ποσοστό 95,18% της προσαρμογής των «αντικειμένων» στις πέντε κλάσεις των δύο ταξινομήσεων είναι υψηλό, τα δεδομένα του πίνακα Τ(1721,3) εντός των κλάσεων αυτών μπορεί να θεωρηθούν ότι έχουν την απαραίτητη επιθυμητή ομοιογένεια, σε σχέση με τις τιμές των τριών μεταβλητών.

Συγκρίνοντας τις δύο αξιολογήσεις των ταξινομήσεων αφενός με τα κριτήρια Δ1-Δ3, αφετέρου με τα κριτήρια Δ4-Δ9, αναδεικνύεται ότι οι 1721 επισκέπτες της Θεσσαλονίκης είχαν ομοιόμορφη εικόνα της πόλης ως προς τα τρία πρώτα κριτήρια που κλήθηκαν να αξιολογήσουν, ενώ για τα υπόλοιπα έξι κριτήρια δεν παρουσίαζαν ανάλογη ομοιογένεια απαντήσεων.

Αυτό μπορεί να οφείλεται επειδή οι επισκέπτες προερχόταν από 51 διαφορετικές χώρες του πλανήτη, οπότε κρίνεται φυσιολογικό στα κριτήρια Δ1 έως Δ3 να έχουν περίπου την ίδια αντίληψη, ενώ σε κάποια από τα κριτήρια Δ4 έως Δ9 να παρουσιάζουν τελείως διαφορετικές απόψεις λόγω διαφορετικών παραδόσεων που ισχύουν στον τόπο του κάθε επισκέπτη.

6.4 Εκπαίδευση των δεδομένων με την Μηχανή Διανυσμάτων Υποστήριξης – ΜΔΥ-

Εκπαιδεύοντας με 20 επαναλήψεις τα δεδομένα του πίνακα 6.1, με την χρήση της Μηχανής Διανυσμάτων Υποστήριξης), διατηρώντας σε κάθε επανάληψη ένα τυχαίο δείγμα 20% των 1721 τιμών, μια φορά με τις κατατάξεις των «αντικειμένων» με την μέθοδο VACOR και την άλλη με την μέθοδο BENKAR έχουμε τα ακόλουθα αποτελέσματα:

Πίνακας 6.8: Ποσοστά εκμάθησης των 20 επαναλήψεων μετά την εκπαίδευση των ταξινομήσεων των δεδομένων βάσει της μεθόδου VACOR και της μεθόδου BENKAR

Με VACOR 20%		Με BENKAR 20%	
0,6950		0,8363	
0,6862		0,9620	
0,7595		0,7105	
0,7830		0,9415	
0,7009		0,9444	
0,7185		0,9094	
0,7859		0,7515	
0,8211		0,9532	
0,7889	76,54%	0,6462	87,88%
0,8710		0,9035	
0,8328		0,7076	
0,8182		0,9064	
0,6862		0,9474	
0,7830		0,9474	
0,7067		0,8918	
0,8065		0,9737	
0,8035		0,9386	
0,7155		0,7982	
0,7595		0,9561	
0,7859		0,9503	

Από τον πίνακα 6.8 διαπιστώνεται ότι η μέθοδος BENKAR υπερέχει στην σωστή ταξινόμηση των αντικειμένων. Αυτό προκύπτει επειδή το ποσοστό εκτίμησης της εκμάθησης των δεδομένων του πίνακα 6.1, που αφορά την ταξινόμηση με την μέθοδο BENKAR (87,88%) είναι υψηλότερο από εκείνο που προκύπτει με την Ανιούσα Ιεραρχική Ταξινόμηση με την μέθοδο VACOR

(76,54%).Επί πλέον με την μέθοδο BENKAR στις 20 επαναλήψεις εκμάθησης των δεδομένων τα ποσοστά εκτίμησης πάνω από την μέση τιμή είναι κατά πολύ υψηλότερα (13 στις 20 επαναλήψεις πάνω από 90% με μέγιστη τιμή 97,37%) από τα αντίστοιχα ποσοστά εκτίμησης με την μέθοδο VACOR που η μέγιστη τιμή ανέρχεται μόλις στο 87,10%.

Με δεδομένο ότι όπως προαναφέρθηκε η ΜΔΥ δεν επιστρέφει πιθανότητες, ενώ η μέθοδος BENKAR υπολογίζει την πιθανότητα κάθε στατιστικής μονάδας να ανήκει σε μία ορισμένη κλάση, ως εκ τούτου η κατανομή των μέγιστων πιθανοτήτων που προκύπτει από την προτεινόμενη μέθοδο, μπορεί να θεωρηθεί ότι αξιολογεί αντικειμενικά την Ανιούσα Ιεραρχική Ταξινόμηση που απορρέει με την μέθοδο VACOR.

6.5 Συμπέρασμα

Η μέθοδος BENKAR αξιοποιώντας αντικειμενικά κριτήρια, όπως είναι οι συντεταγμένες των σημείων πάνω στους παραγοντικούς άξονες μετά την εφαρμογή στον πίνακα δεδομένων της Παραγοντικής Ανάλυσης των Αντιστοιχιών, αλλά και της τοποθέτησής τους βάσει αυτών των συντεταγμένων στον Ευκλείδειο διανυσματικό χώρο R^p, με την χρήση της Ευκλείδειας μετρικής, παρέχει την δυνατότητα, αντικειμενικής αξιολόγησης της ομοιογένειας των «υποκειμένων» που συμμετέχουν στη διαμόρφωση των κλάσεων της Ανιούσας Ιεραρχικής Ταξινόμησης.

Η δυνατότητα αυτή της μεθόδου BENKAR μπορεί να εφαρμοστεί σε οποιοδήποτε πίνακα T(n,p) του οποίου τα δεδομένα έχουν ταξινομηθεί με οποιοδήποτε κριτήριο συνένωσης, δεδομένου ότι σε κάθε περίπτωση η μέθοδος BENKAR ταξινομεί τα «υποκείμενα», αφού προηγουμένως τα τοποθετήσει σ' ένα ορθοκανονικό σύστημα συντεταγμένων R^p που δημιουργούν οι παραγοντικοί άξονες μετά την εφαρμογή της Παραγοντικής Ανάλυσης των Αντιστοιχιών στον πίνακα δεδομένων.

Η υπεροχή της μεθόδου BENKAR στην κατάταξη των «υποκειμένων» σε k κλάσεις που υπέδειξε η Ανιούσα Ιεραρχική Ταξινόμηση, διαπιστώνεται και με την χρήση της Μηχανής Διανυσμάτων Υποστήριξης, η οποία αξιολογεί την εκμάθηση των δεδομένων σε ποσοστό υψηλότερο από εκείνο που παρέχει η ίδια μηχανή για την ταξινόμηση των ίδιων δεδομένων με την Ανιούσα Ιεραρχική Ταξινόμηση, χρησιμοποιώντας την μέθοδο BENKAR.

Επιπροσθέτως η μέθοδος BENKAR σε αντίθεση με την ΜΔΥ, αλλά και με κάθε άλλη μέθοδο ταξινόμησης, υπολογίζει την πιθανότητα των «υποκειμένων» να ανήκουν στις διαμορφούμενες κλάσεις, στοιχείο που αποτελεί βασικό πλεονέκτημα της συγκεκριμένης μεθόδου, με τελική κατάληξη την αντικειμενική αξιολόγηση της ομοιογένειας των κλάσεων, που είναι ένα από τα ζητούμενα σε κάθε ταξινόμηση.