

# Authority estimation within social networks using regression analysis

Kristína Machová<sup>1</sup> · Jaroslav Štefaník<sup>1</sup>

Received: 27 April 2016 / Accepted: 7 September 2016 / Published online: 21 September 2016  
© The Author(s) 2016. This article is published with open access at Springerlink.com

**Abstract** This paper focuses on methods of machine learning, particularly on regression analysis to solve a problem of authority identification within social networks. Within this paper, linear, polynomial, and non-linear regression types were considered. The aim was to find an approximation of dependency of the authority value on variables representing parameters of the structure and particularly the content of selected web discussions. The approximation function can be used at first for computation of the authority value of a given discussant, at second, for discrimination of an authoritative discussant from non-authoritative contributors to the web discussion. This information is important for web users, who search for truthful and reliable information in the process of decision making about important things. The web users would like to be influenced by some credible professionals. The various regression methods were tested, particularly linear, polynomial, and non-linear regression models. The best solution was implemented in the Application for the Machine Authority Identification.

**Keywords** Authority identification · Social networks · Web mining · Linear regression · Non-linear regression · Web forums

## 1 Introduction

We live in the information era. A volume of information, which is discovered each day, is too large and too time-consuming to be processed by a human. Everybody from us needs sometimes an access to the relevant supporting information for our decision-making. To know the relevance of information we have found, we need information about sources of the obtained information and their credibility. In other words, it is important to know the sources, which are authoritative ones. A web forum discussion can be a repository of various kinds of useful information: facts, opinions, ideas, attitudes, and so on. However, useful information is mixed with non-useful or misleading information. Every web user can join the web discussion, but many of them have not sufficient experiences or theoretical knowledge about the discussed themes. The web discussion often contains an opinion spam and an information trash. Therefore, it is the matter of principal to search for authoritative discussants to let them influence our important decisions. And just the searching for an authority and its machine identification among all discussants of web forum is our challenge.

To achieve our main goal—machine authority identification, we had to do the following three steps:

1. To find such variables—parameters of the structure and content of the web discussion, which are the most related to the authoritative contributing.
2. To define a dependency of the variable “authority” of a web discussion on the independent variables selected in the first step. We tried to find an approximation of this dependency using the linear and non-linear regression [1] based on the method of the ordinary least squares (OLSs) [2].

---

✉ Kristína Machová  
kristina.machova@tuke.sk  
<http://people.tuke.sk/kristina.machova/>  
Jaroslav Štefaník  
jaroslav.stefanik@student.tuke.sk

<sup>1</sup> Department of Cybernetics and Artificial Intelligence, Technical University, Letná 9, 04200 Košice, Slovakia

3. To use this approximation function for the discrimination of the authoritative from non-authoritative contributors to the web discussion.

Before starting the machine authority identification, we had to solve a number of technical problems. The first one was the automatic extraction of the conversation content and structure from the web page with the web discussion. The second one was to extract the values of selected independent variables from previously obtained information about the discussion. Another problem was how to obtain the values of dependent variable “authority” for regression function training. We decided for two alternative ways—to obtain values of “authority” from human “expert” and to extract them directly from the web discussion as so-called “wisdom of a crowd.”

Finally, all learned results were tested using widely used measures of the efficiency—the precision and the recall. The best solution was implemented within the Application for the Machine Authority Identification (AMAI).

## 2 Authority and web discussion

### 2.1 Web discussion group

Our attention was on an authority of a web discussion forum. The discussion group was developed in the society Usenet from the beginning of 80th years of twentieth century [3]. Two computer specialists Jim Ellis and Truscott have come with a new idea to create a system of rules for the contributions creation. Nowadays, WWW society becomes the main organization, which supports and spreads various platforms for Internet discussion groups using various settings up of different web servers. The internet discussion is represented by a web page, where users insert their contributions (opinions and reactions). Within this paper, the web users joining a web discussion will be called the contributors or discussants. They add their opinions, ideas, and attitudes to the web discussion, and in this way, they create the so-called “conversational content.” The authority identification represents the mining of this conversational content and its internal structure. There are different types of Internet discussion forums according to their scope [4]:

- *Discussion connected to a web article* In this case, the discussion is only an additional function to the content of the article to enable feedback. The subject of such discussion can be the text of the article or the concerning theme or product.
- *Guestbook* A place on a web site dedicated to reactions on the given web project, for example, a personal web page or a web page relevant to some theme.

- *Discussion forum* A part of a more extensive discussion project. It enables to establish new discussion pages and to sectionalize them into groups according to themes. It is a place, where users can leave the contributions. These contributions (e.g., news) are often longer than one line as they are within chatrooms. They are temporarily archived. An approval of new contributions by moderator of the discussion can be desired before the contributions became visible for all users.
- *Questions and answers* Some public institutions offer the public answering of question, suggestions or complaints on special web pages. In this case, a pronouncement of a responsible representative of the institution is expected.

There are many other social web platforms, where the conversational content is cumulated, for example, chats, Internet Relay Chat, blog and micro-blog platforms, and so on. However, the paper focuses on the web discussions dedicated to some given theme.

### 2.2 Authority identification in general

The concept “authority” comes from the Latin word “augere.” It denotes a person, whose opinions, attitudes or decisions are respected by other members of the group and whose decisions and advices are expected by other members of the group. The authority is derived from the relations between people (web users), positions, and hierarchies [5]. There are many kinds of authorities. For example, according to prestige, authority can be:

- *Formal (functional) authority* It represents a measure of influence of some person following from his formal position regardless of personal properties. It is leadership of the person, who is mandated to make decisions. It is obviously the result of a position, title or function of some person within an organization (an arbiter, teacher, politician, and so on). A leader could require submission, although this person is not honest or brave or predictable or able of a quick decision-making.
- *Informal (natural) authority* It is based on human and personal properties and professional assumptions of a person. It is the result of a personal profile, capability, adequate self-confidence, and social activities. Such person has natural, spontaneous influence on others, because of his/her persuasiveness and good experiences with his/her advices and decisions. The people, who let an authority to lead them, enforce the weight of this authority.

An important characteristic of the authority is using of no pressure and no force. The process of obtaining an authority is

demanding and time-consuming; nevertheless, more difficult thing is to maintaining it. The formal authority can be at the same time the informal one. The formal authority can sometimes change his status to informal and vice versa.

### 2.3 Authority of a web discussion

The virtual web authority has different characteristics as the authority in real life. It is related to the structure of the web, which is based on hyperlinks among web pages. The Google has discovered very complicated relations among web pages and references. Well-known tool for the web page authority calculation is PageRank [6]. Other known approaches to the web page authority calculating are hyperlink-induced topic search (HITS) algorithm [7] and stochastic approach for link-structure analysis (SALSA) [8]. These approaches are also based on an input and output hyperlinks of the evaluated web page. There are also tools of the respected portal “Seomoz,” for example, MozTrust (Moz’s global link trust score) [9] and Open Site Explorer [10]. All these tools cannot be easily used for calculating of an authority of the web discussion forum. There is also interesting work [11] concerned on a qualitative analysis of discussion forum. However, this work has not the aim to estimate the value of Authority of web discussant.

The authority identification from web discussion forums is a similar problem as web page authority calculation, because authority identification from web discussion is concentrated on web page, the discussion runs on. On the other hand, it is also a different problem, because no input or output references between this page and other pages are considered. Only references inside this page between various discussants are considered. These references are represented by reactions on contributions. All mentioned methods (PageRank, HITS, SALSA, MozTrust, and Open Site Explorer) calculate authority of each web page separately. One page leads to one measure of authority. Within the authority mining from the web conversation, not only one but all contributions of the given discussant are evaluated. All information about all contributions related to one discussant has to be concentrated and used for the authority estimation. Nevertheless, we can inspire ourselves by these techniques and take into account the number of references as reactions on an actual contribution.

In our previous work [12], we have taken into account mentioned number of reactions on all contributions of evaluated discussant, but also the number of all contributions of this discussant, the number of reaction of the discussant on the bottom level of the conversation tree (Fig. 4), the polarity matching between opinion of the discussant and opinion of all discussion, the positions of contributions in the conversation tree and the length of his/her contributions. Some of these variables have appeared to be not so important for the precise estimation of the authority. Another problem of

this approach was in way of the estimation function generation. For these reasons, we decided to modify the set of variables—arguments of the conversational structure and to use the regression methods for training the authority estimation function.

## 3 Used methods

We tried to solve the problem of the authority estimation within the web discussion forum using a machine learning method based on regression analysis. Within scientific works, there is often quantitative evaluation of two or more variables (for example,  $x$  and  $y$ ) and a function relation  $f$  among these variables has to be determined. There is a mutual statistical correlation  $z$  between these variables, as can be modeled by:

$$y = f(x), \quad z = \varphi(y, x). \quad (1)$$

The regression analysis can be:

- A simple regression, which is represented in (1),
- A multiple regression—we are searching for a relation of one dependent variable ( $y$ ) on a set of independent variables ( $x_1, x_2, \dots, x_n$ )—see Eq. (2). These independent variables are called “regressors” or “predictors”:

$$y = f(x_1, x_2, \dots, x_n). \quad (2)$$

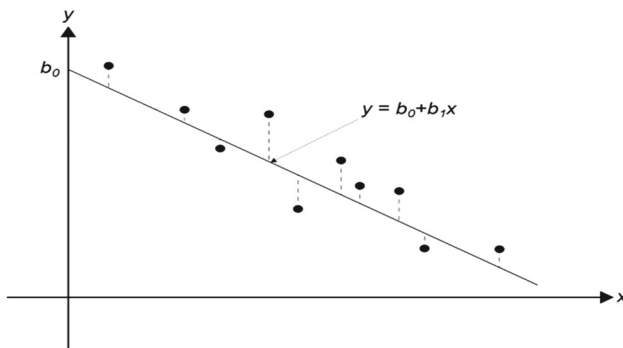
Within the regression analysis, it is very important to realize, which one of variables is dependent and which are independent. The goal of the regression analysis is to describe the relation between a dependent variable ( $y$ ) and independent variables ( $x_1, x_2, \dots, x_n$ ) by a suitable mathematical model, for example by linear or non-linear function. The result will be a regression curve, which should optimally match the empirical polygon [13].

### 3.1 Linear and non-linear regression

A regression function can be considered as a linear function in the case, when it is a linear function of the unknown parameters. Some examples of linear regression functions are as follows:

- $y = b_0 + b_1x$  (basic linear regression)
- $y = b_0 + b_1x + b_2x^2$  (quadratic regression)
- $y = b_0 + b_1x + \dots + b_r x^r$  (polynomial regression)

All these equations represent linear regression, because any unknown constants are not in the exponent. They are linear from the point of view of the regression analysis. Within



**Fig. 1** Linear regression in two-dimensional space [14]

the problem of the authority estimation, we have used the basic linear and polynomial regressions. The basic linear regression for two-dimensional space is shown in Fig. 1.

The goal is to find such values of constants  $b_0, b_1, \dots, b_n$  in the formula (3) (in two-dimensional space  $b_0$  and  $b_1$  of the linear line, see Fig. 1) to achieve the optimal matching between the linear line and the point graph, which consists of  $m$  points (observations). These constants can be dedicated from the point estimation using the ordinary least squares (OLS) method [2]:

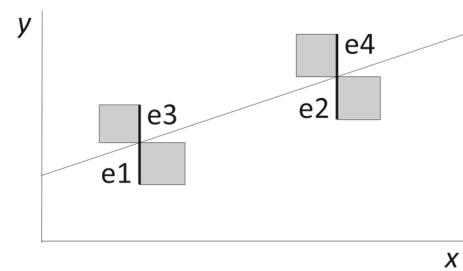
$$y_i = b_0 + b_1 x_{i1} + \dots + b_n x_{in} + \varepsilon_i. \quad (3)$$

Sometimes, it is not possible to find a satisfactory precise linear relation. In this case, the relation can be modeled by some non-linear function, the most frequently exponential function ( $y = be^{cx}$ ) or logarithmic function ( $y = b_0 + b_1 \ln x$ ) [1]. Within the problem of the authority estimation, we have used the non-linear modification of the polynomial function in the form  $y = b_0 + b_1 x^{c_1} + \dots + b_n x^{c_n}$ . In this function, not only constants  $b_i$  but also exponents  $c_i$  represent the searched constants. It is a more general form of the polynomial function with parameters in its exponents, where exponents need not to be integer values.

### 3.2 Ordinary least squares method

This OLS method belongs to mathematical and statistical methods. Through it, it is possible to solve the tasks of both types of regressions, linear and non-linear. In general, the method minimizes the sum of square errors (see Fig. 2). The sum arises when the differences between theoretical and empirical values exist. The theoretical values are calculated using the regression function, and the empirical values are obtained by a measurement or by an observation [2].

At first, values of parameters  $\hat{a}_0$  ( $b_0$  in Sect. 3.1) and  $\hat{a}_1$  ( $b_1$  in Sect. 3.1) in two-dimensional space are found. These values represent a point estimations of parameters. For these parameters, a residual sum of squares SSE (sum of square errors) is calculated according to (4). This sum is shown by



**Fig. 2** The sum of square errors within OLS method in two-dimensional space [14]

gray squares in Fig. 2. The basic principle of the OLS method is the minimization of this sum:

$$SSE(\hat{a}_0, \hat{a}_1) = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \hat{a}_0 - \hat{a}_1 x_i)^2. \quad (4)$$

After all necessary operations, the parameters  $\hat{a}_0, \hat{a}_1$  are calculated according to the following:

$$\hat{a}_1 = \frac{n \cdot \sum_{i=1}^n x_i \cdot y_i - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i}{n \cdot \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}, \quad (5)$$

$$\hat{a}_0 = \frac{\sum_{i=1}^n y_i - \hat{a}_1 \cdot \sum_{i=1}^n x_i}{n} = \bar{y} - \hat{a}_1 \cdot \bar{x}. \quad (6)$$

### 3.3 Specification of the variables of a discussion structure

We have selected 120 discussants from the portal “<http://www.sme.sk>”. Consequently, the following variables for each discussant were extracted from all his contributions:

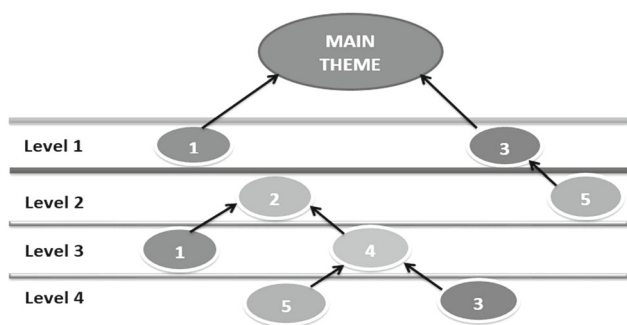
- *AE* Average evaluation of the contribution.
- *K* Value of the karma of the user, which is the contribution author.
- *NCH* Number of characters within his/her contributions.
- *AL* Average layer in the conversation tree (see Fig. 4).
- *ANR* Average number of reactions on his/her contributions.
- *NC* Number of contributions of given discussant.

These variables were used to form the training set (is shown in Fig. 3) for selected regression method.

*Average evaluation of the contribution (AE)* is represented by the ratio of the sum of all reactions [agree (+) and disagree (-)] on the contributions of given discussant to the number of all his contributions. This average evaluation is available on the web discussion page. The range of the *AE* is the number from 0 to 80.

Nickname	AE	K	NCH	AL	ANR	NC
Peter 2	60	108	26,0	0	1	1
V12	80	182	220,0	2	0,5	2
fero	80	171	548,5	3	2,5	2
sandokan555	80	162	57,5	4	0,5	2
Peter_5	50	99	112,5	6	0	2
darkman	80	167	117,0	3	0	1
Jesse Pickman	40	74	210,5	1,5	1,5	2
mm	60	108	22,0	1	1	1

**Fig. 3** Each line of the training set represents one discussant and contains the values of variables *AE*, *K*, *NCH*, *AL*, *ANR*, and *NC*



**Fig. 4** The conversation tree has four levels, the main theme is in the root and reactions are situated on levels 1–4, and all reactions of the same discussant have the same tint of the gray color

Value of *karma* (*K*) of the discussant is also available on the discussion web page. The karma is a number from 0 to 200, which represents activity of the discussant from last 3 months (within the portal “<http://www.sme.sk>”).

*Number of characters* (*NCH*) represents the average length of discussant contributions. It penalized authors with too short and so less informative contributions. We assume that an authoritative contributor does not insert extremely short contributions.

*Average layer* (*AL*) in the conversation tree (see Fig. 4) is the average number of all layers, which the contributions of the discussant are situated in. The conversation tree is a graphical representation of the web discussion. The *AL* represents the information, when the discussant joined the discussion, from the beginning or at the end.

*Average number of reactions* (*ANR*) on the all contributions of the given discussant is the number of reactions per one his contribution.

*Number of contributions* (*NC*) is simply the whole number of contributions of the given discussant. The parameter *NC* penalizes authors, which join a discussion rarely.

**Table 1** Average deviation of four versions of authority estimation function

Version	Average deviation
L-EXPERT	17.3489
L-CROWD	3.2998
PL-EXPERT	24.0123
PL-CROWD	8.7912
NL-EXPERT	18.1131
NL-CROWD	6.5618

All these parameters, taking separately, indicate rather chatty contributors than authoritative ones. However, taking them together as one entity, the emergency phenomenon arises. This phenomenon can indicate the authoritative contributors.

It may happen that a good contribution of already well-known authority finishes the discussion on the Web. It is truth that in such a case, there is no reaction on this contribution. It does not disturb the measure of the authority, because of high probability that there were more previous contributions of this contributor with many reactions within the given discussion. These reactions can balance the lack of reactions on the finishing contribution.

All these variables were considered to be independent variables. The dependent variable of the regression function *Y* was dedicated from:

1. Evaluation of each discussant by “human expert”;
2. Evaluation of each discussant by other discussants and it represents “wisdoms of the crowd.”

### 4 Implementation and testing

The authority value  $A \equiv Y$  was estimated by a linear and non-linear function of selected variables (*AE*, *K*, *NCH*, *AL*, *ANR*, and *NC*). The six regression functions for authority estimation were generated in the process of machine learning:

1. Linear function learned from the “human expert” (L-EXPERT) is represented by:

$$A = 0.4383AE + 0.0746K + 0.0281NCH - 2.1932AL - 3.4386ANR + 8.0102NC \tag{7}$$

2. Linear function learned from the “wisdoms of the crowd” (L-CROWD) is represented by:

$$A = 0.4385AE + 0.325K + 0.002NCH - 0.2928AL - 0.0853ANR + 1.0728NC \tag{8}$$



**Table 2** Values of precision and recall of six versions of regression functions were obtained in the three-time cross validation

Test	Version	Precision		Recall	
		Expert	Crowd	Expert	Crowd
Cross val. 12_3	Linear regression	0.78	0.99	0.69	0.99
	Polynomial regression	0.77	0.84	0.65	0.97
	Non-linear regression	0.72	0.99	0.66	0.88
Cross val. 13_2	Linear regression	0.65	0.98	0.65	0.93
	Polynomial regression	0.63	0.77	0.60	0.91
	Non-linear regression	0.67	0.97	0.67	0.86
Cross val. 23_1	Linear regression	0.68	0.97	0.67	0.67
	Polynomial regression	0.62	0.72	0.58	0.95
	Non-linear regression	0.69	0.97	0.69	0.67
Average	Linear regression	0.70	0.98	0.67	0.80
	Polynomial regression	0.67	0.78	0.61	0.94
	Non-linear regression	0.67	0.97	0.67	0.80

3. Polynomial function learned from the “human expert” (PL-EXPERT) is represented by:

$$A = 0.0001AE^3 - 0.0004K^2 + 0.0303NCH - 1.5539AL - 2.0557ANR + 12.1589NC \quad (9)$$

4. Polynomial function learned from the “wisdoms of the crowd” (PL-CROWD) is represented by:

$$A = 0.0001AE^3 - 0.0009K^2 + 0.0043NCH + 0.7473AL + 1.9875ANR + 6.7507NC \quad (10)$$

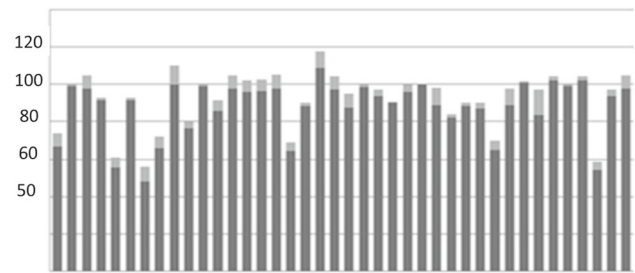
5. Non-linear function learned from the “human expert” (NL-EXPERT) is represented by:

$$A = 0.0382AE^{1.7192} - 0.3295K^{0.959} + 0.4470NCH^{0.681} + 0.1825AL^{0.0001} - 0.6269ANR^{3.2394} + 20.2509NC^{0.2977} \quad (11)$$

6. Non-linear function learned from the “wisdoms of the crowd” (NL-CROWD) is represented by:

$$A = 0.0185AE^{1.8135} + 141.5704K^{-78.39} + 0.0018NCH^{1.0457} - 0.0011AL^{3.7717} - 0.5562ANR^{0.0001} + 37.6642NC^{0.0038} \quad (12)$$

All these functions were created using standard MATLAB functions: “regress” in the case of linear and “lsqnonlin” in the case of non-linear relations. No auxiliary regularization method was used, because the input data matrix was regular. The input data can hardly be considered as noise data obtained, for example, from a device. These used input data map the structure of the given web discussion using defined variables. In the case of non-linear regression, also exponential parameters were elicited from the training data using the function “lsqnonlin.” It solves non-linear least-squares (non-linear data-fitting) problems and uses numerical optimization method “Trust-Region-Reflective Least Squares Algorithm.”



**Fig. 5** The illustration of estimated values of authority of some particular contributors (dark gray color column for each contributor) and deviations (light gray color) for some of tested discussants for the best version L-CROWD. The authority value can be from 0 to 120 (it is the range of  $Y$  values)

The default settings were used, and only the number of iterations was extended.

All the versions of the regression function for authority estimation (from (7) to (12)) were tested. The concise results of these tests are shown in Tables 1 and 2.

At first, the average deviations were calculated. According to the results in Table 1, the better functions were obtained by learning from the “crowd” than by learning from the “expert”. The deviations for some of tested discussants for the best version L-CROWD are shown in Fig. 5.

At second, these six versions of regression function were tested using obvious measures of a machine learning efficiency: precision and recall. The regression problem, when the value of  $A$  (authority) attribute should be estimated from the interval  $(0, 100)$  using formulas (7)–(12), was adopted to classification problem in the following way. A threshold  $T$  has been stated experimentally ( $T = 70$ ) and discussants were classified into categories: “authority” and “non-authority”. The discussants were classified to the class “authority” when their value of  $A$  was equal to or greater than

$T$  and they were classified to the class “non-authority” when their value of  $A$  was smaller than  $T$ . The precision  $\pi$  and recall  $\rho$  were calculated according to the following equations:

$$\pi_j = \frac{TP_j}{TP_j + FP_j}, \quad (13)$$

$$\rho_j = \frac{TP_j}{TP_j + FN_j}, \quad (14)$$

where TP is the number of true positives [the method classifies these examples as positive (authority) and they are truly positive according to the expert’s (crowd’s) opinion]. FP is the number of false positives [the method classifies these examples as positive (authority), but they are not positive according to the expert’s or crowd’s opinion]. FN is the number of false negatives [the method classifies the examples as negative (non-authority), but they are positive according to the expert’s (crowd’s) opinion].

Some key and the most important achieved results of tests are presented in Table 2.

The linear regression learned from the “crowd,” with the best test results, was implemented in the Application for the Machine Authority Identification (AMAI). This application provides the list of all discussants with the actual value of their authority. The AMAI also displays the value of the authority of the discussant, which was selected by a user. This value is from the interval  $(0, 100)$ . The application provides not only the binary decision whether the discussant is or is not the authority, but also it provides a precise numeric value of its authority.

## 5 Conclusions

The design of solving the problem of the authority identification from conversational content using the linear and non-linear regression was presented. The measure of the authority  $A$  was estimated as dependency on variables (AE, K, NCH, AL, ANR, and NC)—parameters of the structure and content of given web discussions. Another parameter which could be relevant is the vocabulary of the discussant—the literary language with scientific concepts can determine the level of the discussant and thus indicate the authority of the writer. In addition, the type of emoticons used in the discussion could be helpful. On the contrary, the dirty language may reflect a low level of the discussant. This language can be identified using some prepared special dictionary of dirty words. We would like to involve these parameters in the future.

The six generated estimation functions were tested. According to the values of average deviations (see Table 1), the best solution is the linear function learned from crowd (L-

CROWD). The second one is the non-linear function learned for crowd (NL-CROWD). Linear and non-linear functions learned from a single human evaluator—expert—seem to be worse. The same conclusions can be deduced from the resulting average values of precision and recall in Table 2. It is surprising that the linear model is better than the higher order model. It can be caused by a character of input data—parameters of the web discussion. Together with an increasing of values of these parameters, also the value of authority increases. Therefore, linear model is sufficient for the authority estimation.

It can be hardly said who is the expert on the authority identification. In addition, an opinion of a psychologist may be also subjective. On the other hand, combined opinion of many discussants can be objective.

There are other existing authority identification methods, as Klout, TwentyFeet, My Web Carrer [14], and our previous work [12]. All these methods use formulas for authority estimation, but these formulas were generated more experimentally without considering a theoretically based way. For this reason, we tried to generate the relation between the authority and the structure of web discussion using the classic mathematical approach based on the linear and non-linear regressions. For the future, we plan to elicit the constants of linear and non-linear equations using evolutionary algorithms [15, 16] to calculate not only constant values but the form of a non-linear regression function as well.

The presented approach can be used also in weighted opinion analysis of some discussions on social networks. Within the classic opinion analysis, the whole discussion is recognized to be positive (or negative) when there are more positive (or negative) contributions within it and each contribution has the same weight within determination of the summarized opinion. The weighted opinion analysis could multiply the measure of positivity of a given contribution with the weight represented by the estimated authority value of the contributor, who is the author of the given contribution. Thus, the opinions of authoritative contributors would have greater influence on the summarized opinion. We would like to apply the weighted opinion analysis in a domain of recognizing personality aberration from written text [17]. The designed approach and its implementation can be used to solve the problem of the decreasing of a web user cognitive load [18].

**Acknowledgements** The work presented in this paper was supported by the Slovak Grant Agency of the Ministry of Education and Academy of Science of the Slovak Republic under VEGA Grant No. 1/0493/16.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Pazman, A., Lacko V.: Lectures from Regression Models (in Slovak), vol. 132. University of Comenius Bratislava, Bratislava, Slovakia (2012). ISBN: 978-80-223-3070-1
- Pohlman, J.T., Leitner, D.W.: A comparison of ordinary least squares and logic regression. *Ohio J. Sci.* **103**(5), 118–125 (2003)
- What is Usenet?. <http://www.usenet.org>. (2016). Accessed 15 Sept 2016
- Machová, K., Penzész, T.: Extraction of web discussion texts for opinion analysis. In: IEEE 10th jubilee international symposium on applied machine intelligence and informatics, SAMI 2012, Herl'any, 26–28 January 2012, pp. 31–35. Óbuda University, Budapest, Hungary (2012) (ISBN 978-1-4577-0195-5)
- Chavalkova, K.: Authority of ateacher (in Czech). Philosophic-faculty of the University of Pardubice, Pardubice, Czech republic (2011)
- Fiala, D.: Time-aware PageRunk for bibliographic networks. *J. Infometrics* **6**(3), 370–388 (2012)
- Li, L., Shang, Y., Zhang, W.: Improvement of HITS-based algorithms on web documents. In: 11th International Conference on the WWW, pp. 527–535. ACM, Hawaii, USA (2002)
- Lempel, R., Moran, S.: The stochastic approach for link structure analysis (SALSA) and the TKC effect. *Comput. Netw. Int. J. Comput. Telecommun. Netw.* **33**(1–6), 387–401 (2000)
- Hallur, A.: MozRunk and MozTrust: everything you should know. <http://www.gobloggingtips.com/mozrank-and-moztrust/> (2016). Accessed 15 Sept 2016
- Fishkin, R.: Open site explorer news link building opportunity section (2016). <http://moz.com/blog/open-site-explorers-new-link-building-opportunities-section>. Accessed 20 April 2016
- Azevedo, B.F.T., Behar, P.A., Reategui, E.B.: Qualitative analysis of discussion forums. *Int. J. Comput. Inf. Syst. Ind. Manag. Appl.* **3**, 671–678 (2011). ISSN: 2150-7988
- Machová, K., Sendek, M.: Authoritative authors mining within web discussion forums. In: 9th International Conference on Systems, pp. 154–159. International Academy, Research and Industry Association, Nice, France (2014)
- Introduction to regress analysis (in Czech) (2016). [http://www.statsoft.cz/file1/PDF/newsletter/2014\\_26\\_03\\_StatSoft\\_Uvod\\_do\\_regresni\\_analyzy.pdf](http://www.statsoft.cz/file1/PDF/newsletter/2014_26_03_StatSoft_Uvod_do_regresni_analyzy.pdf). Accessed 20 April 2016
- Štefaník, J.: Approximation of the relation of an authority on the parameters of the structure of web discussion(in Slovak). Technical University of Košice, Košice, Slovakia (2015)
- Mach, M.: Evolution algorithms—problems solving (in Slovak). FEI Technical University, Košice, p. 135 (2013). ISBN: 978-80-553-1445-7
- Čádrík, T., Mach, M.: Evolution classifier systems (in Slovak). Electrical Engineering and Informatics IV. In: Proc. of the FEI Technical University of Košice, Košice, pp. 168–172 (2013). ISBN: 978-80-553-1440-2
- Šaloun, P., Ondrejka, A., Malčík, M.: Personality disorders identification in written texts. In: International Conference on Advanced Engineering Theory and Applications, Ho Chi Minh City, Lecture Notes in Electrical Engineering, vol. 371, no. 1, pp. 143–154. Springer, New York (2016). ISBN: 978-3-319-27245-0, ISSN: 1876-1100
- Machová, K., Klimko, I.: Classification and clustering methods in the decreasing of the internet cognitive load. *Acta Elektrotech. et Inf.* vol. 6, no.2, pp. 52–56. FEI TU Košice (2006). ISSN: 1335-8243