Acta Morphologica Generalis
AMG Vol.5 No.1 (2016)
ISSN 2001-2241

EEZY: A Gaokao Recommendation System Using General Morphological Analysis over Big Data

Shan Lu, Hao Zhang, Jie Wang*

Department of Computer Science, University of Massachusetts

Abstract: College admission criteria in mainland China depend, to a large extent, on a single test score on the annual national college entrance examination called Gaokao (literally "High Exams" in Chinese). To obtain admission to a college, each student participates in Gaokao and submits a common application to their provincial Gaokao office, listing a small fixed number of universities and majors they intend to study. Each province adopts one of the following three admission models: parallel, gradient, and hybrid. No matter what models are used, it is always possible that a qualified applicant would end up being rejected. We present an automated recommendation system, called EEZY, using general morphological analysis (GMA) on large volume of data in previous years, to help students make informed decisions for achieving the best match of admissions (EEZY stands for "Yi Yi Zhi Yuan" in Chinese, meaning "personalized selections"). We then present a number of case studies to show how EEZY works and analyze why the recommendations made by EEZY make sense.

Keywords: General Morphological Analysis, Cross-Consistency Assessment, Gaokao, Recommendation System

1. Introduction

Gaokao is the mandatory annual national college entrance examination in mainland China for admission into all four-year universities. Applicants must choose to take one of the two types of exams, one called the "Li-Ke" exam (science exam) and the other called the "Wen-Ke" exam (liberal-arts exam). There are about 2,400 universities and colleges in mainland China, which are officially categorized into three tiers based on the programs they offer and the qualities of their programs. Each university sets an admission quota for each province each year, breaking down into majors. To avoid admitting a student by multiple universities, each province sets its own rules how universities access applications. These rules may be grouped into three admission models: parallel admission, gradient admission, and hybrid admission. We will explain these models in the next section.

Under any of these models, each student will only be accepted by one university, or, alternatively, not accepted at all. "Admission mismatch" thus becomes a common problem. If applicants apply to universities inappropriately, they may end up receiving no offer or an offer that is a poor match of their abilities or interests. Both Gaokao and college admission are conducted once a year. Once a student is admitted by a university to a particular major, it is almost impossible to change majors after admission. Thus, selecting the best-matched university and a major is critical in the admission process. We present an automated system called EEZY using the General Morphological Analysis (GMA) method to analyze large volumes of data which we collected from previous years of Gaokao and help students make informed decisions based on their Gaokao scores and interests and a baseline recommendation index that measures each EEZY recommendation to each student.

* Corresponding author: wang@cs.uml.edu

GMA is a method for identifying and investigating the total set of configurations contained in multi-dimensional, non-quantifiable problem complexes, attempting to derive all the solutions of any given problem in an unbiased manner. While we could use any quantifiable method (such as operation research and statistics) in any way we can, using GMA may help us discover new relationships or configurations that may not be so evident or might have been overlooked by other methods. To the best of our knowledge there has not been any work done on Gaokao recommendation system using the GMA methodology.

This paper is organized as follows. In Section 2 we describe in detail the three admission models currently used in college admission in mainland China. In Section 3 we present GMA analysis to generate recommendations of universities and majors. In Section 4 we describe two case studies and conclude the paper in Section 5.

2. Admission Models

In mainland China, students do not submit their admission applications directly to universities. Instead, they submit them to their provincial Gaokao office. A university does not get to evaluate all the applications that have it listed. Exactly how a university gets applications depends on the admission model adopted by that province.

Let u denote a university, m a major that u offers, and p a province. Let $Q(u,p)$ denote u's admission quota in province p. In what follows we will omit p with the understanding that we are dealing with a particular province. Thus, we will use $Q(u)$ to denote u's admission quota (in province p). Let $Q(u,m)$ denote u's admission quota for major m. Let $M(u)$ denote the list of majors that u makes available to students. We have $Q(u) = \sum_{m \in M(u)} Q(u,m)$.

Under all admission models, students are allowed to specify a number of universities and majors in an application form in three sections, one for tier-1 universities and majors, one for tier-2 universities and majors, and one for tier-3 universities and majors. Under each section, students are allowed to enter a small fixed number of universities in empty slots, typically four universities labeled as A, B, C, D in the order of preference; and a fixed number of majors for each university, typically six majors for each university, labeled as 1, 2, 3, 4, 5, 6 in the order of preference. For convenience, we will call the university listed in the A-slot in an application the A-university for that application. Similarly, we can define B-university, C-university, and D-university.

2.1. Parallel admission

The parallel admission model follows the principle of "scores first, according to preference". It proceeds as follows: The provincial Gaokao office releases applications to universities in the descending order on Gaokao scores of the applicants, indexed from 1 to N, where N is the total number of applications in the underline province. The value of N is typically in the range of 200,000 to 400,000 for a given province.

Let $N(u,i)$ denote the number of applications that have been released to u before the i-th application, and $N(u,m,i)$ the number of applications that have been released to u for major m before the i-th application. Suppose that u is listed in the i-th application. We say that u is available if the following two conditions are satisfied:

1. $N(u,i) < (1+d_u)Q(u)$, where d_u is a small positive fractional number.
2. $N(u,m,i) < (1+d_{u,m})Q(u,m)$, for some m, where $d_{u,m}$ is a small positive fractional number. The major m here is called an available major.

The first condition indicates that the quota for admission by u has not been met, and the second indicates that the quota for majors by u has not been met. While u will in general accept applications released to it, it might still ask for a slightly larger number of applications than its admission quota to have some flexibility for selecting students.

Initially, set $N(u,1) = 0$ and $N(u,m,1) = 0$ for all university u and all majors m. For each application i in the list, where i starts from 1 to N, the Gaokao office repeats the following procedure: Start from the A-university down to the D-university, find the first university u that is available, and release the application i to u (where u will consider the application for the first available major it lists).

2.2. Gradient admission

The gradient admission model follows the principle of "preference first, according to scores", which proceeds in a number of rounds. The provincial Gaokao office first groups, for each university u, all the applications listing u as A-university into one group, and we call it the A-group for u. The applications in the A-group for each university are further divided into subgroups according to the first major m listed. For each subgroup of applications, the applications are sorted in descending order of Gaokao scores. The Gaokao office releases to u the top $(1+d_{u,m})Q(u,m)$ applications for each subgroup with major m, where $d_{u,m}$ is a small positive fractional number.

After the first round, if there are applications in the A-group not released and the university u still has not met its admission quota, then the remaining applications are grouped according to the next major. The procedure continues until either university u has filled up its admission quota, or there are no universities left in the A-group for all the applications.

If there are universities that have not fulfilled their admission quotas in this round, then the Gaokao office waits for all applications to have been looked at by their A- universities, and then repeats the same procedure for the B-universities, and in the same way, the C-universities, and finally the D-universities.

2.3. Hybrid admission

Some provinces follow a hybrid admission model. For example, Qinghai in 2015 followed the gradient admission model for the first- and second-tier universities, and the parallel admission model for the third-tier universities. Chongqing in 2014 followed the parallel admission model for the first-tier universities and the gradient admission model for the second- and third-tier universities, where students need to select three universities for the first preference and three universities for the second preference, but within the three universities in the same preference group, it followed the parallel admission model.

3. Finding the Best Suited Universities and Majors Using GMA

Using computer assisted GMA, we can compute all possible combinations of majors and universities for students according to their Gaokao scores and their interests under the admission model in their province, with the requirements that these recommendations are "well suited" according to a baseline recommendation index we will define in Section 3.2. We then partition the combinations of majors and universities into a number of recommendation categories. The number of recommendation categories for different provinces may be different.

Let LU denote the set of labels for university slots in an application form. Let J the number of universities a student is allowed to specify for each tier, and K the number of majors a student is allowed to specify for a university. For example, it is typical that $J = 4$ and $K = 6$.

For convenience, in what follows we will use Alice to represent a student and XYU to represent a university.

.

Alice enters her Gaokao score and other information into EEZY to obtain recommendations for X-universities, where $X \in LU$, and we call them X-recommendations. For example, when J = 4, we have A-, B-, C-, and D-recommendations, respectively. Universities listed in A-recommendations are competitive for Alice, but Alice still has a chance to be accepted. Universities listed in B-recommendations would present a good match of Alice's ability and interests, which means that Alice would have a good chance to be accepted. Universities listed in C-recommendations are conservative choices for Alice, which means that Alice would have a very good chance to be accepted. Universities listed in D-recommendations are the safest choices for Alice, which means that Alice would have a near 100% chance to be accepted.

3.1. Identify and define the most important parameters of the problem

The conceptual framework of EEZY consists of a so-called GMA *overlay model* consisting of a number of common parameters by which we can pit *scores and choices of students* against *attributes and requirements of universities*. This overlay creates an inference model which allows us to see how particular student choices are compatible with university attributes and requirements.

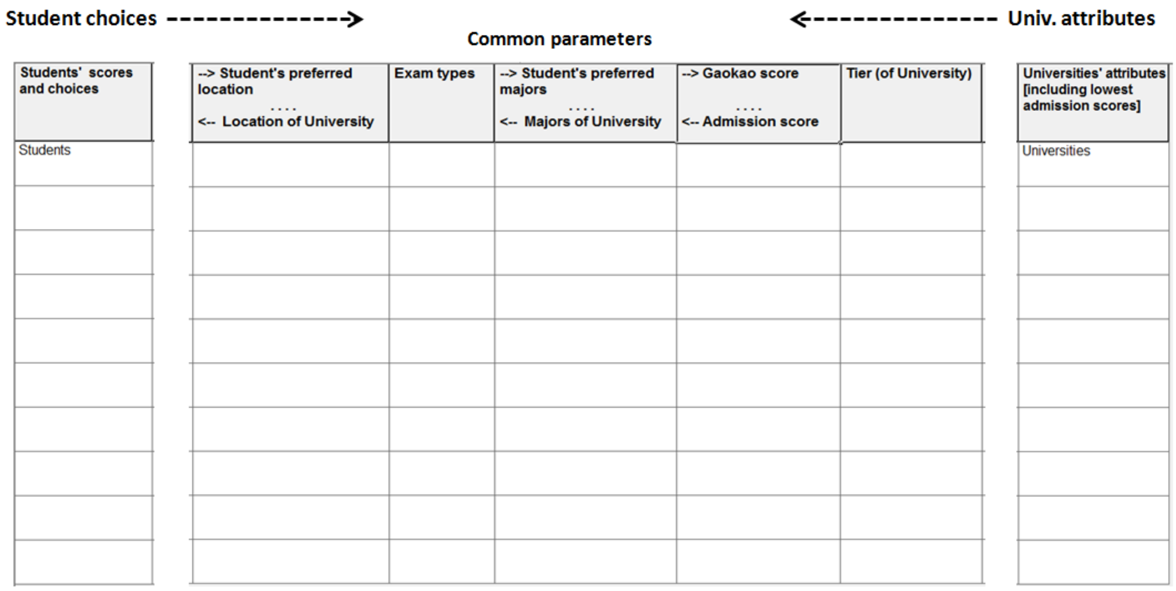

Figure 1: Example of overlay model for student choices vs. university attributes.

In practice, the results for a particular student case are given in tabular form, with 7 parameters each for the *student choices* and the *university attributes* respectively (see below). There are other parameters we would want to consider, such as a university's appropriateness to the student, a student's talents, parents' occupations, annual incomes, rankings of a university's majors, admission quota for each major, immediate employment record for a university, and the cost to attend a university. We present only baseline parameters here for simplicity, and deal with universities' appropriateness to the student in the next section.

The parameters in the student group are:

1. **Gaokao score**. This is Alice's Gaokao score.
2. **Exam type**. This is the type of the exam that Alice takes, which is either Li-Ke or Wen-Ke.
3. **Tier**. This is the tier of the universities that Alice wants to attend.
4. **Preferred location**. This is a list of locations where Alice wants to go to for college.
5. **Disliked location**. This is a list of locations that Alice does not want to go to for college.
6. **Preferred major**. This is a list of majors that Alice wants to study.
7. **Disliked major**. This is a list of majors that Alice does not want to study.

Gaokao Score	Exam Type	Tier	Preferred Location	Disliked Location	Preferred Major	Disliked Major
600	Sciences	First	Liaoning Jilin Hunan Beijing	Sichuan Yunnan	Bioengineering Science Economics	Philosophy Electric Information Material Science

Figure 2: Alice's parameters

Figure 2 shows the data entered by Alice.

The parameters in the university group are:

1. **Admission score**. This is the lowest admission score of XYU in the past year.
2. **Major type**. This is the type of majors that XYU offers.
3. **Tier**. This is XYU's official tier.
4. **Location**. This is XYU's location.
5. **Majors**. This is the list of majors that XYU offers, including (when possible) the lowest, medium, and highest admission scores for each major in the past year, and the total number of expected enrollment for a major for the current year.
6. **Ranking**. This is the ranking of XYU. The first-tier universities are ranked from 1 to 5 with 1 being the highest according to which groups they are in. Likewise, the second-tier universities are ranked from 6 to 7. The third-tier universities have one rank of 8.
7. **Enrollment**. This is the total enrollment of XYU for the current year. (When this number is not known yet, it uses last year's enrollment number.)

Figure 3 shows a few universities and majors that match Alice's Gaokao score and interests, where CSU stands for Central South University, FZU for Fuzhou University, SCU for Sichuan University, HNU for Hunan University, HHU for Hohai University, CAU for China Agricultural University, NEU for Northeastern University, and SWUFE for Southwestern University of Finance and Economics.

Admission Score	Major Type	Tier	Location	Ranking	Enrollment	Major
CSU (594)	Sciences	First	Hunan	3	138	Automotive Engineering Material Chemistry Statistics Applied Chemistry
FZU (571)	Sciences	First	Fujian	4	1686	Electric Information Accounting Automation Finance
SCU (592)	Sciences	First	Sichuan	3	72	Chemistry Software Engineering Computer Science Material Science
HNU (591)	Sciences	First	Hunan	3	54	Automation Chemistry Environmental Sciences Automotive Engineering
HHU (572)	Sciences	First	Jiangsu	4	93	Accounting Computer Science Economy and Trade Automation
CAU (588)	Sciences	First	Beijing	3	51	Landscape Architecture Animal Science Computer Science Biological Sciences
NEU (572)	Sciences	First	Liaoning	3	75	Finance Economics Software Engineering Bioengineering
SWUFE (586)	Sciences	First	Sichuan	4	26	Statistics Computer Science Logistics Management Law

Figure 3: University's parameters. Possible combinations of majors and universities for Alice are marked, respectively, with red, orange, blue, and green for A-, B-, C-, and D-recommendation. Abbreviations on the leftmost column are the names of universities, and the number inside the parentheses by the university abbreviation is the average admission score of that university.

3.1.1 Collecting data

We collected a large volume of data in the previous years of Gaokao data for all provinces in the following three formats: (1) Excel files; (2) PDF files (we were able to obtain PDF files from a small number of provinces); and (3) printed books (we were able to obtain printed books for most provinces). We scanned printed books into PDF files, and used a procedure we developed using an off-the-shell OCR software to convert these files into editable excel files, and checked manually for correctness.

3.2 Baseline recommendations

We now setup baseline recommendations based on baseline parameters to form a solution space. For each university u in a possible solution, let H_u and L_u denote, respectively, the highest and lowest admission scores by u in the previous year. We then divide $D = H_u - L_u$ into $J = |U|$ intervals as follows:

$$[L_u - d, l_1), [l_1, l_2), \cdots, [l_{J-2}, l_{J-1}), [l_{J-1}, H_u + d),$$

where d is a small positive integer (e.g., we may select d = 3) and $l_i = L_u + iD/J$, i = 1, · · ·, J − 1.

- If Alice's Gaokao score is in the range of $[L_u - d, l_1)$, then u is among the most highly competitive university for Alice, yet still has a chance to be accepted. The chance of admission to an A-university is small.
- If Alice's Gaokao score is in the range of $[l_{J-1}, H_u + d)$ then u is among the safest schools for Alice and the chance of admission is close to 100%. On the other hand, Alice may still have comparable peers with similar Gaokao scores in her class (otherwise, u would be a poor choice for Alice).
- The meanings of Alice's Gaokao score falling in one of the middle intervals can be similarly determined.

For example, when J = 4, we have

$$A = [L_u - d, l_1),\ B = [l_1, l_2),\ C = [l_2, l_3),\ D = [l_3, H_u + d)$$

If Alice's Gaokao score is in the interval A and u offers some of Alice's preferred majors, then u will be included in the A-recommendations for Alice. The rest of the recommendations are similarly defined.

To determine which recommendation is better for Alice, we define an individual Base line Recommendation Index (BRI) for each university recommended to Alice. The universities with BRI under 50% are considered unsuitable to Alice. Let $X(s)$ denote the X-recommendation for s, where $X \in U$.

3.2.1 University groupings

We first characterize the first-tier universities into the following five groups:

1. Group G_1 consists of the two super universities: Peking University and Tsinghua University. They are the best funded and most reputable universities in China. The Chinese government designates both universities as project-985 universities. There are 39 universities in mainland China with this designation, which are the national key universities.
2. Group G_2 consists of the top ten universities after Peking and Tsinghua. They are also project-985 universities. Members of G_2 may change from year to year.
3. Group G_3 consists of all the remaining 27 project-985 universities.

4. Group G_4 consists of all the officially designated project-211 universities, excluding project-985 universities. These are universities having top programs in certain areas. There are 73 universities in G_4.
5. Group G_5 consists of the remaining first-tier universities.

The second-tier universities can also be further characterized into the following two groups:

1. Group G_6 consists of provincial key universities.
2. Group G_7 consists of the remaining universities in this tier.

Finally, group G_8 consists of all the universities in the third tier.

3.2.2 Measure of ranking

Denote by $I_X(s,u)$ the BRI for university u contained in the X-recommendation for student s. We will want $I_X(s,u)$ to reflect an appropriate value, which depends on the following three factors:

1. The group ranking of the universities in $X(s)$;
2. The matching of majors and the number of majors offered by u that are similar to a major selected by s;
3. The admission quota of u.

These three factors should carry about the same weight. Since ranking of a university is in general more important to many applicants, we will want the university ranking to carry slightly more weight.

We first devise a mechanism to reflect the ranking of a university in a recommendation. Let $r(u)$ denote the ranking of u, where $r(u) \in \{1,2,\cdots,8\}$. Let $X \subseteq U$. We say that $X(s)$ is in case l for some k, where $l = 1,2,3,4$ and $k \in [1, 9-l]$, if $X(s)$ contains universities only in G_{k+j} for each $j = 0,\cdots l-1$; that is, if for each $j = 0,\cdots l-1$ we have

$$X(s) \cap G_{k+j} \neq \emptyset$$

$$X(s) \cap \left(\bigcup_{j=0}^{l-1} G_{k+j} \right) = X(s)$$

Recall that U is the set of labels for university slots and $J = |U|$.

Define a ranking measure for u within the X-recommendation, denoted by $R_X(u)$, as follows:

$$R_X(u) = \begin{cases} 50, & \text{if } X(s) \text{ is in case } 1, \\ \left(5 - \frac{5}{2}\big(r(u) - k\big)\right) \cdot 10, & \text{if } X(s) \text{ is in case 2 for some } k, \\ \left(5 - \frac{3}{2}\big(r(u) - k\big)\right) \cdot 10, & \text{if } X(s) \text{ is in case 3 for some } k, \\ \big(5 - \big(r(u) - k\big)\big) \cdot 10, & \text{if } X(s) \text{ is in case 4 for some } k. \end{cases}$$

3.2.3. Measure of major matching

Next, we devise a mechanism for measuring matching of majors. In mainland China, areas of studies are officially classified into a hierarchy of three classes. The Class-1 category consists of eleven general areas of studies:

(1) Philosophy, (2) Economics, (3) Law, (4) Education, (5) Literature, (6) History, (7) Science, (8) Engineering, (9) Agriculture, (10) Medicine, (11) Management.

Each area in Class 1 (referred to as class-1 subject) often consists of a number of subjects referred to as class-2 subjects. For example, Science is a Class-1 subject, which consists of 12 Class-2 subjects:

(1) Math, (2) Physics, (3) Chemistry, (4) Astronomy, (5) Geographical Sciences, (6) Atmospheric Sciences, (7) Ocean Sciences, (8) Geophysics, (9) Geology, (10) Biological Sciences, (11) Psychology, (12) Statistics.

Each Class-2 subject further consists of a few subdisciplines referred to as Class-3 subjects. For example, Math is a Class-2 subject, which consists of two Class-3 subjects: 1. Mathematics and Applied Mathematics, 2. Information and Computing Science. Each subject in any class is uniquely identified by a subject code.

Alice is allowed to select six subjects in her application as her preferred list of majors. For each slot m_i of majors, she must specify a class-3 subject, or leave it unspecified (that is, she needs to check the box on her application that she is willing to major in any subject).
Recall that K denotes the number of majors s is allowed to specify for u. Let $m_i(s,u)$ denote the matching score for the i-th major that student s specifies for a university u, $i = 1, 2, \cdots, K$.

EEZY allows students to specify majors at the Class-1 level, Class-2 level (after Class 1 specified), or the Class-3 level (after Class 2 is specified). Let (a,b,c) denote a specification of major, where a is a subject in Class 1 (which could be empty), b a subject in Class 2 (which could be empty), and c a subject in Class 3 (which could be empty). Note that if a is empty, then b and c must be empty. Likewise, if b is empty then c must be empty. Given a major specification (a,b,c) entered by s, we define the following terms:

1. We say that a match occurs at level 3 for student s with university u if one of the following conditions are satisfied:
   (a) The university u offers c.
   (b) The university u offers b, and c is empty (in this case, any Class-3 subject offered by u under b is deemed specified by s).
   (c) The university u offers a, and b is empty (in this case, any Class-3 subject offered by u under a is deemed specified by s).
   (d) The specification (a,b,c) is empty (in this case, any Class-3 subject offered by u is deemed specified by s).

2. We say that a match occurs at level 2 for student s with university u, if b is offered by u, but c (not empty) is not offered by u.

3. We say that a match occurs at level 1 for student s with university u, if a is offered by u, but b (not empty) is not offered by u.

In addition to matching of majors, we would also like to put more weight on university u if it offers more majors under a given Class-2 subject, for it provides more related disciplines of studies for student s. For a particular Class-2 subject b, let n_b denote the number of Class-3 majors u offers under b.

Given a major specification (a_i, b_i, c_i) for the i-th slot, where $i = 1, 2, \cdots, K$, we define $m_i(s,u)$ by the following formula:

$$m_i(s,u)=\begin{cases}\left(43+\min\left\{n_{b_i},5\right\}\right)/K, & \text{if a match ocurrs at level 3,}\\ \left(24+\min\left\{n_{b_i},5\right\}\right)/K, & \text{if a match ocurrs at level 2,}\\ \left(12+\min\left\{n_{b_i},5\right\}\right)/K, & \text{if a match ocurrs at level 1.}\end{cases}$$

Let

$$m(s,u)=\sum_{i=1}^{K} m_i(s,u).$$

We note that the maximum value of $m(s,u)$ is 48.

3.2.4. Measure of admission quota

Finally, we define a mechanism to credit admission quota. Let G_k denote universities in group k, where $k=1,2,\cdots,8$. Recall that Q(u) denotes the admission quota of university u for the current year. Let

$$Q_k=\max_u\left\{Q(u)\middle|u\in G_k\right\}.$$

That is, Q_k is the largest admission quota of the universities in G_k for the current year. We note that $Q(u)$ could range from dozens to thousands, where the top-tier national universities tend to have a smaller admission quota for a particular province. Large enrollment typically happens at low-tier universities or at universities that are located in the same province. Thus, enrollment larger than a certain number (e.g, 2,500) will no longer be significant for influencing the BRI.

Define:

$$e(x)=\begin{cases}\lfloor x/2\rfloor, & \text{if } 1\le x\le 20,\\ 5(i-1)+\lfloor x/2^i\rfloor, & \text{if } 2^{i-1}+1\le x\le 2^i\times 10, \text{where } i=2,3,\cdots,8,\\ 45, & \text{if } x>2{,}560.\end{cases}$$

We will use $q(u)=e(Q(u))$ to measure enrollment credit for university u. Note that universities included in $X(s)$ may either include universities of the same rank or universities of consecutive ranks. Because universities with ranks from 1 to 4 could have close admission scores for certain majors, it is possible that $R_A(s)$ may include universities with ranks from 1 to 4.

3.2.5. The baseline recommendation index

We now define $I_X(s,u)$ as follows, where $X\in U$

$$I_X(s,u)=\begin{cases}\dfrac{R_X(s,u)+m(s,u)+q(u)}{50+e(Q_k)+48}\cdot 100\%, & \text{if X(s) is in case 1 for some k,}\\ \dfrac{R_X(s,u)+m(s,u)+q(u)}{50+\max\left\{e(Q_{k+j})\middle|j=0,1\right\}+48}\cdot 100\%, & \text{if X(s) is in case 2 for some k,}\\ \dfrac{R_X(s,u)+m(s,u)+q(u)}{50+\max\left\{e(Q_{k+j})\middle|j=0,1,2\right\}+48}\cdot 100\%, & \text{if X(s) is in case 3 for some k,}\\ \dfrac{R_X(s,u)+m(s,u)+q(u)}{50+\max\left\{e(Q_{k+j})\middle|j=0,1,2,3\right\}+48}\cdot 100\%, & \text{if X(s) is in case 4 for some k,}\end{cases}$$

4. Case Studies

We present two case studies, where students Bob and Jill are residents in the Fujian province. We will analyze the B-recommendations for Bob, the C-recommendations for Jill, and justify that EEZY-recommendations make sense. We use data in the year of 2013 for demonstration.

4.1 Case A

Bob obtains a Gaokao score of 530 in the Wen-Ke exam (i.e., the liberal-arts exam). He is interested in pursuing a major in Human Resource Management, and specifies the following chain of majors in EEZY:

(Management, Business Administration, Human Resource Management).

Recall that (a,b,c) represents a major chain, where a is a Class-1 subject, b (could be empty) is a Class-2 subject under a, and c (could be empty) is a Class-3 subject under b. We will use asterisk" *" to denote an empty subject (i.e., a wildcard). Note that if b is empty then c must be empty. An empty subject means any subject at that level. Bob also wants to major in Law (class 1) with any class-3 subject, and so he specifies (Law, *, *) as his second favorite subject.

Bob, however, does not want to major in Economics (class 1). Thus, he checks (Economics, *, *) as his disliked major.

The B-recommendations returned by EEZY for Bob is shown with English translation in Table 1.

We will use the following notations in all the tables below, where "AN" stands for" Admitted Number", "AS" for "Average Score", "HS" for "Highest Score", and "LS" for "Lowest Score".

Table 1: B-recommendations for Bob with his favored Class-3 subject of Human Resource Management and Class-1 subject of Law; and a disliked Class-1 subject of Economics

University	BRI	Major	AN	AS	HS	LS
Shanghai Normal University	83%	Tourism Management	5	524	532	520
		Human Resource Management	4	525	527	522
		Law	5	525	532	520
		Public Administration	6	519	520	517
		Philosophy	5	516	517	514
		Television Broadcasting Science	5	521	528	517
		Advertising	5	520	523	518
Total			35			
Tianjin Foreign Studies University	61%	Advertising	2	525	525	524
		International Politics	1	520	520	520
		Law (attorney)	2	526	527	525
		Marketing	1	522	522	522
		Indonesian	1	530	530	530
		Translation	1	530	530	530
		Journalism (International)	2	528	529	527
Total			10			

We can see that the average admission scores for each major of the two universities in the B-recommendations are all close to, but less than or equal to Bob's Gaokao score.

Moreover, Bob's Gaokao score is higher than most of the highest admission scores with only two exceptions, where Bob's Gaokao score is two points under. Thus, B-recommendations are indeed a good fit for Bob and Bob has a solid chance to be accepted. Since Shanghai Normal University has a much larger admission quota and has a higher reputation, its BRI is also a lot higher.

4.2. Case B

Jill obtains a Gaokao score of 525 in the Wen-Ke exam. Table 2, depicts the C- recommendations returned by EEZY for Jill without specifying any preference.

Table 2: C-recommendations for Jill without specifying any preference

University	BRI	Major	AN	AS	HS	LS
Minnan Normal University	83%	Economics	73	528	536	523
		Japanese	15	519	524	515
		Television Broadcasting Science	20	524	533	520
		Editing And Publishing	40	519	535	514
		Advertising	35	519	532	515
		Tourism Management	6	523	526	521
		Labor and Social Security	12	518	520	516
		Culture Industry Management	9	522	524	520
		Human Resource Management	12	526	533	523
		Marketing	21	522	526	521
		Translation	29	518	529	513
		English	45	525	541	521
		Chinese Literature	224	517	535	507
		Law	48	523	531	519
		Politics and Administration	68	515	529	508
		Social Work	23	515	528	513
Total			680			
Hunan Agricultural University	76%	Education	2	522	522	521
		Marketing	2	522	522	522
		Investment	2	524	524	523
		English	2	524	525	523
		Japanese	2	523	524	521
		Administration Management	2	523	523	523
		Law	3	522	525	520
		Land Resources Management	2	523	523	522
Total			17			
Shenyang Jianzhu University	73%					
		Business Administration	6	526	533	515
		Law	2	523	529	516
Total			8			

Now suppose that Jill does not want to study the following majors: (Management, Business Administration, Marketing Management), (Management, Public Administration, *), (Literature, Foreign Literature, English), and (Education, *,*). Moreover, she does not want to go to schools in Fujian. Under these constraints the C-recommendations for Jill are shown in Table 3. We see that Minnan Normal University is no longer on the list because it is in Fujian. More than half of the majors in Hunan

Agricultural University are related to Jill's disliked majors and so they are also removed, reducing its BRI for Jill from 76% to 72%, which is lower than that of the Shenyang Jianzhu University.

Table 3: C-recommendations for Jill with disliked majors: (Management, Business Administration, Marketing Management), (Management, Public Administration, *), (Literature, Foreign Literature, English), and (Education, *, *); and disliked location: Fujian

University	BRI	Major	AN	AS	HS	LS
Shenyang Jianzhu University	73%					
		Business Administration	6	526	533	515
		Law	2	523	529	516
Total			8			
Hunan Agricultural University	72%	Investment	2	524	524	523
		Japanese	2	523	524	521
		Law	3	522	525	520
Total			7			

5. Conclusions

Gaokao is a major national event in mainland China, affecting about 10 million students each year. We presented an automated system called EEZY to help students using general morphological analysis over big data to make informed decisions which universities and majors to apply to so that they can achieve the best possible match of their ability and interests. Our case studies showed that the recommendations provided by EEZY make sense. Moreover, using EEZY students may try all possible combinations of their preferences easily and explore the best matches for them.

Acknowledgements

Ming Jia, Jingwen (Jessica) Wang, Liqun (Catherine) Shao, Hualiang Xu, You (Stephanie) Li, Wuping Zheng, Fanyu Kong, Zhiyong Luo, and a few other people provided assistance on various aspects in the early stage of this project. The authors thank them for their help.

Background literature

Mingers, J. & Rosenhead, J. (2004). “Problem Structuring Methods in Action”, European Journal of Operational Research, 152:530–554.

Ritchey, T. (2002). “General Morphological Analysis – A general method for non-quantified modeling”, Proceedings of the 16th Euro Conference on Operational Analysis, Brussels, July 1998.

Ritchey, T. (2006) “Problem Structuring using Computer Aided Morphological Analysis”. Journal of the Operational Research Society, 57:792-C801.

Ritchey, T. (2011). Wicked Problems/Social Messes: Decision support Modelling with Morphological Analysis. Springer, Berlin.

Thalheim, B. (2010). Towards a Theory of Conceptual Modelling. Journal of Universal Computer Science, 16(20):3102–3137.