

# Εργασία 4<sup>η</sup>

Διονύσης Κοτζαΐτης

Μελετήστε την εργασία

*Pavlik, P.I., Cen, H. & Koedinger, K.R. (2009). Performance Factors Analysis -- A New Alternative to Knowledge Tracing. Presented at 14th International Conference on Artificial Intelligence in Education 2009.*

και απαντήστε τις παρακάτω ερωτήσεις. (Αποφεύγετε να αντιγράφετε σε μεγάλο βαθμό αυτούσια τμήματα της εργασίας ως απαντήσεις στις ερωτήσεις. Οι απαντήσεις να είναι ακριβείς, περιεκτικές. Αποφεύγετε αυτόματες μεταφράσεις όπως από Google translate.) Πληροφορίες για το overlay model θα βρείτε μετά τις ερωτήσεις. **ΟΙ ΑΠΑΝΤΗΣΕΙΣ ΘΑ ΠΡΕΠΕΙ ΝΑ ΔΟΘΟΥΝ ΣΤΑ ΕΛΛΗΝΙΚΑ. ΜΠΟΡΕΙΤΕ ΝΑ ΔΙΑΤΗΡΗΣΕΤΕ ΟΡΟΥΣ ΣΤΑ ΑΓΓΛΙΚΑ ΑΝ ΔΕΝ ΓΝΩΡΙΖΕΤΕ ΤΗ ΔΟΚΙΜΗ ΑΠΟΔΟΣΗ ΤΟΥΣ.**

## 1. Διατυπώστε το στόχο της εργασίας/μελέτης (Introduction)

Στόχος της μελέτης είναι η εύρεση ενός συστήματος μοντελοποίησης και πρόβλεψης γνώσης, που θα είναι closed loop. Αυτά τα συστήματα ονομάζονται έτσι, καθώς απαιτούν πολύ λίγη τεχνική γνώση για να προστεθούν αντικείμενα στο σύστημα, να μαζέψουμε δεδομένα για τα αντικείμενα αυτά και να τα χρησιμοποιήσουμε για να εκπαιδύσουμε μοντέλα με τελικό σκοπό να κρατήσουμε μόνο τα αντικείμενα που θα είναι καλώς αποτυπωμένα στην μαθητική διαδικασία. Με λίγα λόγια, οι ερευνητές καλούνται να κάνουν την μετατροπή ενός ήδη υπάρχοντος μοντέλου εξόρυξης δεδομένων διδασκαλίας και να συγκριθεί με τα κλασικά μοντέλα (LFA, KT)

## 2. Το αρχικό εκπαιδευτικό μοντέλο εξόρυξης δεδομένων είναι το LFA. Ποιους βασικούς παράγοντες αποτυπώνει

Το βασικό εκπαιδευτικό μοντέλο που μελετάνε οι συγγραφείς είναι το LFA (Learning Factor Analysis), το οποίο χρησιμοποιείται για διάφορους σκοπούς ως εργαλείο εξόρυξης δεδομένων (πχ ως αλγόριθμος αναζήτησης για να χωρίσει τα συστατικά μάθησης σε διάφορους παράγοντες, έτσι ώστε να καθοριστεί την αντιστοίχιση του συστατικού μάθησης με την γνώση/γνωστικό αντικείμενο). Γι' αυτό το μοντέλο έχουμε 3 βασικούς παράγοντες:

- Την ικανότητα του μαθητή σε μια πολύ συγκεκριμένη παράμετρο για κάθε μάθημα.
- Την ευκολία του συστατικού μάθησης που μελετάμε.
- Τον ρυθμό μάθησης για κάθε συστατικό μάθησης, χρησιμοποιώντας μία μόνο παράμετρο για κάθε συστατικό.

## 3. Ποιο είναι η αδυναμία του LFA

Παρ' όλο που το LFA έχει σημαντική δυνατότητα να μπορεί να εκπαιδευτεί με τα δεδομένα μέσω αυτής της διατύπωσης, δεν μπορεί να αλλάξει δυναμικά ανάμεσα σε διαφορετικά συστατικά μάθησης, καθώς δεν λαμβάνει υπόψιν τις σωστές και λάθος απαντήσεις του μαθητή. Με λίγα λόγια, θεωρεί ότι οι μαθητές μαθαίνουν όλοι με τον ίδιο τρόπο. Αυτό κάνει το σύστημα αρκετά ευαίσθητο στον παράγοντα της συχνότητας εξάσκησης.

## 4. Ποιες είναι οι 4 παράμετροι του Knowledge Tracing (KT); Ποιο είναι το πλεονέκτημα αυτού του μοντέλου; Τι αναφέρουν σχετικά με την προσαρμογή αυτών των παραμέτρων; Πως συνδέεται το KT model με τα Knowledge Components (KC) σε συστήματα διδασκαλίας (tutors) ;

Οι 4 παράμετροι του Knowledge Tracing είναι:

- Αρχικός ρυθμός μάθησης.
- Ρυθμός μάθησης.
- Παράμετρος τυχαίας απάντησης.
- Παράμετρος ατυχούς απάντησης.

Ένα πολύ θετικό είναι ότι αυτές οι παράμετροι είναι εύκολα ερμηνεύσιμες και έτσι είναι εύκολο να αντιληφθούμε τα αποτελέσματα τους στο προβλεπτικό μοντέλο. Για να γίνει fit των παραμέτρων στο μοντέλο, χρησιμοποιούμε παρελθοντικά δεδομένα των μαθητών για κάθε συστατικό μάθησης σε κάποιο τομέα, έτσι μπορούν να χρησιμοποιηθούν ιστορικά δεδομένα του μαθητή, τα οποία μπορούν να καταγραφούν από ένα σύστημα διδασκαλίας και να αξιοποιηθούν εκ των υστέρων στην εκπαίδευση του μοντέλου σε ατομικό επίπεδο, για κάθε μαθητή προσωποποιημένα.

## 1. Performance Factors Analysis

### 5. Δώστε τις εξισώσεις για το LFA model. Εξηγείστε το ρόλο κάθε παραμέτρου

Οι εξισώσεις είναι:

- $m(i, j \in KCS, n) = a_i + \sum_{j \in KCS} (\beta_j + \gamma_j n_{i,j})$ 
  - Το  $m$  είναι η logit τιμή που αντιπροσωπεύει την συσσωρευτική γνώση του μαθητή  $i$ .
  - Η παράμετρος  $a$  είναι η ικανότητα του μαθητή σε συστατικά μάθησης (ένα ή περισσότερα).
  - Η παράμετρος  $\beta$  είναι η ευκολία του κάθε συστατικού μάθησης.
  - Το όφελος της συχνότητας των πρότερων ενεργειών για κάθε συστατικού μάθησης εκφράζεται με τα  $n$  προηγούμενα observations και την σταθερά  $\gamma$ , για κάθε συστατικό μάθησης.
- $p(m) = \frac{1}{1+e^{-m}}$ 
  - Το παραπάνω εκφράζει την πιθανότητα, μετατρέποντας τις τιμές του  $m$  για κάθε μαθητή σε πιθανότητα.

Το fit των παραμέτρων γίνεται με την τεχνική max likelihood.

### 6. Δώστε τις εξισώσεις για το PFA model. Εξηγείστε το ρόλο κάθε παραμέτρου

Οι εξισώσεις είναι:

- $m(i, j \in KCS, s, f) = \sum_{j \in KCS} (\beta_j + \gamma_j s_{i,j} + \rho_j f_{i,j})$ 
  - Το  $m$  είναι η logit τιμή που αντιπροσωπεύει την συσσωρευτική γνώση του μαθητή  $i$ .
  - Η παράμετρος  $\beta$  είναι η ευκολία του κάθε συστατικού μάθησης.
  - Το  $s$  εκφράζει τις επιτυχίες του μαθητή  $i$  στο γνωστικό αντικείμενο  $j$  και το  $f$  είναι αντίστοιχο για τις αποτυχίες.
  - Το  $\gamma$  και το  $\rho$  είναι σταθερές με τις οποίες γίνεται scale των επιτυχιών/ αποτυχιών.
- $p(m) = \frac{1}{1+e^{-m}}$ 
  - Το παραπάνω εκφράζει την πιθανότητα, μετατρέποντας τις τιμές του  $m$  για κάθε μαθητή σε πιθανότητα.

Το fit των παραμέτρων γίνεται με την τεχνική max likelihood.

### 7. Το LFA σε τι διαφέρει από το LFAns;

Η διαφορά ανάμεσα στο LFA και στο LFAns είναι ότι ενώ το LFAns είναι παρόμοιο με το LFA, δεν περιέχει την ικανότητα του μαθητή για το συγκεκριμένο component ( $\alpha$ ). Είναι πιο απλό μοντέλο.

### **8. Ποιος είναι ο σκοπός της μελέτης σύγκρισης των 3 μοντέλων που διενεργούν οι συγγραφείς;**

Ο σκοπός της σύγκρισης των 3 αυτών μοντέλων είναι ότι επιτρέπει τους συγγραφείς να δουν πόσο καλύτερο είναι το PFA σε σχέση με το LFAns μοντέλο, το οποίο είναι πανομοιότυπο με το LFA αλλά λαμβάνοντας υπόψιν μόνο τις νέες επιδόσεις, με τους συγγραφείς να αναμένουν κατώτερο επίπεδο όσον αφορά την αποδοχή του μοντέλου. Από την άλλη το LFA, μπορεί να «πιάνει» τις διαφορές ανάμεσα στους μαθητές (λόγω α) και η σύγκριση με το PFA μας δίνει μια γνώση για το πόσο προσαρμοστικό είναι το μοντέλο.

### **9. Ποια σύνολα δεδομένων χρησιμοποίησαν για τη μελέτη τους;**

Χρησιμοποιήθηκαν 4 σύνολα δεδομένων για την μελέτη των μοντέλων. Τα δεδομένα ήταν από διαφορετικές πηγές και ήταν τα παρακάτω:

- Κλάσματα. Τάξεις 5-8. Πηγή: Bridge to Algebra Cognitive Tutor, Carnegie Learning.
- Άλγεβρα. Τάξεις 5-8. Πηγή: Bridge to Algebra Cognitive Tutor, Carnegie Learning.
- Γεωμετρία. Τάξεις 9-12. Πηγή: Angles and Areas units of the Carnegie Learning Geometry Cognitive Tutor.
- Φυσική. Τάξεις 9-12. Πηγή: Andres Physics tutoring system, VanLehn.

Τα σετ για την γεωμετρία και την φυσική, έγιναν download από το Pittsburgh Science of Learning Center DataShop. Όλα τα σετ δεδομένων, είχαν ένα προκαθορισμένο πεδίο ορισμού για το περιεχόμενο του γνωστικού συστατικού για κάθε επίδοση.

### **10. Ο Table 1 τι παρουσιάζει;**

Το table1 παρουσιάζει την οργάνωση των αρχείων για ένα μικρό παράδειγμα για κάποιο συστατικό μάθησης σε έναν μαθητή. Έχουμε το id του μαθητή, έναν υποστόχο του προβλήματος, για κάθε εγγραφή του KC, αν είναι σωστή η απάντηση, το skill το οποίο συσχετίζεται με την ερώτηση και μια τιμή πρόβλεψης η οποία υπολογίζεται για τον μαθητή από το KC, τον υποστόχο και το αν ήταν σωστή η απάντηση.

### **11. Ο Table 2 τι παρουσιάζει;**

Ο πίνακας 2 παρουσιάζει την σύγκριση των μοντέλων για τα 4 διαφορετικά σετ δεδομένων. Στην σύγκριση έχουμε κάποια χαρακτηριστικά:

- Τον αριθμό των παραμέτρων.
- Τον αριθμό των παρατηρήσεων.
- Το log likelihood για το κάθε μοντέλο σε κάθε dataset.
- Το Bayes Information Criterion για το κάθε μοντέλο σε κάθε dataset.
- Την μέση απόλυτη απόκλιση για μία 7-fold cross validation, που δείχνει αν το μοντέλο γενικεύεται.

### **12. Από τα αποτελέσματα του Table 2 ποια συμπεράσματα συνάγουν οι συγγραφείς; (Στην απάντησή σας θα πρέπει να αναφέρετε τις διάφορες μετρικές που χρησιμοποιούν οι συγγραφείς για να διατυπώσουν τα συμπεράσματά τους).**

Τα αποτελέσματα που προκύπτουν είναι:

- Από το LL: Το LFA είναι οριακά καλύτερο στα  $\frac{3}{4}$  των σέτ δεδομένων (Φυσική, Γεωμετρία και Κλάσματα). Αυτό είναι αναμενόμενο, εφόσον το LFA έχει την παράμετρο α, μπορούμε να δούμε ότι το PFA είναι αρκετά κοντά στο LL, και σε μία περίπτωση το ισοφαρίζει.
- Από το BIC: Γενικά και εδώ έχουμε ότι στα  $\frac{3}{4}$  των dataset το LFA είναι καλύτερο, με την διαφορά ότι στα κλάσματα το PFA είναι καλύτερο απ' το LFA. Γενικά φαίνεται ότι το overfit κινείται στα ίδια επίπεδα σε όλα τα μοντέλα.

Γενικά και στις 2 μετρικές, φαίνεται το LFA να είναι καλύτερο, αλλά πολύ οριακά σε σχέση με το PFA.

- Από το MAD CV: Είναι δείκτης που μαζί με το BIC, δείχνει το πόσο μπορεί να γενικευτεί το μοντέλο, κάτι που φαίνεται ότι το PFA κάνει αρκετά καλά.

Εν τέλει, το PFA φαίνεται να είναι μια καλή εναλλακτική για την διαδικασία εντοπισμού και αντίδρασης στην κατά την μάθηση σε κάποιο «παιδαγωγό».

## **2. Performance Factors Analysis compared to Knowledge Tracing**

### **13. Σε τι διαφέρουν τα δεδομένα που χρησιμοποιήθηκαν στη μελέτη σύγκρισης των 3 μοντέλων (LFA, PFA, PFAns) από τα δεδομένα στη μελέτη σύγκρισης των μοντέλων KT1, KT2, PFA**

Τα δεδομένα που χρησιμοποιήθηκαν και προηγουμένως χρησιμοποιήθηκαν και τώρα, παρ' όλα αυτά, η διαφορά ανάμεσα στα 2 είναι ότι στα πρώτα είχαμε πολλαπλά γνωστικά συστατικά. Από την άλλη, στην δεύτερη περίπτωση, έχουμε τα δεδομένα να σπάνε σε μοναδικά γνωστικά συστατικά, ανά πέρασμα μοντέλου, καθώς το KT δεν μπορεί να εκπαιδευτεί με περισσότερα από ένα γνωστικό συστατικό ανά πέρασμα.

### **14. Οι συγγραφείς αναφέρουν «Unlike PFA, which takes the form of standard logistic regression and therefore allows fast optimization by computing the true solution gradient to maximize the loglikelihood, KT has less well explored options for finding parameters....» Με βάση αυτήν τη διαπίστωση γιατί στη συγκριτική μελέτη χρησιμοποιούν 2 υλοποιήσεις του μοντέλου KT; Και τι παραδοχές κάνουν όσον αφορά τις τιμές των παραμέτρων για τα μοντέλα που συγκρίνουν;**

Επειδή ο PFA χρησιμοποιεί στανταρ λογιστική παλινδρόμηση για την εκπαίδευση του μοντέλου μέσω max log likelihood, ενώ το KT όχι, οι συγγραφείς έψαξαν πιο σύνθετες λύσεις. Χρησιμοποιήθηκαν 2 διαφορετικές υλοποιήσεις για το KT1 μοντέλο και μια για το KT 2 μοντέλο. Οι δύο αυτές υλοποιήσεις, χρησιμοποιήθηκαν ώστε να μην υπάρχει περίπτωση να ειπωθεί ότι το fit της μεθόδου μας με τις παραμέτρους είναι αποτέλεσμα οποιουδήποτε πλεονεκτήματος βρήκαμε. Επιπλέον, επειδή το PFA σπανίως παράγει αποτελέσματα κοντά στο 0 ή στο 1, σε αντίθεση με το KT που μπορεί να παράξει τα νούμερα αυτά (προσεγγιστικά) και να κάνει δύσκολη την σύγκριση των μοντέλων, χρησιμοποιούμε 2 παραδοχές για να μπορέσουμε να κάνουμε σύγκριση των τιμών:

- Βάζουμε «ταβάνι» το 0.99 και «πάτωμα» το 0.01 για την τιμή της πρόβλεψης στο μοντέλο -όταν εφαρμόζουμε max log likelihood στους αλγορίθμους.
- Το PFA οριοθετείται ανάμεσα στις τιμές:
  - Minimum=0 για τον δείκτη  $\gamma$ , ώστε να αποφύγουμε το overfit από αρνητικά learning rates.
  - Max πιθανότητα για απροσεξία ορίστηκε το 0.1.
  - Max πιθανότητα για τυχαία απάντηση ορίστηκε το 0.3.
  - Στο KT2 ορίστηκαν max για αρχικό ρυθμό μάθησης το 0.85 και max για πιθανότητες μετάβασης (δεδομένου ότι το μοντέλο είναι βασίζεται σε ένα Μαρκοβιανό μοντέλο 2 καταστάσεων) ίση με 0.3.

### **15. Ο Table 3 τι παρουσιάζει;**

Το table 3 παρουσιάζει την σύγκριση των μοντέλων KT1, KT2, KT2LL (για την γεωμετρία και τα κλάσματα) και PFA για όλα τα σέτ δεδομένων. Οι μετρικές που βλέπουμε στο πίνακα 3 είναι:

- Αριθμός παραμέτρων ανα μοντέλο.
- Αριθμός παρατηρήσεων.
- Log Likelihood του κάθε μοντέλου (LL).
- Bayes Information Criterion (BIC).
- r statistic.
- A' δείκτης .

### **16. Γιατί υλοποίησαν και το KT2LL μοντέλο ως παραλλαγή του KT2;**

Αρχικά οφείλουμε να αναφέρουμε ότι το KT2 μοντέλο χρησιμοποίησε εξαντλητική οριοθετημένη αναζήτηση, με fitness function το άθροισμα των τετραγωνικού λάθους (διαφορά ανάμεσα στην πραγματική τιμή και στην αναμενόμενη τιμή και υψωμένη στο τετράγωνο). Για να δουν οι συγγραφείς αν η διαφορά ήταν υπεύθυνη για την διαφορά στο fit των μοντέλων, τροποποιήσαμε το KT2 μοντέλο για να εκπαιδευτεί μέσω Log Likelihood, αντί στο μέσο τετραγωνικό σφάλμα. Έτσι από την τελική σύγκριση του KT2 και του KT2LL φαίνεται ότι όταν χρησιμοποιείται εκπαίδευση των δεδομένων με μέσο τετραγωνικό σφάλμα δεν οδηγεί σε λάθος εκπαίδευση σε σχέση με τα άλλα μοντέλα.

### **17. Από τα αποτελέσματα του Table 3 ποια συμπεράσματα συνάγουν οι συγγραφείς ;**

Τα αποτελέσματα του table3 δείχνουν ότι:

- Οι διαφορές ανάμεσα στα μοντέλα δεν είναι μεγάλες.
- Το μοντέλο PFA έχει καλύτερους δείκτες LL, BIC, A' και r. Αυτό, σε συνδυασμό με το αποτέλεσμα της ερώτησης 16, δείχνει ότι το μοντέλο PFA είναι γενικά οριακά καλύτερο.

### **18. Ποια συμπεράσματα συνάγουν από την ανάλυση των καμπυλών μάθησης (learning curves) κάθε μοντέλου.**

Στην σύγκριση καμπυλών μάθησης των δύο μοντέλων ανάμεσα στα δεδομένα της γεωμετρίας και των κλασμάτων, και συγκεκριμένα την προβλεπτική ικανότητα για κάθε επαναλαμβανόμενη παρατήρηση με δεδομένη συνθήκη για κάθε γνωστικού αντικειμένου την προηγούμενη απάντηση στο αντικείμενο αυτό, οι συγγραφείς ανέπτυξαν τα εξής αποτελέσματα:

- Το KT μοντέλο είχε την τάση να προβλέπει πολύ χειρότερα, αν υπάρχει αποτυχία, σε αντίθεση με το PFA.
- Ο παρατηρήσιμος μέσος όρος των δεδομένων βρίσκεται ανάμεσα στις προβλέψεις των δυο μοντέλων αλλά πολύ πιο κοντά στο PFA. Αυτό το αποτέλεσμα προκύπτει από την ιδιότητα του KT μοντέλου ότι μία αποτυχία σε απάντηση υποδεικνύει ότι η δεν έχουμε μάθει το γνωστικό συστατικό.

## **3. Conclusion**

### **19. Στην ενότητα Conclusion οι συγγραφείς συνοψίζουν τα συμπεράσματα της εργασίας τους. Επιλέξτε να αναφέρετε το σημαντικότερο, κατά τη γνώμη σας, συμπέρασμα.**

Κατά την άποψη μου, το σημαντικότερο συμπέρασμα είναι ότι το PFA μοντέλο παράγει logit αποτέλεσμα το οποίο μπορεί να μετατραπεί σε πρόβλεψη της επίδοσης της καθυστέρησης ή της διάρκειας. Αυτή η ικανότητα, «αφήνει» το μοντέλο να χρησιμοποιηθεί σαν εκπαιδευτική μηχανική, αφού μας δίνει του κόστους κάθε δράσης. Ξέροντας το παραπάνω, μας δίνει την δυνατότητα να πάρουμε αποφάσεις με περισσότερη βέλτιστη γνώση επι του θέματος.