Analysis of Demographical Factors' Influence on Websites' Credibility Evaluation*

Maria Rafalak[1], Piotr Bilski[2], and Adam Wierzbicki[1]

[1] Polish-Japanese Institute of Information Technology, (PJIIT), Warsaw, Poland
[2] Institute of Radioelectronics, Warsaw University of Technology, Warsaw, Poland
{maria.rafalak,awierzbicki}@pjwstk.edu.pl,
pbilski@ire.pw.edu.pl

Abstract. The paper presents results of an experiment conducted in 2013 via Amazon Mechanical Turk Platform (www.mturk.com) aimed at creating a classifier predicting online content credibility evaluation misjudgement tendencies. The rough sets based module processes demographic variables describing each participant and predicts his/her misjudgement tendency. Data collection method, data-set preparation are described in detail. Next the rough set methodology is introduced explaining the process of training and validating using available data. Experimental results are presented in detail showing the classification accuracy for various configurations of rough-sets algorithms. The analysis of importance of subsequent demographic variables on prediction efficiency is discussed as well. The paper is concluded with future prospects and future applications of implemented methodology.

Keywords: Demographic variables, rough sets, classification, artificial intelligence, credibility assessment.

1 Introduction

Internet has undeniably become a vital part of our everyday lives. Among many versatile functions that it serves, probably the most important aspect of Internet use is the functionality of information search. Currently www technologies allow for free and simple content transfer between users. In *Web 2.0* paradigm, every user is the author of potentially important information. The credibility of such data is often dubious and should be treated with precaution. This also refers to the opinions and judgments made about the Internet content. The tendency to overestimate the web pages' credibility and usefulness of their information was discovered recently. There is the psychological effect causing better judgment (higher grades) than the page actually deserves. The examples are the online movie database (www.imdb.com), book community *Books Crossing*, music review systems or *Amazon* – online retailer [1]. The reason for this behaviour is unknown, but the bias, i.e. the constant shift towards the positive opinions was confirmed. People

* Research supported by the grant "Reconcile: Robust Online Credibility Evaluation of Web Content" from Switzerland through the Swiss Contribution to the enlarged European Union.

M. Kurosu (Ed.): Human-Computer Interaction, Part III, HCII 2014, LNCS 8512, pp. 57–68, 2014.

rarely verify information found online with other sources like books or newspapers [2, 3]. As the number of websites grows exponentially, the tool for distinguishing the relevant content from the unimportant one is required. The topic of this paper is the analysis of the websites assessment made by Internet users related to their demographic characteristics (such as age, sex, education, living location or political views). Based on these features the authors determine the users' tendency to over- or underestimate the credibility of selected websites.

1.1 Definition of Credibility

Credibility can be understood as a personal belief that certain piece of information can be trusted. It is worth noticing that using concept of truth is purposely avoided here. Adopted definition refers to subjective conviction of individual making judgment rather than construct of objective truth. Existence of the latter, from our point of view, remains in the sphere of philosophical inquiries.

Research trends dedicated to credibility of online content so far have concentrated on two big issues: technical characteristics of websites (such as i.e. design, functionality or thematic area) and individual differences of users. Such factors as familiarity with the topic [1, 2], the first impression effect [3,4] or preferred heuristics in making evaluations [5] are claimed to moderate final credibility judgments of online content.

Defining credibility evaluations as subjective implies that identical stimulus is likely to elicit different reactions depending on characteristics of a person making judgment. One of the recently most popular theoretical frameworks giving foundations for scientific investigation of credibility evaluations is *Prominence - Interpretation Theory* (*P-I*) proposed by Fogg [4]. In general, according to *P-I* final credibility evaluation depends heavily on particular elements user notices on the website (prominence) and the meaning he/she attaches to them (interpretation). Among the factors affecting *prominence* Fogg lists user's task, motivation, experience, type of website content (informational vs transactional) and individual differences (literacy level, cognitive abilities, psychological traits etc.). *Interpretation* depends on users' knowledge, assumptions, goals and context of making evaluation. The process of noticing website element and interpreting it has an iterative character of subjectively defined stop criterion and leads to formulating website's overall credibility evaluation.

Basing on factors influencing final judgments Fogg [4] distinguishes four types of credibility: *presumed*, *earned*, *reputed* and *surface credibility*.

Presumed credibility reflects general assumptions and stereotypes held in evaluator's mind i.e. those connected with site identifiers (.gov, .org, .edu, etc.). *Surface credibility* is connected with general impression about the website including its design, usability and presence of advertisements. *Earned credibility* refers to experience user has with the website including its responsiveness. Finally, *reputed credibility* stands for seals of approval, links to other reputed websites etc. This distinction is not disjunctive and which of those factors are taken into account by user depends strongly on individual differences.

1.2 Credibility Evaluation Aid

There are two main approaches of giving Internet users guidance about how to assess website credibility. The probably most intuitive way are checklists which are specially designed lists of statements, questions or phrases supposed to sensitize user on not credible content. One of the main advantages of this approach is teaching a habit to search for certain website elements that may suggest whether the source should be trusted. This method however is time consuming and requires some effort from user. Following tools may serve as an example of checklist approach: *University of Maryland Library checklist*[1] , *Berkeley checklist*[2], *Health On the Net*[3], *Widener University checklist*[4] and *Discern*[5].

Second option for giving users clues about credibility of online content bases on crowdsourcing approach. Recommender or rating systems gather votes from websites' reviewers and combine them into a credibility rating. Users registered in such systems are able to vote and view websites ratings. This approach does not consume user's time unless they wants to make a contribution. Following tools may serve as an example of crowdsourcing approach: *ReConcile*[6], *MyWot*[7], *Factlink*[8], *Hypothes.is*[9].

1.3 Positive Bias in Credibility Evaluations

Adopting crowdsourcing approach in website credibility estimation is not free of drawbacks. It has been noted that wisdom of crowd can sometimes lead to overenthusiastic credibility evaluations. Vassilis Kostakos [6] in his paper showed that rating systems of such popular online services like *Amazon*, *BooksCrossing* or *Imdb* are negatively skewed. The same pattern has been distinguished in restaurant [7] or music review systems [8]. As opposed to online auction systems giving reviews does not imply any reciprocity. This is why phenomenon of overenthusiastic positive evaluation in this context cannot be explained in terms of strategic reasoning as suggested by [9]. The mechanism explaining users' online behaviour is yet to be determined.

1.4 Our Study General Idea

Studying credibility evaluations is a challenging task as by definition the construct is subjective and easily influenced by many factors. Polish-Japanese Institute of Technology is currently working on creating an online platform (www.reconcile.pl) that

---

1 http://www.lib.umd.edu/binaries/content/assets/public/usereducation/evaluating-web-sites-checklist-form-fall-2012.pdf
2 http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/EvalForm.pdf
3 http://www.hon.ch/
4 http://www.widener.edu/about/campusresources/wolfgram library/evaluate/original.aspx
5 http://www.discern.org.uk/
6 http:// www.reconcile.pl
7 https://www.mywot.com/
8 https://factlink.com/
9 http://hypothes.is/

would support evaluation of web content. This system is designed to take into account both credibility evaluations of registered users and websites' characteristics. However, it is feared that specific misjudgement tendencies presented by the users (i.e. positive bias) may reduce the effectiveness of credibility estimations given by the system. In order to avoid such effect we aspire to create a mechanism that would identify users with tendencies to overestimate or underestimate websites' credibility and in the future enable to correct their votes in the reconcile system.

We hypothesize that individual differences which according to Fogg's *P-I theory* [4] determine prominence of particular website elements and influence credibility evaluation, may be connected with demographic background of users. Demographic variables such as sex, age or education are easily accessible in many systems as they usually need to be defined in registration forms. Therefore proving the usability of demographic data in predicting specific credibility misjudgement tendencies would be beneficial not only for *reconcile* project but for all automated systems designed to aid credibility evaluations.

2 Experiment Design

Experiment was conducted via *Amazon Mechanical Turk* online platform (www.mturk.com) in the first quarter of 2013. Participants were asked to evaluate credibility of presented websites on five point Likert scale (where 1-very poor, 5-very high). Websites were assigned to participants at random. Every participant could decide how many websites he/she wanted to evaluate, getting monetary reward for every vote (0.35-0.55 /evaluation, depending individual total evaluation number). In order to eliminate the effect of websites' design on credibility evaluations we decided to create big website corpus so that individual characteristics of websites' could be neglected in final data analysis. Websites selected for the experiment (N = 4699) represented five broad topics: 'medicine', 'personal finance', 'healthy lifestyle', 'politics, economy, ecology', 'entertainment'. Websites' external credibility index (www.mywot.com) was monitored. The index can take values from 0-100 range. Mean reference credibility index in the experiment website corpus was 72.

Participants were recruited from registered mTurk American users with at least 75% approved assignments. Apart from giving credibility evaluations they shared some basic demographic information including: sex, birth year, education, political views, state of residence, and wage level.

We decided to use artificial intelligence method, namely rough sets, to build a classifier aimed at predicting specific misjudgement tendencies.

3 Data-Set Preparation

Over two thousand mTurk users (N = 2075) took part in our study (986 males, 1419 females). Median age of participants was 27 (minimum age 16, maximum age 90). On average every person evaluated three websites.

In order to prepare data-set to the rough set analysis the following steps have been applied to the original dataset. As the reference website credibility index (www.mywot.com) varies from 0 to 100 the distribution of websites' reference credibility has been transformed to standardized distribution ($M = 0$; $SD = 1$).

According to the experiment design every participant evaluated declared number of randomly assigned websites. Therefore for every person we calculated median of evaluations given on five point Likert scale. Next we transformed this distribution of median values to standardized distribution ($M = 0$; $SD = 1$).

For every *website-participant* pair we calculated the following difference:

$$\textit{standardised median evaluation of person} - \textit{standardised credibility index} \ .$$

For every participant we calculated median of the abovementioned difference. After examining the distribution of those median values we assigned respondents to three distinct categories: *under-raters* (*UR*; showing tendency to underestimate websites' credibility), *over-raters* (*OR*; showing tendency to overestimate websites' credibility), *average-raters* (*AR*; giving adequate judgments about websites' credibility). We arbitrarily decided that people obtaining scores below or equal the first quartile of the abovementioned distribution (Q_1 = - 0.944) were be classified as *under-raters* (*UR*; $N = 515$), people obtaining results over or equal third quartile (Q_3 = 0.859) were classified as *over-raters* (*OR*; $N = 1041$) and all other participants were qualified as *average-raters* (*AR*; $N = 519$). We chose such cut points in order to distinguish users with most extreme misjudgement tendencies.

4 Description of the Data Analysis Method

Rough Sets (*RS*) are the standard classification tool of large data sets, used widely in the biology [10a] and, financial [10,11] engineering [12] sciences. Their advantage is considering the measurement uncertainty (such as missing data or incorrect values of attributes), which is typical in real-world applications. As opposed to the classical theory of sets (where the object can belong to only one set), in *RS* it is possible to have objects belonging to multiple sets "in some degree", which expresses the complexity of real-world problems.

The basis for the analysis of data is the decision table, i.e. the structure $\boldsymbol{D_T}$, containing the information system $\boldsymbol{I_S}$ (1) supplemented with the information about the category of objects $\boldsymbol{d}$.

$$\boldsymbol{I_S} = (\boldsymbol{U}, \boldsymbol{A}) \tag{1}$$

It is the pair consisting of the set of objects **A** represented by their attributes (here demographic variables), while **U** is the universe, i.e. the set of all possible objects (Internet users).

This way the original data set, described in section 3, can be treated as $\boldsymbol{D_T}$ (2), where with objects are described by the *under-*, *average-* or *over-rater* category $\boldsymbol{d}$, based on the analysis presented in section 3.

$$\boldsymbol{D_T} = (\boldsymbol{I_S}, \cup \{d\}) = (\boldsymbol{U}, \boldsymbol{A} \cup \{d\}) \tag{2}$$

The *RS* operation uses the indiscernibility relation, findings groups of attributes describing examples, which makes objects non-distinguishable from each other. The equivalence relation between objects describes the redundancy existing in the decision table, which is the requirement for extracting knowledge from the available data.

The application of *RS* as the classification module requires extracting knowledge from the training data *L* and verification of the generalization abilities using the validating set *V*, both having the form (2). This approach falls into the supervised learning scheme, where the desired output *d* of the classification module for the selected set of input data is known. The *RS*-based module produces the classification hypothesis c for each object oi from *V*, which is confronted with *d*. It is crucial for *L* and *V* to be disjunctive, therefore each object (participant) should belong to only one set. The classification quality is measured by the following sample error es, calculating the fraction of incorrectly classified objects by the module with respect to the number of objects in *V*:

$$e_s(\boldsymbol{V}) = \frac{\left|o_i \in \boldsymbol{V} : d(o_i) \neq c(o_i)\right|}{|\boldsymbol{V}|} \cdot 100\% \tag{3}$$

The alternative measure of the classification quality, also used in our research, is the supplement 1-$e_s(\boldsymbol{V})$. Because the *RS* algorithms are parameterized, the training process must be repeated multiple times for different configurations, until their optimal set (minimizing the error (3)) is found. The training of *RS* ad knowledge extraction operation is presented in Fig. 1. The subsequent operations are parameterized by vectors $\boldsymbol{p_d}$ (for discretization) and $\boldsymbol{p_r}$ (for reduction). Their content depends on the particular applied algorithm.

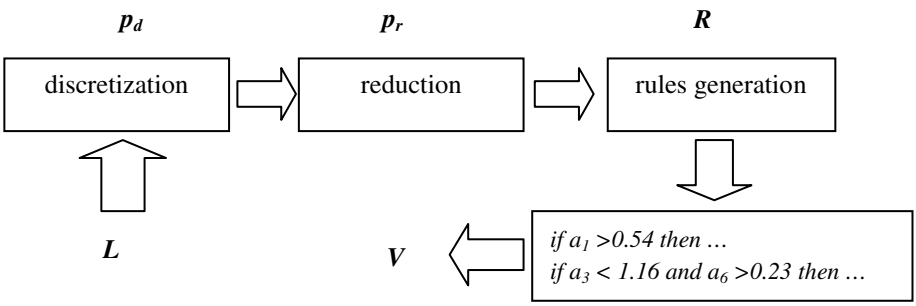

Fig. 1. Training scheme of the RS-based classification module

The following steps are implemented here:

1. Discretization of selected attributes from $\boldsymbol{D_T}$ to exchange their continuous values with the discrete numbers of intervals. The latter are described by the index of the attribute and threshold values. There are many discretization algorithms available, such as *Boolean Reasoning*, *Equal Frequency Binning* (*EFB*), *naive* or *semi-naive approaches*. It was proven in [13] that the discretization has the decisive impact on

the classification efficiency of *RS*, therefore the proper algorithm and its parameters have to be carefully selected. The characteristic feature of the demographic data is their discrete nature. Such attributes, as sex or level of education, have only a few different values, therefore are not suitable for this operation. The only attribute, which had to be discretized was the birth year of analysed participants. Although it is also discrete, the cardinality of its values range makes it quasi-continuous. The most important discretization parameter $\boldsymbol{p_d}$ is the number of intervals, dividing the continuous range of attribute values. In multiple applications *EFB* approach was considered the most useful, but the optimal number of intervals had to be selected heuristically. Other methods, especially naïve and semi-naïve algorithms generate too many intervals, leading to the classification deterioration, disregarding the reduction method.

2. Calculating reducts, i.e. the subsets $\boldsymbol{R}$ of attributes describing categories $\boldsymbol{d}$ as accurately as the whole information system. Because usually multiple reducts can be found in the set, it is desired to find as many of them as possible. The reduct is represented as the binary vector of the length equal to the number of attributes in $\boldsymbol{D_T}$ (1), with ones on the positions corresponding to the particular attributes. For example, the reduct containing attributes No. 2,3 and 5 (out of six) has the form:

$$\{0,1,1,0,1,0\}$$

Searching for reducts is the *NP-hard* problem [14], therefore no exact algorithms can be used for this purpose. Instead, approximate methods are employed, which find as many reducts as possible in reasonable time. The most popular ones are *genetic*, *Johnson's* and *Holte's* algorithms, differing in the method of searching through the space of possible atrributes' combinations. Because the analyzed sets have six attributes, it is possible to generate all combinations using the exact approach. In our experiment we tested approximate methods as well. The parameters vector $\boldsymbol{p}_{\mathrm{r}}$ refers mainly to the genetic reduction, where the size of the population, mutation and cross-over probability can be determined.

3. Generating rules. Based on the discrete reducts $\boldsymbol{R}$, all structures of the following form are created:

If premises then decision

The premises part contains the set of conditions that must be met by the particular attributes to fire (activate) the rule. As the consequence, the decision $\boldsymbol{c}$ about the object category is made. The conditions check if the value of the selected attribute is above or below the particular threshold, separating neighbouring intervals. The *RS* produce large sets, with hundreds or even thousands of rules.

4. The rules are next used to make decision about the category of users from **V**. The classification is based on the voting mechanism, with multiple rules pointing at the same category. During the analysis, each rule is analysed. If the premises part is fulfilled, the rule is fired. After analysing all rules in the set all activated are grouped according to the categories they point at. The category supported by the maximum number of active rules is the output of the *RS*-based module.

The scheme from Fig. 1 is repeated multiple times for various sets of parameters $\boldsymbol{p_d}$ and $\boldsymbol{p_r}$ in searching for the minimal value of $e_s(\boldsymbol{V})$ [15,16]. Also, the problem in our research was relatively small number of available objects (compared to the number of all Internet users). Therefore the division of all objects into the learning and testing set was of the great concern. To make sure the knowledge extracted during the training process would be useful in the analysis of new users, the cross-validation procedure was applied. It consists in repeated dividing the original data set into $\boldsymbol{L}$ and $\boldsymbol{V}$, training and validating the module for the subsequent versions and averaging the result. The *Repeated-Random Sub-Sampling Validation* (*RRSSV*) was used. The following steps were implemented:

1. Random selection of n objects (without replacement) from the set $\boldsymbol{D_T}$ (containing all available participants) and storing them in the set $\boldsymbol{V}$. Remaining examples form the set $\boldsymbol{L}$. Here n is determined to cut the predefined fraction of objects from $\boldsymbol{D_T}$. For instance, if there are 1500 users in $\boldsymbol{D_T}$, n can be set to 500. This way one third objects belongs to $\boldsymbol{V}$ and 1000 – to $\boldsymbol{L}$.
2. Training and validating the RS-based module using the scheme from Fig. 1.
3. Repeating the procedure from step 1 k times, creating different $\boldsymbol{L}$ and $\boldsymbol{V}$ from $\boldsymbol{D_T}$.
4. Calculation of the averaged sample error (or its supplement) as the mean of the n results.

The alternative to *RRSSV* is the k-Fold validation, which is susceptible to the position of objects in the original decision table, therefore was not used here. We implemented $n = 5$ repeats of the training/validating procedures, which is the compromise between the time of computations and the quality of results.

Although the cross-validation assumes the selection of objects from $\boldsymbol{D_T}$ to $\boldsymbol{L}$ and $\boldsymbol{V}$ with the uniform distribution, we decided to additionally implement the probabilities depending on the distribution of subsequent categories in the original data set. Because the number of average-rating people was assumed twice as high as the number of under- and over-rating, the chance of selecting the former to $\boldsymbol{V}$ was 0.5, while the latter – 0.25. In the experimental section both approaches are compared. This is justified if the assumed fraction of under- and over-rating users in the real world is the same as in the analysed set.

5 Experimental Results

Data processing was implemented using Rosetta software [17]. The tests of the *RS*-based module included optimization of discretization $\boldsymbol{p_d}$ and reduction $\boldsymbol{p_r}$ parameters to minimize the sample error $\boldsymbol{e_s}$. Also, the significance of the particular attributes for the distinguishing between the separate categories was checked. The mean classification accuracy (supplement to (3)) was below 50 percent, which reveals low correlation between the demographic variables and the evaluation of webpages by users. Probably introduction of psychological variables would increase the number of correct classifications. The method of selecting objects to $\boldsymbol{L}$ and $\boldsymbol{V}$ has strong impact on the results. The uniform distribution of objects gives about 40 percent accuracy, while the distribution reflecting the frequency of subsequent categories in $\boldsymbol{D_T}$ gives

almost 48 percent. In contrast, the fully random category selection gives about 35 percent. The results for particular combinations of discretization /reduction algorithms are in Table 1. The abbreviation *US* stands for the *Uniform Selection* of objects to the validating set, while *NUS – Non-Uniform Selection*. Abbreviations *EFB9*, *EFB10*, *EFB12* stand for *equal frequency binning* discretisation of "birth year" variable into 9, 10 and 12 intervals respectively.

All discretization and reduction methods presented in section 4 were tested. Results confirm the conclusions from previous research [13] about the usefulness of particular methods. The most effective discretization approach is *EFB*, although the optimal number of intervals must be individually adjusted in each case. For the "birth year" attribute the number of intervals ensuring the best classification results varies between 8 and 12. The influence of the number of intervals on the outcome (assuming the genetic reduction) is in Fig. 2. Increasing the number of intervals increases the classification accuracy to some point, although their too large number leads to the similar effect as in the naïve discretization. Finding the optimal number of intervals is always the problem, requiring the repeated computations for the varying number of intervals.

Table 1. Classification accuracy for selected combinations of discretization/reduction algorithms

discretization / reduction	EFB9		EFB10		EFB12		Naive		Semi-naive	
CV Object selection	US	NUS	US	NUS	US	NUS	US	NUS	US	NUS
Genetic	38.31	46.26	40.0	45.57	40.48	47.50	39.27	45.30	31.08	37.18
Johnson	32.53	40.24	33.73	35.18	28.91	37.10	31.08	36.62	30.0	33.46
Holte	51.08	65.06	51.08	65.06	51.08	65.06	51.08	65.06	51.08	65.06

The optimal discretization/reduction combination is *EFB/genetic algorithm*. The *Holte's* reduction tends to generate majority of rules pointing at the most frequent category. Therefore although the percentage of correctly classified objects is not lower than for genetic and *Johnson's* reduction, it can't be used in practice. Various values of genetic reduction (i.e. crossover and mutation probabilities and the population size) were tested. Crossover probability was tested in range between 30 and 80 percent. Mutation probability was between 0.05 and 0.3. Analysed population size varied from 50 to 200. No changes in parameters modified the results. Increasing them causes longer optimization time, but does not lead to the improvement of the classification efficiency. This means the algorithm quickly finds the optimal configuration, but no improvement is possible, even after expanding the search space and introducing more solutions analysed simultaneously. In any case, the genetic algorithm is still supreme over other methods.

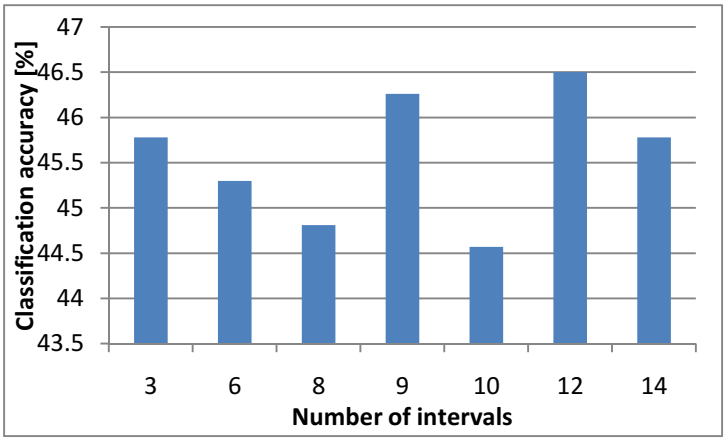

Fig. 2. Influence of the number of EFB intervals on the classification efficiency (genetic reduction, improved object selection for RRSSV)

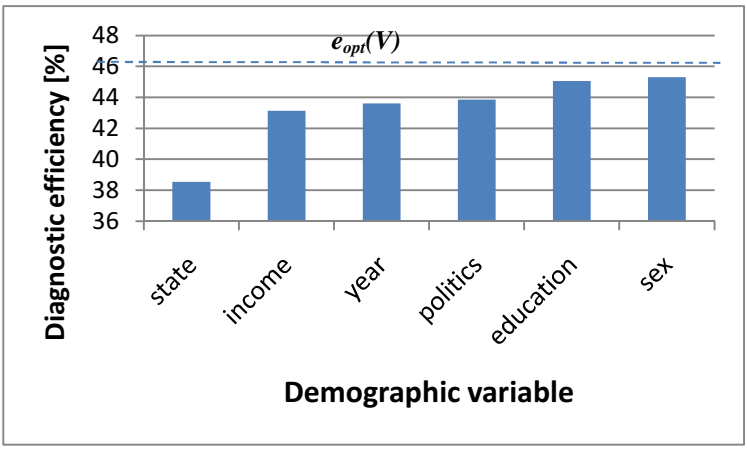

Fig. 3. Importance of the demographic variables for prediction of the user's web page credibility assessment

The mechanism of reducts generation allows for minimizing the number of attributes correlated with the categories. If the particular attribute is not present in any reduct, its elimination should not degrade the classification efficiency. All demographic variables appear to be important. The experiment consisted in eliminating the selected attribute from the set and checking the classification accuracy on the reduced table (using the genetic reduction method). Fig. 3 shows attributes in the descending order (the ones to the right being the least important). The greatest deterioration is for the set after eliminating the "state" attribute, while the smallest – for the "sex" variable. The dotted line determines value of the $e_{opt}(V)$, i.e. the maximum classification accuracy, obtained by the optimal combination of discretization and reduction methods.

6 Discussion

The rough set based prediction strongly depends on the applied discretization and reduction algorithms. Experiments show that finding dependencies between demographic variables and Internet user's tendency to *over*- or *under*-estimate webpage credibility is difficult. The obtained efficiency, although better than random category selection is still on unsatisfactory level. The solution to increase the classification accuracy would be the addition of psychological variables – probably better correlated with decision making presented in this paper. Also implemented strategy for the object selection to the validating set during the cross-validation gives promising results. However its usefulness is based on the assumption that category distribution in the analysed data-sets conforms to the actual distribution among Internet users. Results of conducted data analysis suggest that "education level" and "sex" are demographic variables most strongly influencing final results of *rough set* classifications of misjudgement tendencies in evaluating online content. We hypothesize that "education level" may be a factor connected with user's expertize, which strongly influences credibility evaluations [18]. It is likely that "education level" may be also connected with intelligence and in this sense have moderating effect on making judgements.

When it comes to "sex" we hypothesize that obtained result might indicate individual differences in psychological traits level. Differences between men and women in intensity of certain psychological traits is documented in literature. We suspect that level of risk taking [19, 20] or general trust [21] might be worth examining in the context of credibility evaluation in future studies. Therefore we prepare to replicate this study and include psychological measures together with expertise level. To sum up, application of artificial intelligence methods in the context of credibility evaluations so far have focused mainly on examining websites' characteristics. We suggest that in order to fully explain the phenomenon of trust in online content *human factor* should be included in such studies. In [22] it was shown that predicting personality traits from online behaviour (Facebook likes) can be successful. Using artificial intelligence methods to predict credibility evaluation misjudgement tendencies would be a powerful tool applicable in consumer studies or crowdsourcing systems.

References

1. Lucassen, T., Schraagen, M.: The inuence of source cues and topic familiarity on credibility evaluation. Computers in Human Behavior 29(4), 1387–1392 (2013)
2. Rimmer, T., Weaver, F.: Different questions, different answers? media use and media credibility. Journalism & Mass Communication Quarterly 64(1), 28–44 (1987)
3. Morris, M., Counts, S., Roseway, A., Ho, A., Schwarz, J.: Tweeting is believing?: understanding microblog credibility perceptions. In: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work, pp. 441–450. ACM (2012)
4. Fogg, B.J.: Prominence-interpretation theory: explaining how people assess credibility online. In: CHI 2003 Extended Abstracts on Human Factors in Computing Systems, pp. 722–723 (2003)

5. Metzger, M., Flanagin, A., Medders, R.: Social and heuristic approaches to credibility evaluation online. Journal of Communication 60(3), 413–439 (2010)
6. Kostakos, V.: Is the crowd's wisdom biased? a quantitative analysis of three online communities. In: Computational Science and Engineering International Conference, vol. 4, pp. 251–255. IEEE (2009)
7. Weijia, D., Jin, G., Lee, J., Luca, M.: Optimal Aggregation of Consumer Ratings: An Application to Yelp.com. Harvard Business School Working Paper 13, 42 (2012)
8. Salganik, M.J., Dodds, P.S., Watts, D.J.: Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market. Science 311(5762), 854–856 (2006)
9. Bolton, G., Greiner, B., Ockenfels, A.: Engineering trust: reciprocity in the production of reputation information. Management Science 59(2), 265–285 (2013)
10. Hvidsten, T.R., Komorowski, J.: Rough Sets in Bioinformatics. In: Peters, J.F., Skowron, A., Marek, V.W., Orłowska, E., Słowiński, R., Ziarko, W.P. (eds.) Transactions on Rough Sets VII. LNCS, vol. 4400, pp. 225–243. Springer, Heidelberg (2007)
11. Tay, E.H., Shen, L.: Economic and Financial Prediction Using Rough Sets Model. European Journal of Operational Research 141(3), 641–659 (2002)
12. Sikora, M., Sikora, B.: Rough Natural Hazards Monitoring, Rough Sets: Selected Methods and Applications in Management and Engineering. In: Advanced Information and Knowledge Processing, pp. 163–179 (2012)
13. Bilski, P., Wojciechowski, J.: Rough-Sets-Based Reduction for Analog Systems Diagnostics. IEEE Transactions on Instrumentation and Measurement 60(3), 880–890 (2011)
14. Tan, S., Cheng, X., Xu, H.: An Efficient Global Optimization Approach for Rough Set Based Dimensionalty Reduction. International Journal of Innovative Computing, Information and Control 3(3), 725–736 (2007)
15. Pawlak, Z.: Rough sets. International Journal of Parallel Programming 11(5), 341–356 (1982)
16. Pawlak, Z., Grzymala-Busse, J., Slowinski, R., Ziarko, W.: Rough Sets. Communications of the ACM 38–11, 88–95 (1995)
17. Öhrn, A., Komorowski, J.: ROSETTA: a rough set toolkit for analysis of data. In: Proceedings of the Third Joint Annual Conference on Information Sciences, Durham, NC, pp. 403–407 (1997)
18. Metzger, M., Flanagin, A.: Credibility and trust of information in online environments: The use of cognitive heuristics. Journal of Pragmatics 59, 210–220 (2013)
19. Jianakoplos, N., Bernasek, A.: Are women more risk averse? Economic Inquiry 36, 620–630 (1998)
20. Dwyera, P., Gilkesonb, J., List, J.: Gender differences in revealed risk taking: evidence from mutual fund investors. Economics Letters 76(2), 151–158 (2002)
21. Buchana, N., Crosonb, R., Solnickc, S.: Trust and gender: An examination of behavior and beliefs in the Investment Game. Journal of Economic Behavior & Organization 68(3-4), 466–476 (2008)
22. Kosinski, M., Stillwell, D.J., Graepel, T.: Private traits and attributes are predictable from digital records of human behavior. Proceedings of the National Academy of Sciences (PNAS) (2013)