A Study of Top-K Measures for Discrimination Discovery

Dino Pedreschi
Dipartimento di Informatica
Università di Pisa, Italy
pedre@di.unipi.it

Salvatore Ruggieri
Dipartimento di Informatica
Università di Pisa, Italy
ruggieri@di.unipi.it

Franco Turini
Dipartimento di Informatica
Università di Pisa, Italy
turini@di.unipi.it

ABSTRACT

Data mining approaches for discrimination discovery unveil contexts of possible discrimination against protected-by-law groups by extracting classification rules from a dataset of historical decision records. Rules are ranked according to some legally-grounded contrast measure defined over a 4-fold contingency table, including risk difference, risk ratio, odds ratio, and a few others. Due to time and cost constraints, however, only the top-k ranked rules are taken into further consideration by an anti-discrimination analyst. In this paper, we study to what extent the sets of top-k ranked rules with respect to any two pairs of measures agree.

Categories and Subject Descriptors

H.2.8 [**Database Applications**]: Data Mining

General Terms

Algorithms, Legal Aspects

Keywords

Discrimination discovery, classification rules

1. INTRODUCTION

Human right laws [1, 10, 11] prohibit discrimination against protected groups on the grounds of race, color, religion, nationality, sex, gender, marital status, age and pregnancy; and in a number of settings, including credit and insurance; sale, rental, and financing of housing; personnel selection and wages; access to public accommodations, education, nursing homes, adoptions, and health care. Recently, the issue of discrimination analysis has been considered from a data mining perspective [5, 6]. *Discrimination discovery from data* consists in the actual discovery of discriminatory situations and practices hidden in a large amount of historical decision records. The aim is to unveil contexts of possible discrimination on the basis of *legally-grounded* measures of the degree of discrimination suffered by protected-by-law groups in such contexts. Reasoning on the extracted contexts can support all the actors in an argument about possible discriminatory behaviors. A complainant in a case can use them to find specific situations in which there is a *prima facie* evidence of discrimination against groups she belongs to. A decision maker can use them to prevent incurring in discriminatory decisions. Finally, control authorities can base the fight against discrimination on a formalized process of intelligent data analysis.

However, the actual discovery of discriminatory situations and practices may reveal an extremely difficult task. The reason is twofold. First, a huge number of possible contexts may, or may not, be the theater for discrimination. To see this point, consider the case of gender discrimination in credit approval: although an analyst may observe that no discrimination occurs in general, i.e., when considering the whole available decision records, it may turn out that it is extremely difficult for aged women to obtain car loans. Many small or large niches may exist that conceal discrimination, and therefore all possible specific situations should be considered as candidates, consisting of all possible combinations of variables and variable values: personal data, demographics, social, economic and cultural indicators, etc. Second, the interpretation of existing legislations lead to different quantitative measures of discrimination and, *a fortiori*, to different rankings of the possibly discriminatory contexts to be considered by an anti-discrimination analyst.

The first problem, i.e., extracting contexts of possible discrimination, has been considered in [5, 6] by resorting to the extraction of classification rules maximizing some discrimination measure defined over the 4-fold contingency table of the rule. In this paper, we focus on the second problem by studying *how the rankings induced by different discrimination measures affect the results of the analysis*. Due to time and cost constraints, an anti-discrimination analyst can afford to (manually) investigate a limited number of possible contexts/rules, hence typically she will extract the top-k rules with respect to a given discrimination measure. We intend to investigate whether such a set of rules differs significantly from one measure to another, or, if their rankings differs significantly.

This paper is organized as follows. In Sect. 2, we survey the various discrimination measures and recall the approach based on classification rule extraction and ranking. In Sect. 3-5, we study the distributions of values of pairs of measures on a reference dataset of credit applications. Finally, Sect. 6 summarizes the contributions of the paper.

SAC'12 March 25-29, 2012, Riva del Garda, Italy.

group	benefit: denied	benefit: granted	
protected	a	b	n_1
unprotected	c	d	n_2
	m_1	m_2	n

$$p_1 = a/n_1 \quad p_2 = c/n_2 \quad p = m_1/n$$

$$RD = p_1 - p_2 \quad RR = \frac{p_1}{p_2} \quad RC = \frac{1-p_1}{1-p_2}$$

$$OR = \frac{RR}{RC} = \frac{a/b}{c/d}$$

$$ED = p_1 - p \quad ER = \frac{p_1}{p} \quad EC = \frac{1-p_1}{1-p}$$

Figure 1: Discrimination measures.

2. BACKGROUND

Discrimination Measures. Consider a dataset of historical decisions about granting or not a benefit (e.g., a loan, a job, a wage increase, a school admission). A common tool for statistical analysis is provided by a 2×2, or 4-fold, contingency table, as shown in Fig. 1. Different outcomes between two groups are measured in terms of the proportion of people in each group with a specific outcome. Consider a subset of the decisions, which we call a *context*. The proportions of benefit denied for the protected-by-law group (p_1), the unprotected-by-law group (p_2) and the overall subset (p) are considered. A general principle is then to consider *group under-representation* in obtaining a benefit as a quantitative measure of discrimination against a protected-by-law (briefly, protected) group.

Group under-representation can be measured as differences and rates of these proportions, including:

- *risk difference* (RD $= p_1 - p_2$), also known as *absolute risk reduction*,
- *risk ratio* or *relative risk* (RR $= p_1/p_2$),
- *relative chance* (RC $= (1-p_1)/(1-p_2)$), also known as *selection rate*,
- *odds ratio* (OR $= p_1(1-p_2)/(p_2(1-p_1))$),

and the versions of RD, RR, and RC when the protected group is compared to the average proportion p, rather than to the proportion of the unprotected group: *extended difference* (ED $= p_1 - p$); *extended ratio* or *extended lift* (ER $= p_1/p$); *extended chance* (EC $= (1-p_1)/(1-p)$).

Level curves of RD, RR, RC, and OR over the 2-D risk plane [3] are shown in Fig. 2. The degree of observed disproportionate burden over the protected group is monotonic increasing for RD, RR, and OR (resp., ED, and ER), and monotonic decreasing for RC (resp., ED). Since one is interested in contexts of higher benefit denial (resp., lower benefit granting) for the protected group compared to the unprotected group or to the average, the values of interest for RR, OR, and ER are those greater than 1; for RD and ED are those greater than 0; and for RC and EC are those lower than 1. [5] showed that RR ≥ 1 iff OR ≥ 1 iff ER ≥ 1. It is readily checked that this is the case also iff RD ≥ 0 iff ED ≥ 0 iff RC ≤ 1 iff EC ≤ 1. Summarizing, all the measures agree on the contexts to be considered as potentially discriminatory. Our objective is to study whether this holds also when restricting to the top-k contexts.

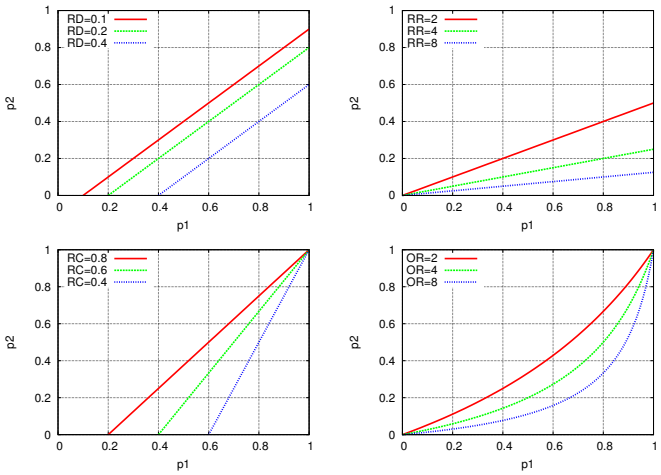

Figure 2: Level curves over the 2-D risk plane.

From a legal point of view, several measures are adopted worldwide. UK law [9](a) mentions risk difference, EU Court of Justice has given more emphasis on the risk ratio (see [8, Section 3.5]), and US laws and courts mainly refer to the selection rate. Notice that the risk ratio is the ratio of the proportions of *benefit denial* between the protected and unprotected groups, while selection rate is the ratio of the proportions of *benefit granting*. Also, the odds ratio has been widely considered in legal research studies.

Discrimination Discovery by Rule Mining. The legal principle of under-representation has inspired existing approaches for discrimination discovery based on pattern mining. Starting from a dataset of historical decision records, [5, 6] propose to extract classification rules of the form $\mathbf{A}, \mathbf{B} \rightarrow \mathbf{C}$, called *potentially discriminatory* (PD) rules, to unveil contexts $\mathbf{B}$ of the dataset where the protected group $\mathbf{A}$ suffered from under-representation with respect to the decision $\mathbf{C}$. $\mathbf{A}$ is a non-empty itemset, whose elements denote a fixed set of protected groups. $\mathbf{C}$ is a class item denoting the negative decision, e.g., credit denying, application rejection, job firing, and so on. Finally, $\mathbf{B}$ is an itemset denoting a context of possible discrimination. The degree of under-representation is measured by a measure from Fig. 1, where the contingency table refers to only those records satisfying $\mathbf{B}$. As an example, RACE=BLACK, PURPOSE=NEW_CAR $\rightarrow$ CREDIT=NO is a PD rule about denying credit (the decision $\mathbf{C}$) to blacks (the protected group $\mathbf{A}$) among those applying for the purpose of buying a new car (the context $\mathbf{B}$). PD rules are ranked by their value of the discrimination measure. The approach has been implemented in [7].

The German Credit Dataset. As a running example of our study, we consider the public domain German credit dataset [4], which consists of 1000 records over bank account holders. It includes attributes on personal properties, past/current credits, employment status, and on personal status. The decision attribute represents the good/bad creditor decision assigned to the bank account holder. The protected groups considered in the analyses include female non-single, foreign workers, and senior people. When not otherwise specified, classification rules are extracted with a minimum support of 2%.

M1	M2	1^{st} M2	k^{th} M2	top missed	bottom included
OR	RR	9.515	2.100	2.477	1.517
RR	OR	17.24	2.867	4.600	2.385
RD	RR	9.515	2.100	6.262	1.400
RR	RD	0.589	0.199	0.364	0.099
RC	RR	9.515	2.100	7.765	1.240
RR	RC	0.215	0.746	0.429	0.892
RD	OR	17.24	2.867	7.851	2.286
OR	RD	0.589	0.199	0.255	0.104
RC	OR	17.24	2.867	10.36	1.687
OR	RC	0.215	0.746	0.524	0.891
RC	RD	0.589	0.199	0.242	0.127
RD	RC	0.215	0.746	0.632	0.790
ED	ER	1.808	1.234	1.360	1.131
ER	ED	0.254	0.076	0.128	0.043
EC	ER	1.808	1.234	1.448	1.062
ER	EC	0.477	0.883	0.706	0.949
EC	ED	0.254	0.076	0.088	0.041
ED	EC	0.477	0.883	0.706	0.905
ER	RR	9.515	2.100	9.515	1.309
RR	ER	1.808	1.234	1.636	1.011
ED	RD	0.589	0.199	0.589	0.102
RD	ED	0.254	0.076	0.142	0.005
EC	RC	0.215	0.746	0.406	0.842
RC	EC	0.477	0.883	0.804	0.992

Table 1: 1^{st} and k^{th} ranked M2 value of classification rules extracted from the German credit dataset, with $k = 1000$. Top M2 value missed by and bottom M2 value included in the top-k rules w.r.t. M1.

3. MEASURES OVER P_1 AND P_2

Let us start by comparing the measures defined over proportions p_1 and p_2, namely RR, OR, RD, and RC.

RR vs OR. Fig. 3 (left) shows the 2-D risk plane for the top-k classification rules, with $k = 1000$, extracted from the German credit dataset and ranked w.r.t. RR and OR. More in detail, in such a plot (and in the other similar 2-D plots) we show:

- as red points those rules (denoted as RR $\cap$ OR) belonging to the top-k set of both measures;
- as green points those rules (denoted as RR $\setminus$ OR) belonging to the top-k set of RR only;
- as blue points those rules (denoted as OR $\setminus$ RR) belonging to the top-k set of OR only;
- the level curve corresponding to the RR of the k^{th} ranked rule (in the plot, RR = 2.10), i.e., such that all top-k rules w.r.t. RR lie below it;
- the level curve corresponding to the OR of the k^{th} ranked rule (in the plot, OR = 2.87), i.e., such that all top-k rules w.r.t. OR lie below it.

OR has been often regarded as a good approximation of RR, due to the fact that for $p_1 \approx 0$, it turns out $1 - p_1 \approx 1$, and $1 - p_2 \approx 1$ (since we assume $p_1 \geq p_2$), which in turn imply RC ≈ 1, and then OR = RR/RC $\approx$ RR.

In Fig. 3 (left), it is readily checked that for $p_1 \leq 0.4$, the level curve of OR is very close to the one of RR. In other terms, the top-k rules w.r.t. RR missed by the top-k rules w.r.t. OR (i.e., the green points in the plot) lie very close to the lowest ranked rule w.r.t RR. Hence, the most relevant rules w.r.t. RR are not missed by looking at the most relevant rules w.r.t. OR. The highest RR value of a missed rule is 2.477, which is close to the RR = 2.10 of the k^{th} ranked rule and well below to the RR = 9.515 of the 1^{st} ranked rule. However, the blue points in Fig. 3 (left) highlight a number of rules in the top-k set w.r.t. OR that do not belong to the top-k set w.r.t. RR. Those rules are ranked as interesting w.r.t. OR but they are not w.r.t. RR. As an example, the top-k set w.r.t. OR contains a rule with a RR of only 1.52.

Table 1 details these statistics for all pairs of measures. It answers the question: "*How interesting M2 values are missed and how irrelevant M2 values are considered when using a measure M1 in place of M2?*". If missed and included values are close to the k^{th} ranked M2 value, then M1 can be considered a good proxy for M2, covering all its top-ranked rules while not covering low-ranked ones.

Finally, it is worth noting that, in general, the level curves of RR vs OR have the form as in Fig. 3 (left). In fact, the level curve of RR is p_1/v, for v being the RR of the k^{th} ranked rule w.r.t. RR, and the level curve of OR is $p_1/(p_1+v'(1-p_1))$, for v' being the RR of the k^{th} ranked rule w.r.t. OR. By basic algebra, they intersect only at $p_1 = 0$ and at $p_1 = (v' - v)/(v' - 1)$.

RR vs RD, RC. Since the level curves of RR and RD (resp., RC) are straight lines, they intersect in one point only, as shown in Fig. 3 (center) (resp., right). The number of rules belonging to the top-k set of only RR or of only RD (resp., RC) is now considerably higher.

Approximating RR by means of RD (resp., RC) in top-k rule extraction leads to miss the green points, with a theoretically upper unbounded RR. In our example, a rule with RR = 6.262 (resp., RR = 7.765) would be missed. Also, rules with RR much lower than the k^{th} ranked rule would be considered. In our example, again, a rule with RR = 1.4 (resp., RR = 1.24) would be considered.

Consider now approximating RD (resp., RC) by means of RR. The rules missed by RR (the blue points) have a RD as high as 0.364 (resp., RC as low as 0.429). Conversely, the rules considered by RR, but not by RD (resp., RC), have a minimum RD of 0.099 (resp., RC of 0.892).

OR vs RD vs RC. Fig. 4 shows the 2-D plots for OR, RD, and RC. Curves of OR and RD intersect in two points[1], while curves of OR and RC and of RD and RC intersect in one point only. The number of rules belonging to the top-k set of only OR (resp., RD) or of only RD (resp., RC) appears moderate, in the same order as in the case RR vs OR. We refer the reader to Table 1 for details of using one of the three measures as an approximation for another.

Intersection and Order Correlation. Starting from the sample $k = 1000$ considered so far, it is now legitimate to ask ourselves how many rules are shared, for generic k's, between the top-k sets of any two measures. The answer also solves the question of how many rules in one of the two top-k sets belong to it only, namely k minus the number of rules that belong to both. Intuitively, for increasing k, the top-k sets tend to converge to the set of all rules. We expect then that the fraction of shared rules is a monotonically increasing function over k. This is confirmed in Fig. 5 (left, center). RD vs RC and RR vs OR show the largest over-

[1]The intersection points occur at the solutions of a quadratic equation in p_1 obtained by equating the level curve of OR ($p_2 = p_1/(p_1+v(1-p_1))$, for v being the OR of the k^{th} ranked rule w.r.t. OR) and RD ($p_2 = p_1 - v'$, for v' being the RD of the k^{th} ranked rule w.r.t. OR).

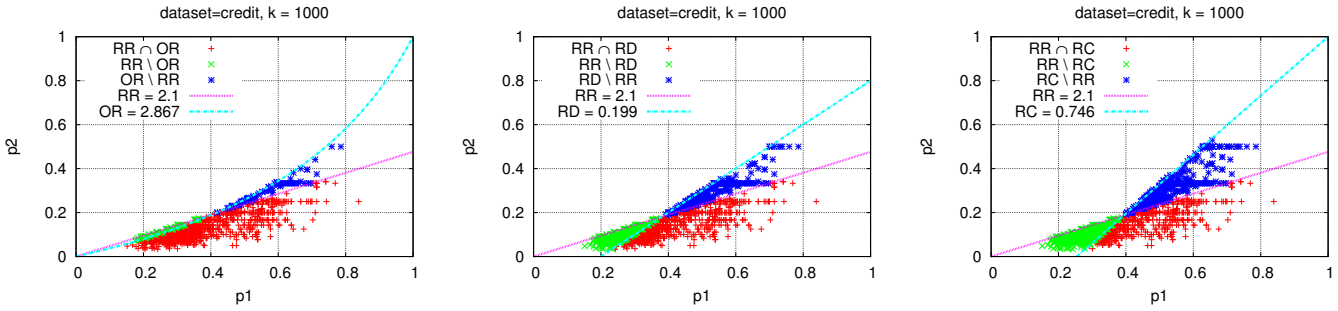

Figure 3: Classification rules over the 2-D risk plane: RR vs OR (left), RR vs RD (center), RR vs RC (right).

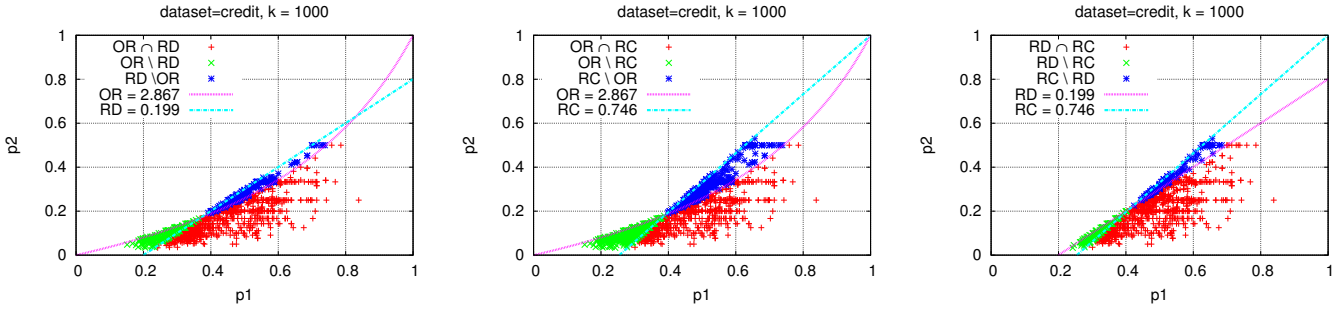

Figure 4: Classification rules over the 2-D risk plane: OR vs RD (left), OR vs RC (center), RD vs RC (right).

lapping set of rules. RR vs RC has the smallest overlapping set. This is rather relevant from a practical (legal) point of view: *EU laws (RR) and US laws (RC) do not lead the anti-discrimination analyst to reason on a significantly overlapping set of potentially discriminatory classification rules.*

Another intuitive question is whether the rankings of two measures are correlated on the set of overlapping rules, i.e., whether the two measures impose the same order on the rules shared by their top-k sets. Kendall's correlation coefficient τ represents the difference between the probability that in the observed data two variables are in the same order versus the probability that the two variables are in different orders. It ranges from -1 (negative correlation, i.e., inverse ordering) to +1 (positive correlation, i.e., same ordering). Fig. 5 (right) shows the Kendall's τ by varying k for three pairs of measures, the top two and the bottom one with respect to the overlapping fraction. Unfortunately, τ appears to be rather unstable, apart for RR vs OR for which it is almost over 0.5. Since the top-k rules are only a *prima facie* evidence of discrimination, they are (manually) screened by the anti-discrimination analyst to validate their legal relevance. The negative conclusion from Fig. 5 (right) is that, *using a discrimination measure to rank the rules to be validated does not help in covering earlier the most relevant rules with respect to a different measure.*

4. MEASURES OVER p_1 AND p

We turn now the attention on measures defined over proportions p_1 and p, namely ER, ED, and EC. The 2-D risk planes in Fig. 6 are defined over p_1 as the x-axis, and p as the y-axis. Compared to their respective (i.e., RD for ED, RR for ER, and RC for EC) plots from Fig. 3 and Fig. 5, it is readily checked that the points:

- are more flattened around the level curve of the k^{th}-ranked rule. This is explained by a smaller variability range of the measures, as shown in Table 1. As an example, ER ranges from 1.808 to 1.234, whilst RR ranges from 9.515 to 2.1;
- are more centered in the 2-D risk plane. Since, in general, $p \geq min\{p_1, p_2\}$, any rule plotted in the p_1-p_2 risk plane would be plotted on a higher (if $p_1 \geq p_2$, otherwise in a lower) y-point in the p_1-p risk plane.

The fraction of overlapping rules between any pair of the ER, ED, and EC measures appears considerably high. This is confirmed in Fig. 7 (left), showing that ED and EC almost yield the same top-k set of rules, even for reasonably low k. Table 1 also shows that the rules missed by ED (resp., EC) would not rank high w.r.t. EC (resp., ED). Nevertheless, we should not draw a general conclusion. So far, we have considered classification rules with a minimum support of 2%. The assumption of extracting rules with a minimum support threshold is natural in discrimination analysis: it amounts at extracting contexts where a minimum number of persons of the protected group are treated unfavourably. Do the trends of Fig. 5 (left, center) and Fig. 7 (left) hold in general? Unfortunately, this is not the case. For lower minimum support thresholds, the fraction of overlapping rules tends to decrease rapidly, as shown in Fig. 7 (center) for $ED \cap EC$. As an example, for a 0.5% minimum support, to reach an overlapping fraction of 50%, we must consider at least 500 rules. Intuitively, this fact can be explained by the exponential growth of the number of classification rules extracted, and, *a fortiori*, of the number of possible contexts where the ED and EC rankings significantly differ.

Finally, Fig. 7 (right) shows that a similar trend occurs for the *adult* dataset [4] as well. Strictly speaking, *adult* is

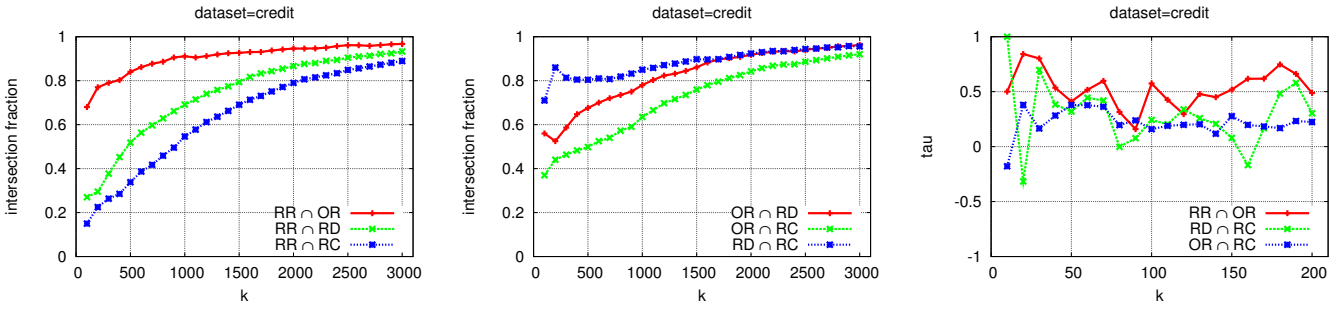

Figure 5: Number of rules in both the top-k sets of two measures over k (left, center). Kendall's τ (right).

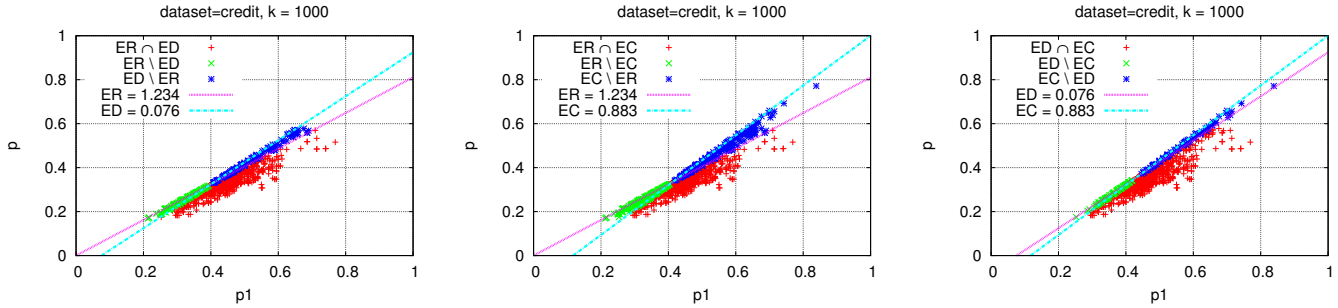

Figure 6: Classification rules over the p_1-p risk plane: ER vs ED (left), ER vs EC (center), ED vs EC (right).

not concerned with decisions, but rather with 48,842 census records related to people income (low/high). The objective of discrimination analysis on this dataset is to discover forms of statistical discrimination [2], namely patterns of lower performances, skills or capacities of protected groups in order to prevent their use in future (possibly, automated) decisions. While a support of 2% is reasonable for the German credit dataset, amounting at 20 records, it becomes too high for *adult*, amounting at 976 records. Niche contexts of possible discrimination are likely to occur for much less than 976 persons. Unfortunately, as shown in Fig. 7 (right), for lower minimum support thresholds, the top ranked rules w.r.t. ED and EC considerably differ for a reasonably low k.

5. COMPARING GROUPS OF MEASURES

After having looked separately at the measures over p_1-p_2 and over p_1-p, we now compare the related pairs of measures RR and ER, RD and ED, RC and EC. Intuitively, ER is defined as RR apart from using p (the proportion over the whole context) as the comparator against p_1 instead of using p_2 (the proportion over the unprotected group). The same holds for ED vs RD, and for EC vs EC. Therefore, by comparing RR and ER we are actually studying the effects of a critical choice from a legal point of view.

Fig. 8 (left) plots the top-k rules w.r.t. the RR and ER measures. In order to separate the rules that occur only in one top-k set, the 2-D risk plane is now defined along the axis p and p_2. Let v be the RR value for the k^{th}-ranked rule w.r.t. RR (for $k = 1000$, $v = 2.1$), and v' be the ER value for the k^{th}-ranked rule w.r.t. ER (for $k = 1000$, $v' = 1.234$). A rule belongs to the top-k set of only RR iff $p_1/p_2 \geq v$ and $p_1/p < v'$. By elementary algebra, this occurs only if $p/p_2 > v/v'$ (in our example, $v/v' = 2.1/1.234 = 1.702$). Similarly, a rule belongs to the top-k set of only ER only if $p/p_2 < v/v'$ Therefore, the line $p/p_2 = v/v'$ separates the two sets. Notice, however, that this is a necessary condition only, not a sufficient one. Rules shared between the two top-k sets lie around such a line.

Fig. 8 (center) considers the RD and ED measures. We reason as before. Let v be the RD value for the k^{th}-ranked rule w.r.t. RD (for $k = 1000$, $v = 0.199$), and v' be the ED value for the k^{th}-ranked rule w.r.t. ED (for $k = 1000$, $v' = 0.076$). The separation of the rules occurring only in one top-k set is obtained by the system of equalities $p_1 - p_2 = v$, $p_1 - p = v'$, which has solution $p - p_2 = v - v'$ (in our example, $v - v' = 0.199 - 0.076 = 0.123$).

Fig. 8 (right) plots the RC and EC measures. Once again, let v be the RC value for the k^{th}-ranked rule w.r.t. RC (for $k = 1000$, $v = 0.746$), and v' be the EC value for the k^{th}-ranked rule w.r.t. EC (for $k = 1000$, $v' = 0.883$). The separation of the rules occurring only in one top-k set is now obtained by the system of equalities $(1 - p_1)/(1 - p_2) = v$, $(1-p_1)p-p_2 = v$, which has solution $(1-p)/(1-p_2) = v/v'$ (in our example, $v/v' = 0.746/0.883 = 0.845$).

In all three previous cases, the number of overlapping rules is very low. Most of the points in Fig. 5 are either green or blue, very few are red. In terms of the discrimination analysis, this means that taking as a reference proportion in the under-representation principle the set of all people in a context or only the set of the unprotected people significantly affects the top-k set of rules to be looked at by the anti-discrimination analyst.

6. CONCLUSIONS

Classification rules have been adopted for discrimination discovery by interpreting legal measures of under-representation as interestingness measures over a 4-fold contingency table. Extracted classification rules are ranked according to

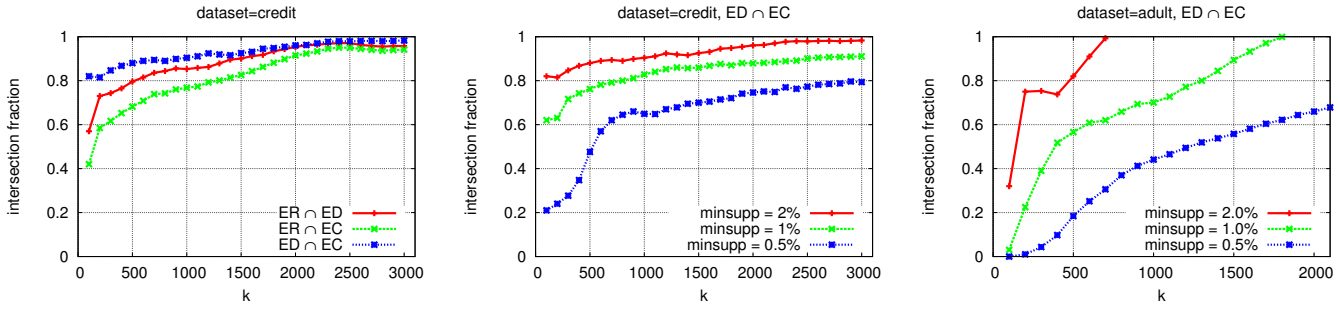

Figure 7: Number of rules in both the top-k sets of two measures over k.

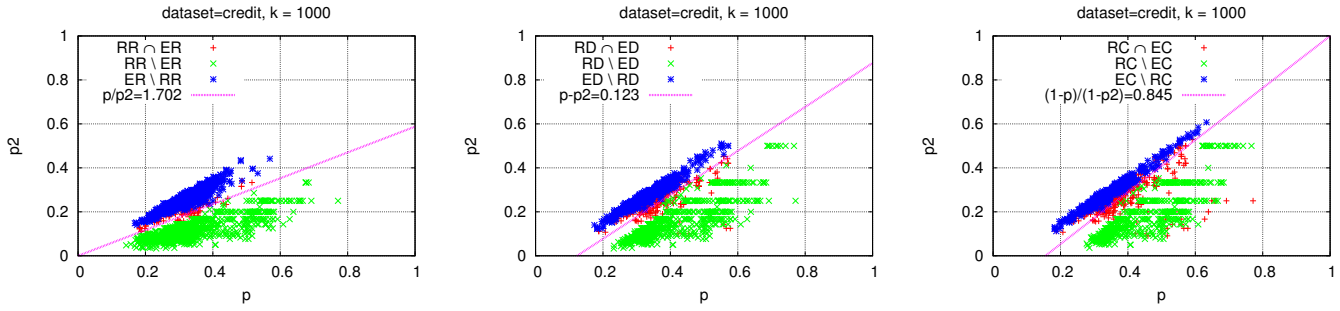

Figure 8: Classification rules over the p-p_2 risk plane: RR vs ER (left), RD vs ED (center), RC vs EC (right).

a measure, and only the top-k rules are taken into (manual) consideration by the anti-discrimination analyst for further investigation. However, which measure has to be considered is not always known in advance, e.g., as in a legal case before a court. This gives raise to the problem of understanding at which extent any two measures yield the same set of top-k rules. We studied such a problem on a reference dataset of credit applications, with some theoretical analyses applicable in general. We can summarize the following specific conclusions of this study: (1) among all the pairs of measures, RR and OR, RD and RC, and ED and EC, show the largest overlapping set of top-k rules; (2) among all pairs of measures, RR (mainly adopted in the US legal system) and RC (mainly adopted in the EU legal system) have the smallest overlapping set of top-k rules; (3) in all cases, however, the set of overlapping rules rapidly degrades as the minimum support threshold of the extracted rules decreases; (4) for none of the pairs of measures there is a stable and relevant correlation between the rankings imposed by the two measures over the set of overlapping rules; (5) pairs of measures that differ only in the comparator proportion (RR vs ER, RD vs ED, RC vs EC) lead to top-k sets with a small percentage of overlapping rules. We believe that the above conclusions represent a relevant contribution of data mining to the legal debate on anti-discrimination, consisting in analytical means to support the choice, e.g., at legislative level, of a quantitative measure of the qualitative legal principle of under-representation.

7. REFERENCES

[1] European Union Legislation. (a) Racial Equality Directive, 2000; (b) Employment Equality Directive, 2000; (c) Gender Employment Directive, 2006; (d) Equal Treatment Directive (proposal), 2008.

[2] H. Fang and A. Moro. Theories of statistical discrimination and affirmative action: A survey. In *Handbook of Social Economics, Vol 1B*. North-Holland, 2010.

[3] J. Li and Q. Yang. Strong compound-risk factors: Efficient discovery through emerging patterns and contrast sets. *IEEE Trans. on Information Technology in Biomedicine*, 11(5):544–552, 2007.

[4] D.J. Newman, S. Hettich, C.L. Blake, and C.J. Merz. UCI repository of machine learning databases, 1998. http://archive.ics.uci.edu/ml.

[5] D. Pedreschi, S. Ruggieri, and F. Turini. Measuring discrimination in socially-sensitive decision records. In *Proc. of SDM 2009*, pages 581–592. SIAM, 2009.

[6] S. Ruggieri, D. Pedreschi, and F. Turini. Data mining for discrimination discovery. *ACM Trans. on Knowledge Discovery from Data*, 4(2):Article 9, 2010.

[7] S. Ruggieri, D. Pedreschi, and F. Turini. DCUBE: Discrimination discovery in databases. In *Proc. of SIGMOD 2010*, pages 1127–1130. ACM, 2010.

[8] D. Schiek, L. Waddington, and M. Bell, editors. *Cases, Materials and Text on National, Supranational and Int. Non-Discrimination Law*. Hart Publ., 2007.

[9] U.K. Legislation. (a) Sex Discrimination Act, 1975, (b) Race Relation Act, 1976.

[10] United Nations Legislation. (a) Universal Declaration of Human Rights, 1948, (c) Convention on the Elimination of All forms of Racial Discrimination, 1966, (d) Convention on the Elimination of All forms of Discrimination Against Women, 1979.

[11] U.S. Federal Legislation. (a) Equal Credit Opportunity Act, 1974; (b) Fair Housing Act, 1968; (c) Employment Act, 1967; (d) Equal Pay Act, 1963.