

Τμήμα Μηχανικών Ηλεκτρονικών Υπολογιστών και Πληροφορικής

Πανεπιστήμιο Πατρών

Μέθοδοι συνάθροισης μερικών κατατάξεων και εφαρμογές

Διδακτορική Διατριβή

Γεώργιος Κριμπάς

Περίληψη

Τα τελευταία χρόνια, όλο και περισσότερες εφαρμογές στο διαδίκτυο εκμεταλλεύονται τη δύναμη που έχουν οι χρήστες (ή πράκτορες) για να αξιολογούν, να βαθμολογούν, να ψηφίζουν, να ανασκοπούν και να ταξινομούν σύνολα αντικειμένων (ή εναλλακτικών επιλογών). Φυσικά, αυτή η εκμετάλλευση χρησιμεύει ως μέθοδος παροχής καλύτερης πληροφόρησης και παροχής υπηρεσιών στο ευρύ κοινό. Θέτοντάς το απλά, οι χρήστες έχουν γνώση για διαφορετικά θέματα και πολλές εφαρμογές στο διαδίκτυο θέλουν να χρησιμοποιήσουν αυτή τη γνώση που πρέπει να συναθροιστεί. Για να γίνει αυτό, γνώστες μέθοδοι όπως ψηφοφορίες και γενικότερα η Θεωρία της Κοινωνικής Επιλογής από τη Μικροοικονομική θεωρία, μας εξοπλίζει με μεθόδους και διαδικασίες για τη συγκέντρωση και συνάθροιση τέτοιου είδους πληροφοριών.

Στην παρούσα διατριβή μελετάμε την εφαρμογή των μεθόδων συνάθροισης κατατάξεων προκειμένου να ανακτήσουμε τη θεμελιώδη αλήθεια. Χρησιμοποιούμε απλούς κανόνες ψηφοφορίας που είναι ευρέως γνωστοί στη Θεωρία της Κοινωνικής Επιλογής, οι οποίοι μας βοηθούν να συναθροίσουμε μικρές σε μέγεθος κατατάξεις, μερικές ή ελλιπείς, σε μια παγκόσμια κατάταξη. Μια άμεση εφαρμογή έχει αντίκρισμα στα μαζικά ανοικτά διαδικτυακά μαθήματα (γνωστά και ως MOOCs) και ιδιαίτερα στην αξιολόγηση μεταξύ ομοτίμων (peer grading), όπου αυτού του είδους οι πλατφόρμες χρησιμοποιούνται εκτενώς από φοιτητές σε όλο τον κόσμο και στις περισσότερες περιπτώσεις η βαθμολόγηση (αξιολόγηση των γραπτών) γίνεται από τους ίδιους τους συμμετέχοντες που είναι εγγεγραμμένοι σε ένα μάθημα. Η ιδέα της αξιολόγησης μεταξύ ομοτίμων είναι το εξής, δίνεται σε κάθε φοιτητή ένας μικρός αριθμός απο γραπτά (άλλων φοιτητών) ώστε να τα αξιολογήσει με τον περιορισμό ότι κανένας φοιτητής δεν θα αξιολογήσει το δικό του γραπτό, ταυτόχρονα ένας ακόμα περιορισμός είναι ότι οι φοιτητές δεν χρησιμοποιούν σκορ για την αξιολόγηση των γραπτών, αλλά κατάσσουν τα γραπτά από το καλύτερο στο χειρότερο. Στη συνέχεια, εφαρμόζουμε κανόνες ψηφοφορίας για να συναθροίσουμε αυτές

τις κατατάξεις προκειμένου να καταλήξουμε σε μία συνολική κατάταξη των γραπτών. Στόχος μας είναι αυτή η συνολική κατάταξη να είναι όσο το δυνατόν πιο κοντά στη θεμελιώδη αλήθεια. Αποδεικνύουμε ότι ένας πολύ απλός κανόνας ψηφοφορίας, συγκεκριμένα ο κανόνας του Borda, μπορεί να ανακτήσει $1 - \mathcal{O}(\frac{1}{\sqrt{k}})$ μέρος της αληθινής κατάταξης (δηλαδή της θεμελιώδους αλήθειας), όπου k είναι ο αριθμός των γραπτών που κάθε φοιτητής πρέπει να αξιολογήσει και να μας επιστρέψει με τη μορφή κατάταξης.

Επίσης, ορίζουμε και μελετούμε μια μεγάλη κλάση κανόνων συνάθροισης, την οποία ονομάζουμε *Type Ordering Aggregation Rules* (Κανόνες Συνάθροισης με τη μορφή Ταξινόμησης Τύπων). Παρουσιάζουμε ένα θεωρητικό πλαίσιο για την αξιολόγηση της απόδοσης κάθε μέλους αυτής της κλάσης. Καταλήγουμε σε στατιστικές πληροφορίες σχετικά με τη αξιολογητική συμπεριφορά των μαθητών και στη συνέχεια το θεωρητικό μας πλαίσιο μπορεί να χρησιμεύσει ως εργαλείο βελτιστοποίησης *εργαλείο βελτιστοποίησης* για την επιλογή του βέλτιστου Κανόνα Συνάθροισης με τη μορφή Ταξινόμησης Τύπων. Αυτό απαιτεί μια ακριβής λύση σε ένα στιγμιότυπο του προβλήματος *feedback arc set*, το οποίο, αν και γενικά είναι ένα NP-δύσκολο πρόβλημα, μπορεί να λυθεί ακριβώς για τις περιπτώσεις που προκύπτουν στη μελέτη μας. Το θεωρητικό πλαίσιο μας επιτρέπει να επιτύχουμε μια σειρά αποτελεσμάτων. Αποδεικνύουμε ότι ο κανόνας του κανόνα Borda είναι ο βέλτιστος κανόνας όταν οι φοιτητές δρουν ως τέλει αξιολογητές και η απόδοσή του είναι πάντα πολύ κοντά στη βέλτιστη. Επιπλέον, η απόφαση για το βέλτιστο κανόνα συνάθροισης εξαρτάται σε μεγάλο βαθμό από τις πληροφορίες σχετικά με τη αξιολογητική συμπεριφορά των σπουδαστών.

Επίσης, σχεδιάζουμε και αναλύουμε αλγορίθμους για την συνάθροιση ελλιπών κατατάξεων που βρίσκουν χρησιμότητα σε εφαρμογές αξιολόγησης επιχειρήσεων. Εφαρμογές αυτού του είδους μπορούν να βρεθούν σε πλατφόρμες και συστήματα συστάσεων. Εδώ, κάθε χρήστης μας “προμηθεύει” με μία ελλιπή κατάταξη εστιατορίων, ξενοδοχείων, ταινιών και πολλών άλλων που έχει χρησιμοποιήσει ως προϊόν. Σχεδιάζουμε αλγορίθμους οι οποίοι βρίσκουν τον καλύτερο θεσιακό κανόνα βαθμολόγησης και ως στόχο έχουμε να συναθροίσουμε τις επιμέρους ελλιπείς κατατάξεις. Η υπόθεση που κάνουμε είναι η εξής, υποθέτουμε ότι κάποια μερική γνώση των σωστών συγκρίσεων εκ των προτέρων, τις οποίες αυτές συγκρίσεις τις χρησιμοποιούμε ως περιορισμούς. Δεδομένου αυτών των περιορισμών και των επιμέρους ελλιπών κατατάξεων, το πρόβλημα βελτιστοποίησης είναι να βρούμε τον καλύτερο θεσιακό κανόνα βαθμολόγησης, το οποίο είναι NP-δύσκολο. Σχεδιάζουμε αλγόριθμο ο οποίος είναι ακριβής και λύνει το πρό-

βλημα βελτιστοποίησης σε χρόνο ο οποίος εξαρτάται εκθετικά μόνο από μία παράμετρο d , όπου d είναι ο αριθμός των εναλλακτικών επιλογών όπου κατατάσσει κάθε χρήστης. Επομένως, ο αλγόριθμός μας τρέχει σε χρόνο πολυωνυμικό όταν το d είναι μία σταθερά. Έπειτα θέλουμε να λύσουμε το πρόβλημα βελτιστοποίησης προσεγγιστικά και αποδεικνύουμε ότι ένα απλός κανόνας βαθμολόγησης όπως ο κανόνας t -approval αποδίδει μία $1/d$ -προσεγγιστική λύση. Έπειτα, σχεδιάζουμε έναν πιο εκλεπτυσμένο προσεγγιστικό αλγόριθμο ο οποίος παραμετροποιείται από ένα θετικό ακέραιο k και αποδίδει μία $\lceil d/k \rceil^{-1}$ -προσεγγιστική λύση. Τέλος, αποδεικνύουμε ότι το πρόβλημα βελτιστοποίησης που μελετάμε είναι δύσκολο να προσεγγιστεί καλύτερα από $23/24$ και παρουσιάζουμε αυτό το φράγμα μη-προσεγγισιμότητας.

Συνολικά, αυτή η διατριβή δεν βασίζεται μόνο σε θεωρητικά αποτελέσματα. Η πρόθεσή μας είναι να εξετάσουμε την απλότητα και την απόδοση αυτών των θεωρητικών αποτελεσμάτων. Έτσι, συμπληρώνουμε τα θεωρητικά μας αποτελέσματα με πειραματικά αποτελέσματα και προσομοιώσεις. Πρώτον, τα εργαστηριακά πειράματά μας εξαρτώνται σε τεχνητά δεδομένα τα οποία δημιουργήσαμε και ως στόχο είχαμε να εξετάσουμε την απόδοση των αλγόριθμών μας. Δεύτερον, σχεδιάσαμε δύο πραγματικά πειράματα (ή πειράματα πεδίου), από τα οποία καταφέραμε να εξάγουμε συγκεκριμένες πληροφορίες από τους συμμετέχοντες με στόχο να τις χρησιμοποιήσουμε για εκτεταμένες προσομοιώσεις. Καταφέραμε έτσι να επιβεβαιώσουμε τα θεωρητικά μας αποτελέσματα. Η διαφορά μεταξύ των εργαστηριακών πειραμάτων και των προσομοιώσεων μέσω των πραγματικών πειραμάτων θα γίνει εμφανής στα επόμενα κεφάλαια.

Department of Computer Engineering and Informatics

University of Patras

Rank aggregation of partial rankings: algorithms and applications

Doctoral Thesis

George Krimpas

Abstract

In the last few years, more and more applications on the internet exploit the power that users (or agents) have, in order to rate, grade, vote, review, and rank sets of services, local businesses, software, etc. (or, more generally, alternatives). Of course, the aim of these applications is to provide better information, and services to the general public. Put it simply, users have knowledge upon different aspects, and internet applications want to use this knowledge which needs to be aggregated. Voting and in general Social Choice Theory from Microeconomics equips us with methods and procedures in order to aggregate this information.

In this thesis we study the application of aggregation methods upon rankings in order to recover a ground truth. We make use variations of *simple* voting rules that are very well-known in Social Choice Theory, which help us aggregate small in size partial or incomplete rankings, into a global one. An immediate application comes from peer grading in Massive Open Online Courses (MOOCs). These platforms are used extensively from students all around the world, and in most cases the grading is performed by the students that are enrolled in a course themselves. The idea behind peer grading in MOOCs is that every student gets a small in size bundle of exam papers (of other students) to grade with the restriction that no student will get its exam paper to grade. We are particularly interested in grading where students do not use numerical scores, but they provide a ranking of the exam papers in their bundle from the best to the worst. Then voting rules are applied so as to aggregate these rankings in order to come up with a full ranking of all exam papers. The efficiency objective requires this full ranking to be as close as possible to the ground truth. We prove that a very simple voting rule, specifically Borda count, can recover a $1 - \mathcal{O}(\frac{1}{\sqrt{k}})$ fraction of the true ranking, where k is the number of exam papers that every student has to rank.

We also define and study a large class of *simple* aggregation rules, which we call *Type-*

Ordering Aggregation Rules. We present a theoretical framework for assessing the performance of each member of this class. We infer statistical information about the grading behaviour of students and then our framework can serve as an *optimization toolkit* for selecting the optimal Type-Ordering Aggregation Rule. This requires an exact solution to an instance of the feedback arc set problem which, albeit NP-hard in general, can be solved exactly for the instances that do arise in practice. Our theoretical framework allows us to obtain a series of results. We prove that Borda rule is the optimal Type-Ordering Aggregation Rule when students act as perfect graders, and its performance is always extremely close to optimality. Furthermore, the decision of the optimal aggregation rule strongly depends on the information about the grading behaviour of the students.

Finally, we design and analyse algorithms for the aggregation of incomplete rankings that are useful in rating applications. Applications of this kind can be found in recommendation systems and platforms. Here, every user can provide us with an incomplete ranking of the restaurants, hotels, or movies that she has used as a product. Then we design algorithms to find the best scoring rule in order to aggregate the individual rankings. The assumption here is that we have some knowledge about the correct comparisons a priori, which are used as constraints. Given these constraints and the individual rankings, the optimisation problem is to find the best positional scoring rule (*OptPSR*) which is NP-hard.

We design an algorithm which is exact and solves *OptPSR* in time that depends exponentially only on the parameter d , where d is the number of alternatives that each user ranks. Hence, our algorithm runs in polynomial time when d is constant. Then we seek to approximately solve *OptPSR*, and we prove that a simple combination of t -approval voting rules yields a $1/d$ -approximate solution. Then, we design a more sophisticated approximation algorithm ApxPSR_k , which is parameterized by a positive integer k and yields a $(\epsilon d)^{-1}$ -approximate solution. Last, we prove that *OptPSR* is hard to approximate and present an explicit inapproximability bound of $23/24$.

In addition to theoretical results, the thesis contains simulation and experiments that aim to test the simplicity and the performance of our algorithms in practical scenarios. We use synthetic data that we have generated, and more importantly, we have designed real-world experiments (or field experiments), in which we managed to extract specific information from real users in order to create datasets that we use in extensive simulations to verify or complement

our theoretical results.