

Μοντελοποίηση της Συμπεριφοράς των Φοιτητών της Πληροφορικής με Χρήση Bayesian Networks

Μιχάλης ΞΕΝΟΣ
Σχολή Θετικών Επιστημών και Τεχνολογίας
Ελληνικό Ανοικτό Πανεπιστήμιο
Πάτρα, Ελλάδα
xenos@eap.gr

ΠΕΡΙΛΗΨΗ

Το άρθρο παρουσιάζει μία επιστημονική προσέγγιση στη μοντελοποίηση της συμπεριφοράς των φοιτητών που σπουδάζουν με τη μέθοδο της Ανοικτής και εξ Αποστάσεως Εκπαίδευσης, με τη χρήση Bayesian Networks. Περιγράφεται το μοντέλο και η εφαρμογή του στο τμήμα του Συντονιστή μίας Θεματικής Ενότητας της Πληροφορικής στο Ελληνικό Ανοικτό Πανεπιστήμιο, καθώς και τα πλεονεκτήματα που δίνει η εφαρμογή του σε συνθήκες αβεβαιότητας. Στα πλαίσια του άρθρου παρουσιάζεται η υλοποίηση του μοντέλου και η εφαρμογή του που έδωσε ικανοποιητικά αποτελέσματα τόσο σε συνθήκες πρόβλεψης (εκτίμησης με βάση την προηγούμενη εμπειρία), όσο και σε περιπτώσεις αποτίμησης (τεκμηρίωσης των λόγων που οδήγησαν σε μία οριστική κατάσταση).

ABSTRACT

This paper presents a scientific and methodological approach towards modeling the behavior of students using the Open and Distance Learning method, based on Bayesian Networks. It describes the structure of the model, its application for modeling the behavior of the student group of the Coordinator of the Informatics Course in the Hellenic Open University, as well as the advantages of the presented method under conditions of uncertainty. The application of this model offered promising result as regards both prediction of student behavior (based on modeled past experience), as well as assessment (i.e. identification of the reasons that led to a given 'current' state).

1. ΕΙΣΑΓΩΓΗ

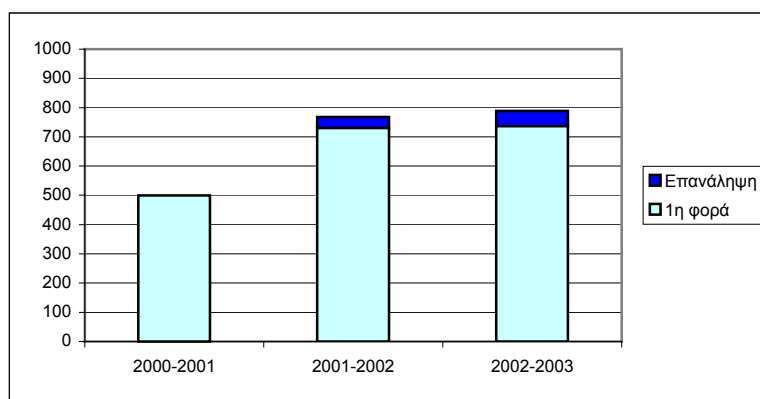
Ένα σύστημα Ανοικτής Εκπαίδευσης πρέπει να παρέχει στους φοιτητές τη δυνατότητα να αποκτήσουν εκπαίδευση υψηλού πανεπιστημιακού επιπέδου λαμβάνοντας υπόψη τις ιδιαιτερότητες της ανοικτής εκπαίδευσης (Race, 1999). Στην ανοικτή εκπαίδευση αρκετοί φοιτητές είναι πιθανό να διακόπτουν τις σπουδές τους και μάλιστα πολλές φορές αυτό θεωρείται μία υγιής επιλογή (Munro, 1987). Παρόλα αυτά, ο φορέας που παρέχει την εκπαίδευση πρέπει να εξετάζει τους λόγους επιτυχίας και αποτυχίας των φοιτητών και στην περίπτωση που αυτοί οι λόγοι είναι ανεξάρτητοι με τις ακαδημαϊκές επιδόσεις και δυνατότητες των φοιτητών, αλλά σχετικές με το σύστημα εκπαίδευσης, να προβαίνει σε βελτιώσεις.

Σε ένα σύστημα ανοικτής εκπαίδευσης που γίνεται με τη μέθοδο της Εκπαίδευσης από Απόσταση ο βασικός φορέας της 'επαφής' με τους φοιτητές και κατά συνέπεια παραλήπτης

αυτού του προβληματισμού είναι ο Σύμβουλος – Καθηγητής (ΣΚ). Ο καλός ΣΚ κρατάει ένα οργανωμένο και τακτικά ενημερωμένο αρχείο με σκοπό την καταγραφή της εμπειρίας του και στόχο την άτυπη εφαρμογή ενός νοητικού μοντέλου που στην ιδανική περίπτωση θα πρέπει να μπορεί να απαντά στα παρακάτω ερωτήματα: α) με βάση τα δεδομένα κάθε φοιτητή ποιες είναι οι πιθανότητες επιτυχίας του, έτσι ώστε να γίνει καλύτερη κατανομή της προσπάθειας ενθάρρυνσης / εμπύχωσης / ακαδημαϊκής βοήθειας; β) με δεδομένη την επιτυχία ή αποτυχία ενός φοιτητή ποιοι παράγοντες συνετέλεσαν σε αυτό το γεγονός; Ένα τέτοιο μοντέλο μπορεί να βρίσκεται στη σκέψη κάθε ΣΚ, αλλά όσο τα δεδομένα γίνονται πολύπλοκα και όσο ο αριθμός των παραμέτρων αυξάνει είναι αδύνατο να τηρείται νοερά. Πόσο μάλλον όταν υπάρχει ανάγκη μοντελοποίησης της συμπεριφοράς των φοιτητών μιας ολόκληρης σχολής και η εκτίμηση εκατοντάδων παραγόντων, όπως στο παράδειγμα που παρουσιάζεται.

Το άρθρο αυτό παρουσιάζει μία επιστημονική προσέγγιση στο θέμα της μοντελοποίησης της συμπεριφοράς των φοιτητών και ειδικότερα στη Θεματική Ενότητα (ΘΕ) ΠΛΗ10 με τίτλο «Εισαγωγή στην Πληροφορική» του προπτυχιακού Προγράμματος Σπουδών (ΠΣ) «Πληροφορική» της Σχολής Θετικών Επιστημών και Τεχνολογίας του Ελληνικού Ανοικτού Πανεπιστημίου (ΕΑΠ). Παρουσιάζεται ένα αυτοματοποιημένο μοντέλο, το εργαλείο ανάπτυξής του και η πρακτική του εφαρμογή. Το μοντέλο είναι επεκτάσιμο και μπορεί να χρησιμοποιηθεί σε πολλά επίπεδα, αλλά η εφαρμογή του έχει γίνει αρχικά πιλοτικά σε ένα μόνο τμήμα, αυτό του Συντονιστή της ΘΕ ΠΛΗ10 για το ακαδημαϊκό έτος 2001-2002.

Στην επόμενη ενότητα του άρθρου παρουσιάζεται το πεδίο εφαρμογής του μοντέλου, δηλαδή οι ιδιαιτερότητες του ΠΣ Πληροφορικής, ενώ στην 3^η ενότητα περιγράφονται τα Bayesian Networks στα οποία βασίζεται το μοντέλο. Στην 4^η ενότητα του άρθρου παρουσιάζεται η φιλοσοφία του μοντέλου και η ανάπτυξή του, καθώς και παραδείγματα και αποτελέσματα από την εφαρμογή του, ενώ στην 5^η ενότητα συμπεράσματα και ιδέες για μελλοντική εξέλιξη της εργασίας που παρουσιάζεται.



Σχήμα 1. Φοιτητές που εγγράφηκαν στην ΠΛΗ10 ανά ακαδημαϊκό έτος

2. ΤΟ ΠΕΔΙΟ ΕΦΑΡΜΟΓΗΣ ΤΟΥ ΜΟΝΤΕΛΟΥ

Οποιοδήποτε μοντέλο αποκτά σημαντικότητα σε σχέση με το πεδίο εφαρμογής του, που στην προκειμένη περίπτωση είναι το ΠΣ «Πληροφορική» του ΕΑΠ. Οι ιδιαιτερότητες και ο χαρακτήρας των σπουδών πληροφορικής παρέχουν πρόσφορο έδαφος για την εφαρμογή ενός τέτοιου μοντέλου. Η Πληροφορική συγκεντρώνει ένα μεγάλο αριθμό φοιτητών από τους οποίους η συντριπτική τους πλειοψηφία επιλέγει τη ΘΕ ΠΛΗ10. Αυτό γίνεται γιατί το ΕΑΠ κατευθύνει τους φοιτητές να συμπεριλάβουν την ΠΛΗ10 στις επιλογές τους. Τον πρώτο χρόνο λειτουργίας του ΠΣ «Πληροφορική» εισήχθησαν 510 φοιτητές, ενώ το δεύτερο 720 και το τρίτο έχουν εγγραφεί 738. Από το δεύτερο ακαδημαϊκό έτος λειτουργίας του ΠΣ όλοι οι φοιτητές επέλεξαν την ΠΛΗ10. Σε αυτούς προστίθεται και ένας αριθμός φοιτητών που επαναλαμβάνουν την ΠΛΗ10, λόγω μη επιτυχίας στις εργασίες ή στις εξετάσεις της προηγούμενης χρονιάς. Κάνοντας την υπόθεση ότι ο αριθμός των εισακτέων θα διατηρηθεί σταθερός η ΠΛΗ10 θα δέχεται περίπου 750-800 φοιτητές κάθε χρόνο, που θα κατανέμονται σε 26-28 τμήματα. Στο σχήμα 1 παρουσιάζονται οι αριθμοί των φοιτητών που ξεκίνησαν για πρώτη φορά ή επανέλαβαν την ΠΛΗ10 (για το ακαδημαϊκό έτος 2002-2003 ο αριθμός είναι εκτίμηση, με μία μικρή ανοχή ± 10 φοιτητών, αφού ο αριθμός των φοιτητών που θα επανεγγραφούν δεν ήταν δεδομένος την περίοδο της συγγραφής του άρθρου και βασίστηκε στα δεδομένα των 2 προηγούμενων ετών).

Με δεδομένο αυτό το μεγάλο αριθμό φοιτητών γίνεται ήδη φανερό ότι δεν είναι εφικτό να γίνει μία ανάλυση περίπτωσης 'από μνήμης' και ότι υπάρχει μία ανάγκη για αυτοματοποίηση κάποιων μετρήσεων και εκτιμήσεων. Η Πληροφορική στο ΕΑΠ αντιμετωπίζει ένα φαινόμενο με ένα σχετικά υψηλό ποσοστό διακοπής σπουδών (dropout rate). Το ποσοστό αυτό είναι της τάξης του 28% έως 35% (Xenos et. al, 2001) και στην πλειοψηφία του αφορά φοιτητές που διέκοψαν τις σπουδές τους λόγω αποτυχίας στη ΘΕ ΠΛΗ10. Το ποσοστό αυτό δεν είναι καθόλου υψηλό, ειδικά για ένα Πρόγραμμα Σπουδών που φιλοδοξεί να καθιερώσει ένα υψηλό επίπεδο αποφοίτων και να μην απαξιώσει το πτυχίο τους, παρόλα αυτά είναι σχετικά υψηλό σε σχέση με τα υπόλοιπα Προγράμματα Σπουδών του ΕΑΠ (Γίνου, 2001). Μάλιστα τα διεθνή δεδομένα μιλούν για ποσοστά διακοπής φοίτησης που κυμαίνονται από 20% έως 30% (Rumble, 1992), ενώ σε Ασιατικές χώρες τα αντίστοιχα ποσοστά φτάνουν έως και το 50% (Shin & Kim, 1999), (Narasimharao, 2000). Ένας αριθμός φοιτητών που διακόπτουν τις σπουδές τους το κάνει για καθαρά ακαδημαϊκούς λόγους. Τέτοιοι λόγοι είναι αδυναμία να αντεπεξέλθουν στις απαιτήσεις του προγράμματος, είτε λόγω κακής εκτίμησης του δικού τους χρόνου, είτε κακής αρχικής εκτίμησης της δυσκολίας του προγράμματος, είτε αλλαγής των συνθηκών με τις οποίες ξεκίνησαν τις σπουδές τους (Xenos et. al, 2002). Παρόλα αυτά υπάρχει ένα ποσοστό φοιτητών οι οποίοι *θα μπορούσαν να συνέχιζαν τις σπουδές τους αν οι περιπτώσεις τους είχαν εντοπιστεί από πριν*. Επίσης, για αυτούς που τελικά διέκοψαν τις σπουδές τους θα ήταν καλό να υπάρχει ένα μοντέλο που να δίνει με εύκολο τρόπο μία καλή *εκτίμηση για τους λόγους που διέκοψαν* με σκοπό να γίνει προσπάθεια αυτοί οι λόγοι να εκλείψουν σταδιακά (εάν τουλάχιστον αφορούν το ΕΑΠ). *Αυτό ακριβώς έρχεται να παρέχει το προτεινόμενο μοντέλο και παρουσιάζει η εφαρμογή του στην ΠΛΗ10.*

Οι σπουδές στην Πληροφορική και ειδικότερα στην ΠΛΗ10 έχουν μερικές ιδιαιτερότητες που κάνουν την εφαρμογή ενός τέτοιου μοντέλου πιο κατάλληλη. Έχουν ένα τυπικό συντονισμό που βασίζεται σε τυποποιημένους κανόνες επικοινωνίας με τους φοιτητές

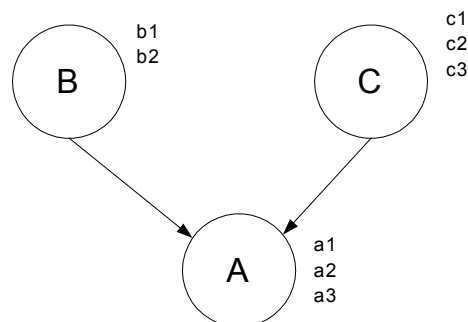
και ενημέρωσής τους, καθώς και τυποποιημένο τρόπο βαθμολόγησης εργασιών και εξετάσεων, παροχής απαντήσεων σε ερωτήσεις, εκδόσεων παράλληλου υλικού, βοηθημάτων, κτλ, κάτι που δεν είναι ούτε εύκολο, ούτε σύνηθες για όλα τα ΠΣ του ΕΑΠ. Η επικοινωνία φοιτητών – καθηγητών είναι κυρίως ηλεκτρονική με κυρίαρχο ρόλο να παίζει ο δικτυακός τόπος (web site) της ΠΛΗ10 (Pli10, 2002). Επίσης ηλεκτρονική είναι και η επικοινωνία των μελών της Ομάδας Διδακτικού Προσωπικού (ΟΔΠ) και γίνεται κυρίως με τη χρήση ηλεκτρονικού ταχυδρομείου και ηλεκτρονικού forum.

Το σύστημα διαχείρισης της ακαδημαϊκής διαδικασίας στην ΠΛΗ10 έχει 6 σημαντικά κομβικά σημεία (milestones) σχετικά με την ακαδημαϊκή πορεία κάθε φοιτητή και αυτά είναι οι επιδόσεις του σε κάθε μία από τις 4 γραπτές εργασίες και οι 2 εξεταστικές περιόδους. Φυσικά κάθε μεμονωμένος φοιτητής μπορεί να έχει σημαντικά κομβικά σημεία στη φοιτητική του πορεία, όπως είναι σημαντικές εργασιακές εξελίξεις ή αλλαγές, αλλαγές στην οικογενειακή του κατάσταση, στην υγεία του ή μελών της οικογένειάς του, κτλ. Όλα τα παραπάνω επηρεάζουν την πορεία του ως φοιτητή, αλλά και την εκτίμησή του για το επίπεδο της δυσκολίας του ΠΣ και τις δυνατότητες επιτυχίας του (Pierrakeas et. al, 2001). Όσον αφορά την ΠΛΗ10, με την ολοκλήρωση του ακαδημαϊκού έτους, ένας φοιτητής μπορεί να βρεθεί σε 3 τυπικές καταστάσεις: α) δεν ολοκλήρωσε με επιτυχία τις εργασίες, κατά συνέπεια πρέπει να επαναλάβει πλήρως την ΠΛΗ10, β) ολοκλήρωσε με επιτυχία τις εργασίες αλλά δεν επέτυχε στις εξετάσεις και θα επαναλάβει για μία χρονιά τις εξετάσεις χωρίς να έχει υποχρέωση εργασιών και γ) ολοκλήρωσε με επιτυχία τις εργασίες και τις εξετάσεις και κατά συνέπεια ολοκλήρωσε επιτυχώς την ΠΛΗ10. Το ποιοι φοιτητές συνεχίζουν και ποιοι όχι καθορίζεται σε μεγάλο βαθμό από την παραπάνω κατηγορία που ανήκουν, σε συνάρτηση με τις άλλες ΘΕ που παρακολούθησαν (εάν παρακολούθησαν). Το ποσοστό δε των διακοπών στην ΠΛΗ10 είναι ένα πολύ μεγάλο μέρος του ποσοστού των συνολικών διακοπών στην Πληροφορική. Τα δεδομένα αυτά λοιπόν γίνεται προσπάθεια να μοντελοποιηθούν ειδικά για την ΠΛΗ10 σε πρώτη φάση και σε αυτά εστιάζει το μοντέλο που εφαρμόστηκε.

3. BAYESIAN NETWORKS

Τα *Bayesian Networks* (BN) –που ονομάζονται επίσης *Belief Networks*, *Causal Probabilistic Networks*, και *Probabilistic Influence Diagrams*– ανήκουν στην κατηγορία των γραφικών μοντέλων, δηλαδή είναι γράφοι στους οποίους οι κόμβοι αναπαριστούν τυχαίες τιμές και οι ακμές τις συσχετίσεις μεταξύ ανεξάρτητων υποθέσεων (Lauritzen & Spiegelhalter, 1990). Ειδικότερα Bayesian Network είναι μια ειδική κατηγορία γραφικού μοντέλου, ένας κατευθυνόμενος άκυκλος γράφος, στον οποίο οι κόμβοι αντιπροσωπεύουν μεταβλητές και οι ακμές τις σχέσεις μεταξύ τους (Jensen, 1996). Είναι δηλαδή ένα δίκτυο που περιγράφει τις σχέσεις πιθανοτήτων μεταξύ των μεταβλητών. Η χρήση ενός BN δίνει τη δυνατότητα να οριστεί ‘λογική’ σε συνθήκες αβεβαιότητας και να συνδυαστούν τα πλεονεκτήματα της διαίσθησης, της εμπειρίας και ενός μαθηματικού μοντέλου. Με τη χρήση των BN είναι δυνατό να οριστεί ‘έξυπνα’ η σχέση μεταξύ των διαφόρων κόμβων (μεταβλητών) και να υπολογιστεί με συνέπεια το πώς επηρεάζουν οι αρχικές πιθανότητες τα αβέβαια συμπεράσματα, όπως τη μελλοντική συμπεριφορά των φοιτητών (*χρήση των BN για εκτίμηση*). Σημαντικό επίσης είναι ότι τα BN μπορούν να χρησιμοποιηθούν με δεδομένη μία τελική και κάποιες ενδιάμεσες καταστάσεις (π.χ. με δεδομένη την αποτυχία ενός φοιτητή και

την οικογενειακή τους κατάσταση) για να πιθανολογήσουν την κατάσταση των αρχικών κόμβων (χρήση των BN για αποτίμηση μίας δεδομένης κατάστασης).



Σχήμα 2. Παράδειγμα Bayesian Network: Παιδί A με γονείς B και C

Για να μπορέσουν να ορισθούν οι σχέσεις μεταξύ των μεταβλητών, σε κάθε κόμβο πρέπει να καθοριστούν οι εξαρτώμενες πιθανότητες οι οποίες περιγράφουν τις σχέσεις μεταξύ ενός κόμβου 'παιδί' και των 'γονέων' του. Αν οι πιθανές τιμές κάθε μεταβλητής είναι διακριτές, τότε σε κάθε κόμβο οι πιθανότητες μπορούν να περιγραφούν σε ένα πίνακα NPT (Node Probability Table). Ο πίνακας αυτός παρουσιάζει τις πιθανότητες να πάρει ο κόμβος 'παιδί' μια από τις καθορισμένες τιμές, για κάθε συνδυασμό τιμών που πιθανόν να έχουν οι κόμβοι 'γονείς' του (Kschischang et. al, 2001). Για παράδειγμα στο σχήμα 2 παρουσιάζονται 2 κόμβοι γονείς B και C και ο κόμβος παιδί A . Ο πίνακας πιθανοτήτων των κόμβου A εκφράζει την πιθανότητα $P(A|B,C)$ για όλους τους πιθανούς συνδυασμούς των A, B, C . Έτσι, στο σχήμα 2, ο κόμβος B έχει 2 πιθανές καταστάσεις (b_1, b_2), ο κόμβος C έχει 3 καταστάσεις (c_1, c_2, c_3) και ο A κόμβος 3 καταστάσεις (a_1, a_2, a_3) τότε ο πίνακας NPT του A θα έχει $3 \cdot 2 \cdot 3 = 18$ στοιχεία.

Το μαθηματικό μοντέλο στο οποίο στηρίζονται τα Bayesian Networks είναι το θεώρημα του Bayes. Το θεώρημα αυτό το ανέπτυξε ο μαθηματικός και θεολόγος Thomas Bayes τον 18ο αιώνα και εκδόθηκε πρώτη φορά το 1763. Η βάση των Bayesian Networks και του θεωρήματος είναι ότι θεωρούν τις πιθανότητες υπό συνθήκη (conditional probabilities) πιο βασικές από τις συνδυαστικές πιθανότητες (joint probabilities). Είναι πιο εύκολο να οριστεί η πιθανότητα $P(A|B)$ χωρίς αναφορά στη συνδυαστική πιθανότητα $P(A,B)$. Από τους ορισμούς (1) και (2), με αντικατάσταση στο δεύτερο μέλος της εξίσωσης προκύπτει το θεώρημα του Bayes (3).

Στις εξισώσεις (1), (2), (3), A είναι η πιθανότητα που θεωρούμε δεδομένη για το γεγονός, και B είναι η επιπλέον πληροφορία. Ο όρος $P(A|B)$ είναι γνωστός ως 'posterior probability' ή αλλιώς η πιθανότητα του A αφού ληφθεί υπόψη το πώς επηρεάζει η επιπλέον πληροφορία B . Ο όρος $P(A)$ ονομάζεται 'prior probability' του A . Ο όρος $P(B|A)$ δίνει την πιθανότητα ενός γεγονότος, δεδομένου της υπόθεσης A και υποθέτοντας ότι οι πληροφορίες από την εμπειρία είναι αληθείς. Ο όρος $P(B)$ είναι ανεξάρτητος από το B και μπορεί να θεωρηθεί ως παράγοντας κανονικοποίησης. Βασική σημείωση είναι ότι αυτές οι πιθανότητες είναι υποθετικές. Ουσιαστικά καθορίζουν το βαθμό της εκτίμησής κάτω από ορισμένες συνθήκες. Έτσι το θεώρημα δεν έχει νόημα χωρίς τις προηγούμενες αναλύσεις για τις

πιθανότητες αυτών των συνθηκών. Έτσι με τη χρήση των ΒΝ ουσιαστικά μπορεί να ορίζεται η αλλαγή μιας 'γνώμης' για την υπόθεση A σε σχέση με το νέο γεγονός B . Πιο συγκεκριμένα η μεταγενέστερη γνώμη (posterior) $P(A|B)$ υπολογίζεται, πολλαπλασιάζοντας την προηγούμενη (prior) γνώμη $P(A)$ με την πιθανότητα (likelihood) $P(B|A)$, ότι δηλαδή το B θα συμβεί αν το A είναι αληθές.

$$P(A | B) * P(B) = P(A, B) \quad (1)$$

$$P(B | A) * P(A) = P(A, B) \quad (2)$$

$$P(A | B) = \frac{P(B | A) * P(A)}{P(B)} \quad (3)$$

Τα Bayesian Networks λοιπόν, παρέχουν τη δυνατότητα μοντελοποίησης μίας σειράς από πιθανά γεγονότα και συνθήκες και τη δοκιμή με επιστημονικό τρόπο υποθέσεων για ένα γεγονός. Όταν υπάρχουν πολλά εναλλακτικά σενάρια τα οποία μπορούν να επηρεάσουν ένα γεγονός, παρέχεται η δυνατότητα να δοκιμάσουμε τα σενάρια αυτά και να δούμε κατά πόσο τα σενάρια αυτά επηρεάζουν το συγκεκριμένο γεγονός καθώς και ποιες είναι οι συνέπειες αυτών των σεναρίων στο τελικό γεγονός. Με τις δοκιμές αυτές καταδεικνύεται πώς (και σε ποιο βαθμό) διάφορα σενάρια μπορούν να επηρεάσουν ένα επιθυμητό γεγονός και ποιοι είναι οι παράγοντες που καθορίζουν τις πιθανότητες να συμβεί. Ακριβώς για αυτό το λόγο είναι κατάλληλα για εφαρμογή του στη μοντελοποίηση της συμπεριφοράς των φοιτητών του ΕΑΠ και ειδικότερα στην ΠΛΗ10 όπως στην εφαρμογή που παρουσιάζεται.

4. ΤΟ ΜΟΝΤΕΛΟ

4.1 Η ΦΙΛΟΣΟΦΙΑ ΤΟΥ ΜΟΝΤΕΛΟΥ

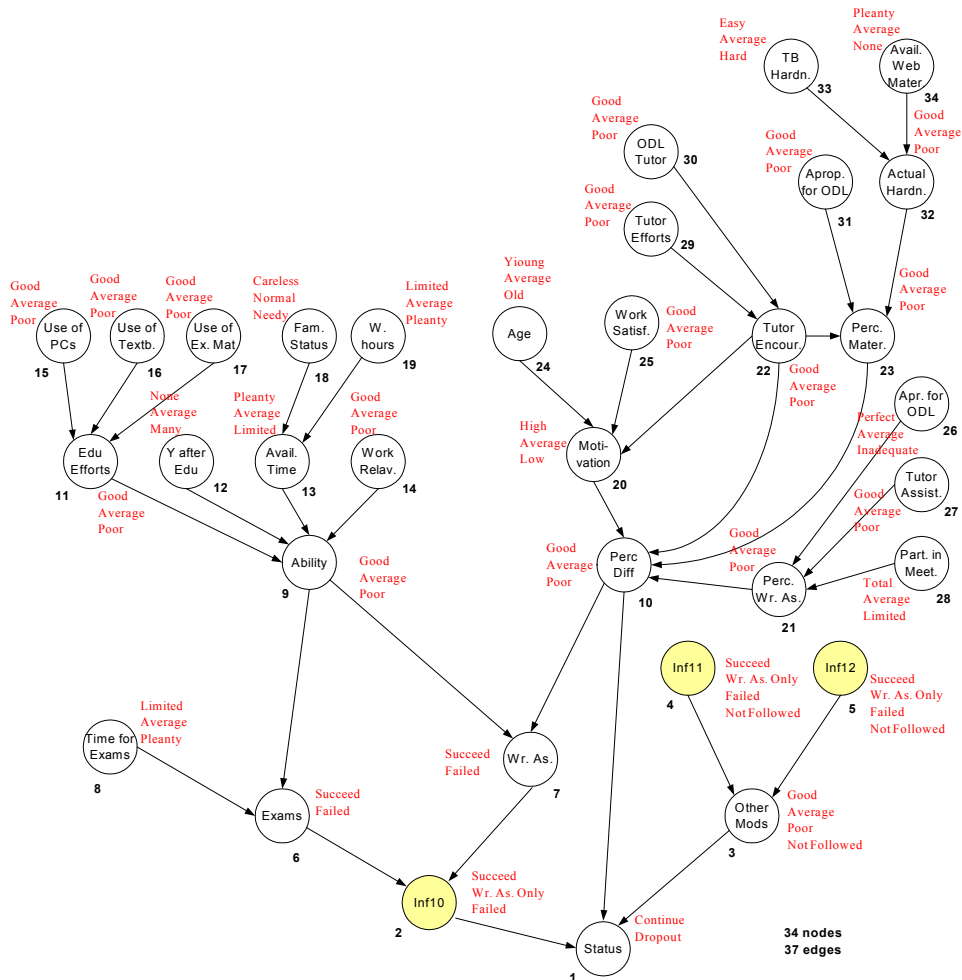
Η φιλοσοφία του προτεινόμενου μοντέλου είναι η δημιουργία ενός, συνεχώς αυξανόμενου δικτύου, το οποίο θα μπορεί να συσσωρεύει την 'εμπειρία' των προηγούμενων χρόνων όπως αυτή θα καταγράφεται από συγκέντρωση και ανάλυση δεδομένων, αλλά και από την εκτίμηση που θα παρέχει το ίδιο το μοντέλο. Στη συγκεκριμένη εφαρμογή παρουσιάζεται η εφαρμογή για ένα τμήμα, αλλά είναι προφανές ότι το ίδιο (μάλιστα με τους ίδιους πίνακες πιθανοτήτων) μπορεί να συμπεριλάβει όλα τα παράλληλα τμήματα για όλες τις ΘΕ. Με αντίστοιχη λογική μπορεί να επεκταθεί σε ολόκληρα Προγράμματα Σπουδών. Βέβαια ένα τέτοιο δίκτυο είναι πολύπλοκο να χτιστεί, αλλά η δυνατότητα τμηματοποίησής του και σταδιακής υλοποίησης όπως στην εφαρμογή που περιγράφεται το κάνει εφικτό.

4.2 ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΚΑΙ ΑΥΤΟΜΑΤΟΠΟΙΗΣΗ

Όπως αναφέρθηκε το μοντέλο που παρουσιάζεται είναι για ένα τμήμα και συγκεκριμένα αυτό του Συντονιστή της ΠΛΗ10. Ο λόγος ήταν να μπορέσει να δοκιμαστεί πιλοτικά η δυνατότητα εφαρμογής του, πριν ζητηθεί να χρησιμοποιηθεί σε ευρύτερη κλίμακα. Ο χρόνος που χρειάζεται για τη σχεδίαση του μοντέλου και την ετοιμασία των πινάκων πιθανοτήτων δεν είναι καθόλου ευκαταφρόνητος και για τους 25 φοιτητές του

τμήματος η εναλλακτική ‘από μνήμης’ ανάλυση δεν θα ήταν καθόλου κακή επιλογή. Όμως από τη στιγμή που το μοντέλο χτίστηκε είναι άμεσα εφαρμόσιμο για όλα τα τμήματα και για τους 800 περίπου νέους φοιτητές της ΠΛΗ10.

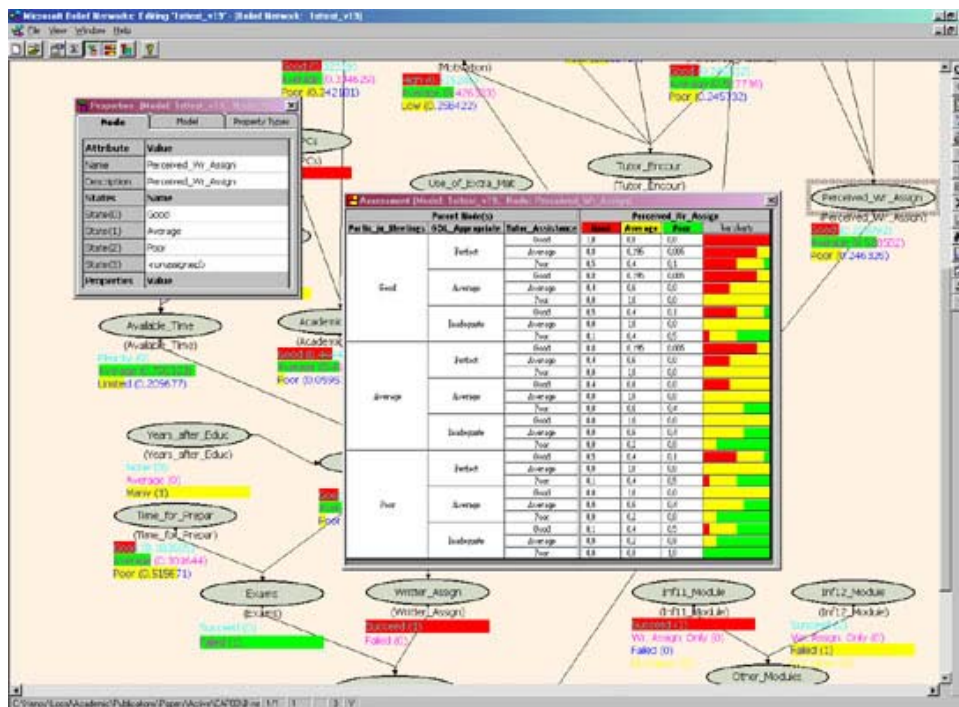
Μια σχηματική αναπαράσταση της σχεδίασης των δικτύων πιθανοτήτων με τις αντίστοιχες καταστάσεις παρουσιάζεται στο σχήμα 3. Στο συγκεκριμένο σχήμα φαίνονται η τελική κατάσταση (αν ο φοιτητής συνεχίζει ή διέκοψε), οι καταστάσεις που επηρεάζουν την τελική (όπως π.χ. η χρήση υπολογιστή, οι ώρες εργασίας, η ενθάρρυνση από τον ΣΚ, κτλ) και οι συσχετίσεις μεταξύ τους. Για λόγους καλύτερης παρουσίασης και αναφοράς σε συγκεκριμένους κόμβους έχουν προστεθεί και οι πιθανές καταστάσεις κάθε κόμβου, αλλά και μια αυθαίρετη αρίθμηση των κόμβων. Αν και το μοντέλο έχει απλοποιηθεί για τους λόγους της πιλοτικής εφαρμογής κάποιοι κόμβοι (όπως π.χ. ο 9 και ο 10) απαιτούν τη συμπλήρωση $3^5=243$ πιθανοτήτων, ενώ άλλοι λιγότερες (π.χ. 64 για τον κόμβο 3, 18 για τον κόμβο 6, κτλ.).



Σχήμα 3. Το σχέδιο του μοντέλου με τις καταστάσεις κάθε κόμβου

Στη σχεδίαση του μοντέλου οι άλλες 2 ΘΕ του πρώτου έτους της Πληροφορικής, δηλαδή οι ΠΛΗ11 και ΠΛΗ12, περιγράφονται ως κόμβοι με 4 δυνατές καταστάσεις. Στην εκτεταμένη σχεδίαση του μοντέλου θα ήταν και αυτές σε πλήρη ανάλυση, αλλά για τις ανάγκες της πιλοτικής εφαρμογής προτιμήθηκε μια συνοπτική εφαρμογή περιορισμένη μόνο στην ΠΛΗ10. Χαρακτηριστικό παράδειγμα της επεκτασιμότητας του μοντέλου είναι ότι οι πίνακες πιθανοτήτων για αυτούς τους κόμβους είναι πρακτικά έτοιμοι, αφού βασίζονται σε αυτούς της ΠΛΗ10 και το μοντέλο θα μπορούσε άμεσα να επεκταθεί (κάνοντας όμως και πιο δύσκολη τη συνοπτική του παρουσίαση στα πλαίσια του άρθρου).

Το μοντέλο υλοποιήθηκε με χρήση του *Microsoft © MSBNx Authoring and Evaluation Tool version 1.4.2*. Η συμπλήρωση των πινάκων πιθανοτήτων βασίστηκε σε δεδομένα και αναλύσεις που έχουν προκύψει από προηγούμενες έρευνες στο ΕΑΠ (Xenos et. al, 2001 και Xenos et. al., 2002), αλλά και από τη διεθνή εμπειρία (Baath, 1994), (Chacon-Duque, 1987), (Kerka, 1996) και που χωρίς αυτά τα δεδομένα θα ήταν αδύνατο να χτιστεί ακόμα και εμπειρικά κάθε πίνακας πιθανοτήτων. Πρέπει βέβαια να τονιστεί, ότι ακόμα και με ελάχιστη πληροφορία οι πίνακες πιθανοτήτων μπορούν να χτιστούν (με μεγάλο ποσοστό αβεβαιότητας) και να χρησιμοποιηθεί το ίδιο το μοντέλο για να οδηγήσει σε βελτιώσεις στους πίνακες, αλλά για καλύτερα αποτελέσματα τίποτε δεν μπορεί να υποκαταστήσει τη συλλογή και ανάλυση δεδομένων από προηγούμενες ακαδημαϊκές χρονιές. Στην εικόνα 1 παρουσιάζεται το γραφικό περιβάλλον του εργαλείου για τη συμπλήρωση του πίνακα πιθανοτήτων ενός 'μέσου' κόμβου, συγκεκριμένα του κόμβου 21 στο σχήμα 3.

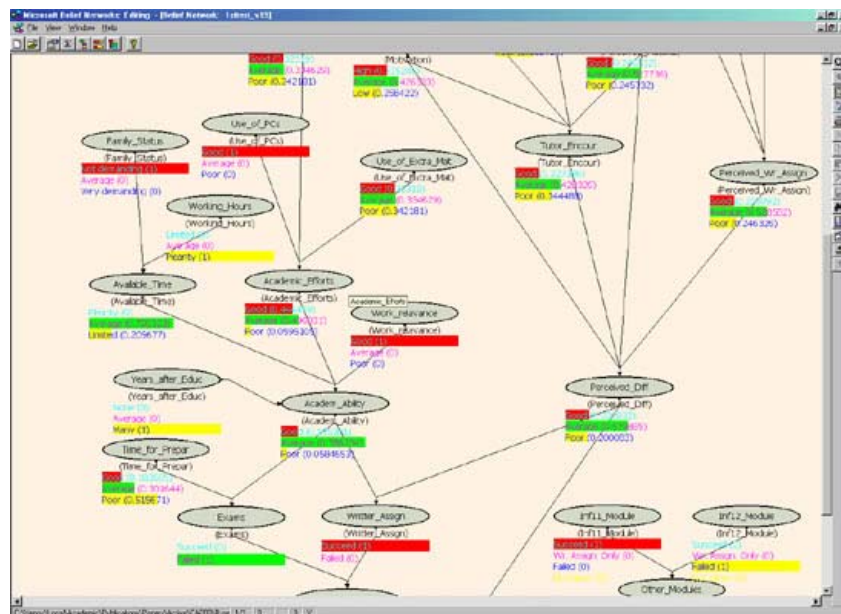


Εικόνα 1. Πίνακας πιθανοτήτων ενός κόμβου

4.3 ΕΦΑΡΜΟΓΗ ΣΤΗΝ ΠΡΑΞΗ

Η εφαρμογή του μοντέλου βασίζεται στην εισαγωγή *στοιχείων (evidence)* για κάποιους κόμβους. Αν κανένα δεδομένο δεν έχει εισαχθεί τότε το μοντέλο δίνει την εκτίμηση –που βασίζεται στην προηγούμενη εμπειρία– ότι κάθε φοιτητής έχει 72% πιθανότητα να συνεχίσει τις σπουδές του και 28% να διακόψει. Η εισαγωγή *στοιχείων* ενεργοποιεί τις πιθανότητες σε κάθε NPT των σχετικών κόμβων με αποτέλεσμα η εκτίμηση να αλλάζει. Για παράδειγμα αν τα μόνα που γνωρίζουμε είναι ότι κάποιος φοιτητής πήρε 2 ΘΕ (ΠΑΗ10 και ΠΑΗ11) και απέτυχε στην ΠΑΗ10 ενώ στην ΠΑΗ11 δεν πέρασε τις εξετάσεις, αλλά κατοχύρωσε τις γραπτές εργασίες, η παραπάνω πιθανότητα μεταβάλλεται σε 32%. Η εισαγωγή και νέων *στοιχείων* θα αλλάξει τις εκτιμήσεις για τους τελικούς κόμβους.

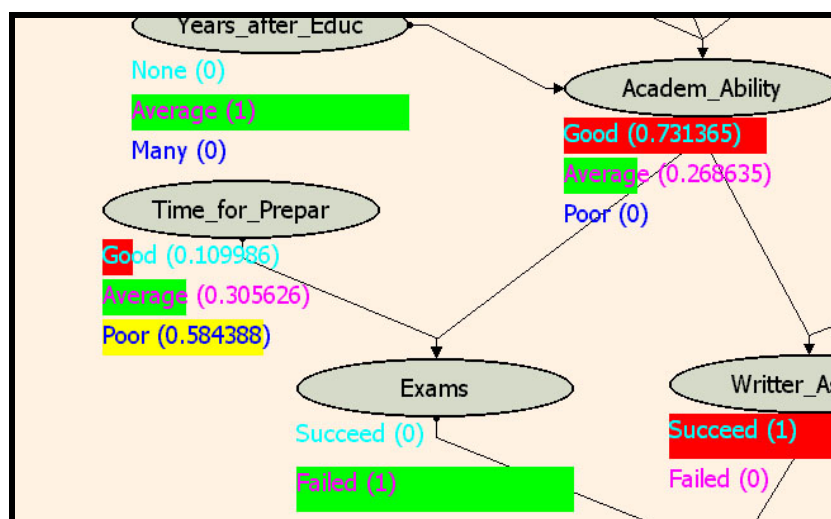
Αυτού του είδους η εφαρμογή του μοντέλου ονομάζεται *forward* και χρησιμοποιείται για να εκτιμηθούν και να προληφθούν καταστάσεις. Για παράδειγμα εάν στη μέση της χρονιάς, και ενώ έχουν εισαχθεί αρκετά *στοιχεία*, κάποιος φοιτητής έχει πιθανότητα διακοπής φοίτησης 45%, αυτό το υψηλό ποσοστό θα μπορούσε να οδηγήσει το ΣΚ να αλλάξει προσέγγιση απέναντι του. Μάλιστα δοκιμάζοντας θεωρητικά κάποιες ενέργειες που κρίνει σκόπιμο να κάνει (τοποθετώντας τις ως *στοιχεία* δηλαδή) μπορεί να δει ποια από αυτές θα είχε το καλύτερο αποτέλεσμα. Από την άλλη, αν ένα ολόκληρο τμήμα είχε εκτίμηση για πολύ υψηλά ποσοστά διακοπής, τότε πιθανόν τα αίτια να ήταν διαφορετικά και να χρειαζόταν συνολική αλλαγή προσέγγισης. Πρέπει να τονιστεί ότι το μοντέλο είναι δυναμικό, δηλαδή σε κάθε ένα από τα κομβικά σημεία (ή σε κάθε σημείο που υπάρχουν νέα *στοιχεία*) τα *στοιχεία* αυτά εισάγονται στο μοντέλο και νέες εκτιμήσεις προκύπτουν. Ένα άλλο σημαντικό σημείο είναι ότι οι πιθανότητες σε κάθε NPT μπορούν να ενημερώνονται συνεχώς (η δυνατότητα αυτή ονομάζεται *propagation*) και έτσι η ποιότητα της εκτίμησης να βελτιώνεται συνεχώς αξιοποιώντας την προηγούμενη εμπειρία.



Εικόνα 2. Παράδειγμα εκτίμησης

Στην εφαρμογή του για *forward* εκτίμηση το μοντέλο έδωσε πολύ ικανοποιητικά αποτελέσματα, εκτιμώντας σωστά όλες τις περιπτώσεις εκτός από μία. Αυτή ήταν η περίπτωση μίας φοιτήτριας που διέκοψε τη χρονική στιγμή που η πιθανότητα διακοπής φοίτησης όπως την έδινε το μοντέλο ήταν στο 16,5%. Ο λόγος διακοπής ήταν μία μετάθεση στο εξωτερικό (Αμερική) και την οδήγησε στην απόφαση να μη συνεχίσει τις σπουδές της στο ΕΑΠ. Ακόμα και σε αυτή την περίπτωση βέβαια, το 16,5% σημαίνει ότι υπάρχει πάντα πιθανότητα κάποιου απρόβλεπτου (μη μοντελοποιημένου) γεγονότος, αν και μία νέα σχεδίαση θα μπορούσε να συνεκτιμά παράγοντες για να προβλέπει ακόμα και περιπτώσεις με τέτοιες ενδείξεις. Στην εικόνα 2 παρουσιάζεται ένα παράδειγμα από την εφαρμογή του μοντέλου. Σε κάποιους κόμβους έχουν δοθεί *στοιχεία* και κατά συνέπεια σε αυτούς κάποια κατάσταση είναι δεδομένη και έχει πιθανότητα 1 (στην εικόνα 2 αυτό φαίνεται για τους αντίστοιχους κόμβους 4, 14, 15, 18 του σχήματος 2), ενώ για τις υπόλοιπες το μοντέλο δίνει εκτιμήσεις.

Αντίστοιχα με τη *forward* χρήση του μοντέλου, υπάρχει και η χρήση του για *αποτίμηση* μίας κατάστασης όταν υπάρχουν οριστικά *στοιχεία* για μία τελική κατάσταση. Αυτή η χρήση του μοντέλου καλείται *backward*. Έστω για παράδειγμα ότι ξέρουμε ότι ένας φοιτητής διέκοψε τις σπουδές του (άρα η πιθανότητα διακοπής φοίτησης είναι 1) και γνωρίζουμε *στοιχεία* για κάποιους αρχικούς κόμβους (όπως π.χ. για το αν χρησιμοποιεί Η/Υ στη δουλειά του, οι ώρες που εργάζεται, εάν παρακολουθεί άλλες ΘΕ, κτλ), τότε μπορούμε να κάνουμε *αποτίμηση* για κάποιους ενδιάμεσους ή αρχικούς κόμβους. Μπορούμε για παράδειγμα να αποτιμήσουμε εάν είχε υψηλό κίνητρο για τις σπουδές του, το πόσο δύσκολες ή εύκολες αντιλαμβανόταν τις γραπτές εργασίες, σε ποιο βαθμό εισέπραττε ενθάρρυνση από το ΣΚ, κτλ.



Εικόνα 3. Παράδειγμα αποτίμησης μίας κατάστασης

Στην εφαρμογή του μοντέλου για *backward* χρήση (δηλαδή αποτίμηση) τα αποτελέσματα είχαν την ιδιαιτερότητα ότι σε μερικές περιπτώσεις δεν υπάρχει μη-υποκειμενικός τρόπος επιβεβαίωσης. Για παράδειγμα, πώς επιβεβαιώνεται αντικειμενικά η κατάσταση του κόμβου 'motivation'; Παρόλα αυτά, σε όσες περιπτώσεις υπήρχαν ενδείξεις,

το μοντέλο ξεπέρασε ακόμα και τις πιο αισιόδοξες τις αρχικές προσδοκίες. Στην εικόνα 3 παρουσιάζεται το παράδειγμα ενός φοιτητή που πέτυχε στις γραπτές εργασίες, αλλά όχι και στις εξετάσεις (αυτά έχουν εισαχθεί ως *evidence*, στους αντίστοιχους κόμβους, όπως και άλλα στοιχεία που ήταν διαθέσιμα). Το μοντέλο για τη συγκεκριμένη περίπτωση έδινε πολύ υψηλή πιθανότητα ελλιπούς προετοιμασίας για τις τελικές εξετάσεις. Όπως φαίνεται στην εικόνα 3, έδινε πιθανότητα καλής προετοιμασίας μόνο 0,109 (δηλαδή περίπου 11%) όταν το αντίστοιχο ποσοστό για άλλους φοιτητές που επίσης δεν πέτυχαν στις εξετάσεις δεν ήταν σε καμία περίπτωση μικρότερο του 20%. Η αποτίμηση αυτή επιβεβαιώθηκε με τηλεφωνική συνέντευξη που αποκάλυψε ότι τον τελευταίο μήνα πριν από τις εξετάσεις υπήρξε σοβαρό πρόβλημα υγείας μέλους της οικογένειάς του με αποτέλεσμα να μην προετοιμαστεί σχεδόν καθόλου.

5. ΣΥΜΠΕΡΑΣΜΑΤΑ – ΜΕΛΛΟΝΤΙΚΗ ΣΥΝΕΧΕΙΑ ΤΗΣ ΔΟΥΛΕΙΑΣ

Συνοψίζοντας, η αξία του μοντέλου σχετίζεται με την ποιότητα της σχεδίασης (που στην πιλοτική εφαρμογή είναι σημαντικά απλοποιημένη) και την ποιότητα των πινάκων πιθανοτήτων. Για να μπορέσει να το χρησιμοποιήσει κάποιος, πρέπει να έχει κάνει αναλυτική συλλογή δεδομένων και μετά να χρησιμοποιήσει τα δεδομένα αυτά για να προσαρμόσει το μοντέλο στις ιδιαιτερότητες της περίπτωσης του, καθώς και να υπολογίσει τους πίνακες πιθανοτήτων. Για τη σχεδίαση και την υλοποίηση απαιτείται σημαντική δουλειά, αλλά το κέρδος στην αυτοματοποίηση εκτιμήσεων και αποτιμήσεων είναι σημαντικό. Ειδικά σε περιπτώσεις που η αποτίμηση πρέπει να λάβει υπόψη της εκατοντάδες δεδομένα από πολλές πηγές (όπως στην περίπτωση των 800 φοιτητών μίας ΘΕ) η χρήση μοντελοποίησης και του αυτοματισμού που παρέχει είναι επιβεβλημένη.

Πρέπει να τονιστεί ότι το μοντέλο δεν κάνει εκτιμήσεις και αποτιμήσεις ‘από μόνο του’! Αυτό που κάνει είναι να μοντελοποιεί την εμπειρία του σχεδιαστή του στην εκπαιδευτική διαδικασία και να αξιοποιεί δεδομένα που έχουν ήδη συγκεντρωθεί ή που προκύπτουν από την ίδια τη χρήση του. Στην εφαρμογή του σε ένα τμήμα έδωσε αποτελέσματα πολύ ικανοποιητικά, αλλά μπορεί να αποτελέσει ακόμα πιο σημαντικό εργαλείο για εξαγωγή συμπερασμάτων σε ευρύτερη κλίμακα, που σχετίζονται με την αξία του εκπαιδευτικού υλικού και των μεθόδων, τους βασικούς λόγους διακοπής της φοίτησης κτλ. Κυριότερο πλεονέκτημα του μοντέλου είναι ότι είναι τμηματοποιήσιμο (*modular*) και μπορεί να χτίζεται τμήμα τμήμα και τα δεδομένα από την εφαρμογή του να ανατροφοδοτούνται στους πίνακες πιθανοτήτων. Αυτό που παρουσιάστηκε στα πλαίσια του άρθρου είναι η πιλοτική του εφαρμογή στα πλαίσια ενός μόνο τμήματος, ενώ η πιο εκτεταμένη εφαρμογή του είναι ένας μελλοντικός στόχος. Άλλος μελλοντικός στόχος είναι η επαύξηση της σχεδίασης σε επίπεδο *module* με την αξιοποίηση της εμπειρίας από τη φετινή εφαρμογή και η ανατροφοδότηση των πινάκων πιθανοτήτων.

ΑΝΑΦΟΡΕΣ

Baath, J., (1994). ‘*Assignments in Distance Education – An Overview*’, *Epistolodidaktika*, No. 1, pp. 13-20.

- Chacon-Duque, F.J., (1987). *'A Multivariate Model for Evaluating Distance Higher Education'*, College Park: Pennsylvania State University Press.
- Ginou E., (2001), *'Διακοπή της φοίτησης: ερμηνείες, εκτιμήσεις'*, 1ο Πανελλήνιο Συνέδριο για την Ανοικτή και εξ Αποστάσεως Εκπαίδευση, Πάτρα.
- Jensen F., (1996), *'An introduction to Bayesian Networks'*, UCL Press.
- Kerka, S., (1996). *'Distance Learning, the Internet and the World Wide Web'*, ERIC Digests, ERIC Document Reproduction Service No. ED395214.
- Kschischang F., Frey B., Loeliger H., (2001), *'Factor graphs and the sum product algorithm'*, IEEE Transactions on Information Theory, February.
- Lauritzen S., Spiegelhalter J., (1990), *'Local computations with probabilities on graphical structures and their applications to expert systems'*, Readings in Uncertain Reasoning, ed. G. Shafer Γ., Pearl Morgan Kaufmann J., San Mateo, California, pp. 415-448.
- Munro J., (1987), *'The Discourse of Dropout in Distance Education: A Theoretical Analysis'*, Annual Conference of the Canadian Association for the Study of Adult Education, Alberta, Canada.
- Narasimharao B., (2000), *'Issues in preparing Open University learners for Open University system'*, Available at: <http://www.cemca.org/ignou-icde/paper23.html>.
- Pierrakeas C., Xenos M., Pintelas P., (2001), *'Εκτίμηση – Αξιολόγηση του Εκπαιδευτικού Υλικού – Συμβούλου Καθηγητή. Απόψεις των Φοιτητών – Σημεία Βελτίωσης'*, 1ο Πανελλήνιο Συνέδριο για την Ανοικτή και εξ Αποστάσεως Εκπαίδευση, Πάτρα.
- Pli10, (2002), *'Σελίδες φοιτητών της Θεματικής Ενότητας Εισαγωγή στην Πληροφορική (ΠΛΗ10)'*, Available at: <http://edu.eap.gr/pli/pli10/students.htm>.
- Race P., (1999), *'Το Εγχειρίδιο της Ανοικτής Εκπαίδευσης'*, Εκδόσεις Μεταίχμιο, Αθήνα, ISBN 960-375-014-X.
- Rumble G., (1992), *'The Management of Distance Learning Systems'*, Paris, UNESCO / ΙΙΕΡ.
- Shin N., Kim J., (1999), *'An exploration of learner progress and drop-out in Korea National Open University'*, Distance Education – An International Journal, Vol. 20, No. 1.
- Xenos M., Pierrakeas C., Pintelas P., (2001), *'Εντοπισμός Ομάδων Φοιτητών του Ε.Α.Π. με Υψηλή Πιθανότητα Διακοπής Φοίτησης - Εμπειρίες και Ιδιαιτερότητες στην Πληροφορική'*, 1ο Πανελλήνιο Συνέδριο για την Ανοικτή και εξ Αποστάσεως Εκπαίδευση, Πάτρα.
- Xenos M., Pierrakeas C., Pintelas P., (2002), *'A survey on student dropout rates and dropout causes concerning the students in the Course of Informatics of the Hellenic Open University'*, International Journal of Computers & Education, Vol. 39 (4), pp. 361-377.