

THEN ONLI NEARMETHOD OF PICTURES REFINEMENT

Melentyeva O.V., Abdullaev A.R.

Federal State Budgeted Education Institution for Higher Professional Education Perm National Research Polytechnic University, Perm, Russia (RF, Perm, Komsomolsky Av. 29, 614990),
e-mail: h.m@pstu.ru, lica90@mail.ru

This article deals with the problem of image processing. This paper proposes a new nonlinear method of full non-linear processing of digital color and grayscale images to improve quality. Compared with the known method of median filtering, which has a number of disadvantages, such as blurring of boundaries, loss of brightness, etc., the method suggested in this paper allows more efficiently to get rid of unwanted noise and noise components: sharpness and contrast are saved, restored fine detail. The proposed algorithm is oriented. The article is a comparative analysis of the test applications of software that implements the algorithm described above. Also proposed a new objective method of assessing the quality of digital images, based on the analysis of variance values before and after treatment of the object.

СРАВНИТЕЛЬНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ ПАРАМЕТРИЧЕСКИХ И НЕПАРАМЕТРИЧЕСКИХ МЕТОДОВ ОЦЕНИВАНИЯ ЗНАНИЙ В СИСТЕМЕ ЗАОЧНОГО ОБУЧЕНИЯ

Михальчук А.А., Арефьев В.П., Филипенко Н.М.

ФГБОУ ВПО «Национальный исследовательский Томский политехнический университет»,
Томск, Россия (634050, г. Томск, пр. Ленина, 30), e-mail: aamih@tpu.ru

Проведен сравнительный статистический анализ параметрических и непараметрических методов оценивания качества заочного технического образования на базе трех переменных: набранных баллов за 4 индивидуальных домашних задания (идз), набранных баллов за тест-экзамен в режиме online (экз) и разности моментов окончания и начала экзамена (dt). На основании критерия хи-квадрат выявлено значимое отличие распределений переменных от нормального закона: от статистически значимого для экз и до высоко значимого для dt. Корреляция между dt и экз, а также между идз и экз оценена как незначимая, то есть экзаменационный результат не зависит значимо ни от времени, затраченного на выполнение экзаменационной работы, ни от качества допуска к экзамену по идз. Корреляция между идз и dt ($r \gg R \gg 0,274$) сильно значимо отличается от 0, тем не менее весьма далека от 1. При этом коэффициенты парных корреляций Пирсона r отличаются от соответствующих ранговых корреляций Спирмена R незначимо. На основании однофакторного параметрического и непараметрического дисперсионных анализов оценена значимость неоднородности переменных dt, экз и идз по трем потокам: сильно значимая для dt, незначимая для экз и высоко значимая для идз по совокупности 3 потоков. Оба критерия также использованы при парном сравнении потоков. Аналогично оценена значимость неоднородности переменных dt, экз и идз по 10 группам. Результаты проведенного дисперсионного анализа могут быть учтены в рамках проходящей реформы высшего образования.

COMPARATIVE STATISTICAL ANALYSIS OF PARAMETRICAL AND NONPARAMETRIC METHODS OF THE ESTIMATION OF KNOWLEDGE IN CORRESPONDENCE COURSE SYSTEM

Mihalchuk A.A., Arefyev V.P., Filipenko N.M.

National research Tomsk polytechnic university, Tomsk, Russia
(634050, Tomsk, Lenin's avenue, 30), e-mail: aamih@tpu.ru

The comparative statistical analysis of parametrical and nonparametric methods of an estimation of quality of correspondence technical education on the basis of three variables is spent: the collected points for 4 individual homeworks (ihw), the collected points for test examination in a condition on-line (ex) and differences of the moments of the termination and the examination beginning (dt). On the basis of criterion a chi-square significant difference of distributions of variables from the normal law is revealed: from statistically significant for ex and to highly significant for dt. Correlation between dt and ex, and also between ihw and ex is estimated as not significant, that is the examination outcome does not depend significantly on time expended on performance of an examination-paper, from quality of tolerance to examination on ihw. Correlation between ihw and dt ($r \gg R \gg 0,274$) strongly significantly differs from 0, nevertheless is rather far from 1. Thus factors of conjugate correlations of Pirson r differ from corresponding grade correlations of Spirmen R not significantly. On the basis of one-factor parametrical and nonparametric analyses of variance the importance of heterogeneity of variables dt, ex and ihw on three streams is estimated: strongly significant for dt, insignificant for ex and highly significant for ihw on a population of 3 streams. Both criteria also are used at paired comparison of streams. The importance of heterogeneity of variables dt, ex and ihw on 10 groups is similarly estimated. Outcomes of the spent analysis of variance can be considered within the limits of passing reform of higher education.

ДИСПЕРСИОННЫЙ АНАЛИЗ РЕЗУЛЬТАТОВ ВХОДНОГО ТЕСТИРОВАНИЯ МАТЕМАТИЧЕСКИХ ЗНАНИЙ В ТЕХНИЧЕСКОМ ВУЗЕ

Михальчук А.А., Арефьев В.П., Филипенко Н.М.

ФГБОУ ВПО «Национальный исследовательский Томский политехнический университет»,
Томск, Россия (634050, Томск, пр. Ленина, 30), e-mail: aamih@tpu.ru

Проведен статистический анализ результатов входного тестирования (ВТ) по математике набора 2011-2013 гг., в котором участвовали студенты-очники первого курса семи институтов Томского политехнического

университета (ТПУ). В рамках 2-х факторной дисперсионной модели ВТ (3-х уровневый фактор «ГОД» и 7-и уровневый фактор «ИНСТИТУТ») исследовано влияние факторов «ГОД» и «ИНСТИТУТ» на ВТ. Результаты ВТ ТПУ являются высоко значимо неоднородными по институтам за счет значимого размаха средних баллов результатов ВТ институтов от 1,4 до 2,5. Для каждого года выделены однородные (различающиеся незначимо) группы институтов. Результаты ВТ математических знаний студентов ТПУ имеют положительную динамику. Различия результатов ВТ ТПУ по годам оценивается как высоко значимое ($p < 0,0005$) за счет высоко значимого отличия результатов ВТ ТПУ 2011г. (средний балл 1,724) от 2012-2013гг. при незначимом ($p > 0,10$) различии результатов ВТ ТПУ 2012г. (средний балл 2,261) и 2013г. (средний балл 2,268). Оценена динамика качества приема на 1 курс по институтам в широком диапазоне значимости: незначимая в период 2012-2013гг. для одного из семи институтов, для остальных институтов в период 2011-2012гг. значимая положительная, но в период 2012-2013гг. три института имеют положительную незначимую динамику, один имеет отрицательную незначимую динамику, а два – отрицательную статистически значимую динамику. Результаты проведенного статистического анализа могут быть учтены в рамках проходящей реформы высшего образования.

ANALYSIS OF VARIANCE OF OUTCOMES OF ENTERING TESTING OF MATHEMATICAL KNOWLEDGE IN TECHNICAL COLLEGE

Mihalchuk A.A., Arefyev V.P., Filipenko N.M.

National research Tomsk polytechnic university, Tomsk, Russia
(634050, Tomsk, Lenin's avenue, 30), e-mail: aamih @tpu.ru

The statistical analysis of outcomes of entering testing (ET) on the mathematician of a gang 2011-2013 in which students-resident of the first course of seven institutes of Tomsk polytechnic university (TPU) participated is spent. Within the limits of 2 factor dispersing models ET (3 level of the factor «YEAR» and 7 level of the factor «INSTITUTE») influence of factors «YEAR» and «INSTITUTE» on BT is investigated. Outcomes of ET TPU are highly significantly inhomogeneous on institutes at the expense of significant scope of mean scores of outcomes ET of institutes from 1,4 to 2,5. For each year are selected homogeneous (differing not significant) groups of institutes. Outcomes ET of mathematical knowledge of students TPU have positive dynamics. Distinction of outcomes of ET TPU on years is estimated as highly significant ($p < 0,0005$) at the expense of highly significant difference of outcomes of ET TPU 2011г. (mean score 1,724) from 2012-2013гг. at insignificant ($p > 0,10$) distinction of outcomes of ET TPU 2012г. (mean score 2,261) and 2013г. (mean score 2,268). Dynamics of quality of reception on 1 course on institutes in a wide range of the importance is estimated: insignificant in phase 2012-2013гг. for one of seven institutes, for remaining institutes in phase 2011-2012гг. significant positive, but in phase 2012-2013гг. three institutes have the positive not significant dynamics, one has negative not significant dynamics, and two - negative statistically significant dynamics. Outcomes of the spent statistical analysis can be considered within the limits of passing reform of higher education.

ОПРЕДЕЛЕНИЕ РЕГИОНА АВТОРА ПО ДАННЫМ ЖИВОГО ЖУРНАЛА

Морозов Е.В., Богданова Д.Н.

ООО «АбиИнфоПоиск», Москва, Россия (127273, Москва, ул. Отрадная, 2Б.6),
e-mail: eugene_m@abby.com

В настоящей работе представлен корпус записей русскоязычных блогов с информацией о местоположении автора, а также проведено исследование методов машинного обучения для автоматического определения региона автора. Для создания корпуса использовалась коллекция текстов блогерской платформы Живой Журнал (<http://livejournal.com>). Регионы авторов были приведены к единому виду, после чего из них были выбраны регионы с наибольшим количеством текстов. Корпус был очищен от выбросов – текстов, не представляющих интереса с точки зрения данного исследования. В данном исследовании были изучены различные наборы признаков, размеры обучающих коллекций и методы машинного обучения. Проведенные эксперименты показали, что большая часть текстов не содержит достаточно информации для определения региональной привязки, однако имеется существенная часть текстов, пригодных для региональной классификации.

GEOGRAPHIC LOCATION PREDICTION IN BLOGS

Morozov E.V., Bogdanova D.N.

LLC «AbiInfoPoisk», Moscow, Russia (127273, 2B.6, Otradnaya, Moscow), e-mail: eugene_m@abby.com

This paper presents research on geographical lexical variation detection. We present a new corpus of Russian blogs labeled with geographical information. The corpus was extracted from LiveJournal (<http://livejournal.com>). Only those blogs that contained enough information about the author were used in the experiments. We have performed outlier detection, and thus, removed spam and other irrelevant data. We have studied various feature sets and performed classification based on Support Vector Machine and Naïve Bayes algorithms. The obtained results show that geographic location prediction is a hard task, and many of the blogs do not contain enough information to determine location of their authors, even though in certain cases accurate classification is possible.