A Modified Scrum Story Points Estimation Method Based on Fuzzy Logic Approach

S. A. Semenkovich <sofya-semenkovich@yandex.ru>
O. I. Kolekonova <okolekonova@gmail.com>
K. Y. Degtiarev <kdegtiarev@hse.ru>
National Research University Higher School of Economics (HSE),
3, Kochnovsky Proezd, Moscow, 125319, Russian Federation

Abstract. Several known methods allow to estimate the overall effort(s) to be used up for the software development. The approach based on story points is preferable and quite common in the context of Scrum agile development methodology. However, it might be rather challenging for people, who are new to this methodology or to a specific Scrum team to estimate the amount of work with story points. The proposed approach involves estimation of features on the basis of linguistic terms that are both habitual and clear for everyone. The presented fuzzy inference system (Mamdani's model) makes it possible to calculate story points using people's opinions expressed as sentences in natural language – the study shows empirically that beginners to Scrum methodology consider the proposed approach to be more convenient and easier in use than the 'plain' story points estimation. Also, four groups of people with different levels of qualification in Scrum were asked to estimate several features of a certain project using the developed approach and common story points approach to prove the relevance of the approach – it was shown that the results of basic story points estimation for Scrum experts differ slightly from the results revealed by proposed approach, while for Scrum beginners such difference is significant. To the opinion of authors, the proposed approach may allow to adapt to Scrum more smoothly, with better understanding of what is implied by story points, grasping the general idea and learning faster their use in practice. The experimental study conducted as a part of the research has shown results approaching the estimations provided by Scrum experts who have been working in real projects and making use of story points for several years. Continuation of the present work can be associated with intensive studies of more complicated methods of aggregation of the experts' opinions, analysis of alternative representation forms of confidence degrees in estimates provided as well as the development of plugin for JIRA tracking system.

Keywords: fuzzy logic; Scrum; story points; expert estimations; aggregation of opinions; fuzzy inference system; Likert scale

DOI: 10.15514/ISPRAS-2017-29(5)-2

For citation: Semenkovich S.A., Kolekonova O.I., Degtiarev K.Y. A Modified Scrum Story Points Estimation Method Based on Fuzzy Logic Approach. Trudy ISP RAN/Proc. ISP RAS, vol. 29, issue 5, 2017, pp. 19-38. DOI: 10.15514/ISPRAS-2017-29(5)-2

1. Introduction

Many software systems relate to large-scaled and rather complex products that embrace, in particular, numerous factors to monitor and control at the development stage. Without a doubt, software development is a multifold process that essentially depends on tangled human activities, thus requiring effective management and planning [1]. Software development effort estimation acts as a key constituent of decision-making support during the process of such planning and further management. In short, effort can be defined in the context of combination «man-time» and expressed as the time (number of units) needed for a man (team's member) to complete a given task [1, 2]. Nowadays, we may address a relatively long list of recognized estimation methods aimed at evaluating efforts needed to be spent in the software development process. In fact, many efforts to categorize such methods are originating from the publications by Barry Boehm on software cost modeling and engineering economics in the early eighties of the previous century. We cannot talk about «the best» from all conceivable standpoints classification, but in rough outline such methods can be divided into three aggregative categories, namely: methods based on expert subjective estimates and views (non-model based methods), formal estimation methods that are grounded on specific or generic models, and combined (or, composite) methods built upon joint use of analysis and processing of available from different sources data along with expert estimates [3]. Amongst others, the first category takes in such approaches as planning poker (also known as Scrum poker) and Wideband Delphi, two similar methods where the provided estimations are based on judgments and expressed opinions of project's stakeholders [1]. In formal estimation models (e.g. Constructive Cost Model (COCOMO), COCOMO II as a generalization of COCOMO, weighted micro function points (WMFP), SLIM, use case modeling, story points) formulas and/or results derived from earlier implemented projects are used.

In the present paper, we consider the method of estimation with story points in the context of Scrum, an agile flexible framework to manage the process of software development. The main goal of Scrum is to deliver new software capability (features) every 1-2 weeks (the duration can be extended), each new version includes the most important features for Product Owner, thus allowing to inspect and adapt product to current conditions. The main Scrum characteristic of the estimation process is that Product Owner defines priorities for the features because the product should be maintained in a tested/integrated state every Sprint (i.e. fixed number of days team works together to produce beforehand coordinated changes in the product), so the work should be broken down to pieces/stages [4]. In case of proper compliance with other Agile principles, the release deadline cannot be missed by the team; if the features were evaluated incorrectly by some reasons, skipping over the less important tasks can be the only noticeable disadvantage as compared to waterfall or pseudo-Agile teams' experience.

In contrast with other approaches, Scrum is concerned with two main factors that are important in estimating development efforts. Firstly, the responsibility for the product falls on the shoulders of the whole team rather than individuals. It means that there are no gradations like «my work» and «your work». The framework attracts attention to cumulative effort(s) per Product Backlog' Item rather than individual effort(s) per feature. Secondly, the tasks are estimated in a relative manner, i.e. they are assessed (compared to each other) in terms of relative units, but not absolute ones. Thus, story points may be employed as such units of measure to express an estimate of the overall effort required to fully implement a product backlog item or any other piece of work [5]. As it is noticed by Joshua Kerievsky [6], "… Many say that story points make us better at estimating because we're estimating the size of work, rather than the time it takes to complete it; … in 2005, one of our customers found story points to be so confusing that he renamed them NUTs (Nebulous Units of Time)".

Such witty testimonial inherently expresses the attitude of newcomers to Scrum development methodology towards story points, their 'fear' of commonly used phrases and statements: «number of points per sprint», or «the estimate in story points is better than estimate in hours», etc. There are many helpful and well-composed electronic and printed sources dedicated to Scrum's set of principles and practices aimed at developing complex systems – books and articles by J. Sutherland, C. Sims, A. Stellman, guides, reports, tutorials on Scrum and other Agile methodologies by AgileRussia.ru, Scrum Alliance®, training courses from ScrumTrek, Scrum.org, LuxSoft, to name a few. Even cursory glance at results of Google search gives cause for being not fully confident indeed in the conception of various word-combinations related to story points enquiries, e.g. «they are cheaper than hours», «relative unit of measure», «estimate of effort», and the like.

In brief, story points are founded on "a short description of a set of features called user stories"; each such story will have a set of story points [7]. When we estimate features with story points, we assign a point value to each item. The raw values we assign are unimportant (we can talk about unit of measure that team's members agreed on), what matters are the relative values. A story that is assigned a value of 2 should require twice as much effort as a story that is assigned a value of 1, and it also constitutes two thirds of a story that is estimated at the level of 3 story points. Because story points represent the effort(s) to develop a story, a team's estimate must cover every aspect that can affect the effort. In general, they bring together as a single whole the amount of work to do, the complexity of the work, any risk or uncertainty in doing the work.

Our research proposes to simplify the process of estimating features with the help of story points. For most people it is rather confusing or even difficult to combine three aforesaid components into one in their mind and give an approximate resultant value. Instead of evaluating the features with story points, we assume that each member of the Scrum team (e.g. expert) provides his/her opinion regarding two factors, namely, these are the amount of work to do and its complexity. Besides, the experts should also specify the level (or, degrees) of their confidence in both such estimates. Experts operate with preset collection of linguistic terms expressed as words or phrases of the natural language. These verbal units are converted after that to proper fuzzy sets used in further processing. The latter provides application of fuzzy inference system (FIS) for each expert's estimations and aggregation of the results obtained into one outcome. What are the reasons to resort to the help of fuzzy approach? Well, we can partly refer to [8] saying that "many fuzzy categories described linguistically appear to be more informative than precise descriptions".

On top of that, a short survey was also conducted with the aim to figure out the opinions of four different groups of people on proposed approach. The core of this activity is the comparison of story points obtained in "experimental" manner and regular story points estimation.

The rest of the paper is organized as follows: section 2 presents basic definitions, terms (type-1 fuzzy set, linguistic variable, inference system, defuzzification, aggregation of estimates) that are used in the subsequent parts of the paper. The proposed approach to obtain story point-based estimates on the basis of defined input variables of Mamdani's fuzzy inference system (FIS) is discussed and visualized in section 3. The results of conducted experiment (empirical study) with several groups of people having different practical skills relative to use of story points estimations are discussed in the section 4. Concluding remarks are drawn in section 5.

2. Basic definitions and general comments

In clear majority of cases humans express their opinions and judgments using statements of natural language; many things that are thus heard or said are vague to a variable degree. According to Stanford Encyclopedia of Philosophy, a term «is vague to the extent that it has borderline cases», and the latter acquires special significance in relation to the vagueness that has to be modeled in adequate way for the case under consideration. In general, such task appears simple enough only at the first glance, and one of practical approaches, at least, from perception-based point of view, relates to fuzzy logic (FL) methodology. It provides ample means to model the perceived meaning of words/phrases conveying the experts opinions (estimates) in a graded fashion. Following seminal paper "Fuzzy Sets" by L.Zadeh [9], the concept of fuzzy set constitutes a class of objects with continuum membership grades.

Definition 1. Let U be a set of elements (objects) that are denoted generically as x (U={x}); fuzzy set $A \subseteq U$ is a set of ordered pairs $\{(x, \mu_A(x))\}$, where mapping $\mu_A : x \to [0,1]$ is a (type-1) membership function of a fuzzy set A. Value $\mu_A(x)$ is a degree (grade) of membership of x in the set A.

In many situations the shape of membership function can be set by a specialist (expert, domain engineer); such manual tuning of function's parameters turns out to be sufficient at the initial stages of model's development and processing. Thus, piecewise linear functions are often chosen due to their usability, expressive power in grasping thoroughly both the knowledge and human's perception of situation, as well as computational efficiency.

Definition 2. Trapezoidal membership function [10] is defined by a 4-tuple $\left(a_1, a_2, a_3, a_4\right)$ of its parameters in the following way:

$$\mu_A(x) = \begin{cases} 0, \ x \in (-\infty, a_1) \\ (x - a_1)/(a_2 - a_1), \ x \in [a_1, a_2] \\ 1, \ x \in [a_1, a_2] \\ (a_4 - x)/(a_4 - a_3), \ x \in [a_3, a_4] \\ 0, \ x \in (a_4, +\infty) \end{cases} \qquad (1)$$

Normalized trapezoidal (and triangular with values $a_2 = a_3$) functions having height $h = \max(\mu_A(x)) = 1, \ \forall x \in U \subset \mathbb{R}^1$, often describe values in the form «close to *b*», «around *b*», where *b* is either a crisp real number $b_{val} \in \mathbb{R}^1$, or the interval $\left[b_{val}^{(1)}, b_{val}^{(2)}\right] \subset \mathbb{R}^1$.

Definition 3. A linguistic variable is characterized by a parameter vector (or, 5-tuple) $\left\langle L_v, T(L_v), U, R_{syn}, R_{sem} \right\rangle$, where parameter L_v is the name of the variable (e.g. $L_v \equiv$"complexity of work"), $T(L_v)$ is the set formed by labels of variable's L_v linguistic values $l_1,..,l_n$ (term-set of L_v ; e.g. 'easy', 'normal', 'difficult', etc.). These names are generated using syntactic rule R_{syn}, whereas the meaning $R_{sem}(l_i)$ is associated with each value l_i, $i = \overline{1, n}$, from $T(L_v)$ by means of semantic rule R_{sem} ; $R_{sem}(l_i)$ is a fuzzy set (respective membership function) defined on a universe of discourse U . Linguistic modifiers (they are also called hedges) 'very', 'more or less' and the like, together with logical connectives 'and', 'or' and negation 'not' are treated as special type operators that modify the primordial meaning of primary values (terms) $l_1,..,l_n$. It results in altered shape of membership functions representing $l_1^{mod},..,l_n^{mod}$ [11].

Definition 4. The fuzzy inference is a process of deriving conclusion from given premises and system's inputs (or, given fact), for which compositional rule of inference (CRI) serves as a core. CRI can be viewed as a generalization of modus ponens argument scheme (the mode that affirms). The premises are represented as a set of If-Then rules forming knowledge base Ω, e.g. If x is A$_i$ Then y is B$_i$, $i = \overline{1, m}$, as a basic case ($A_i \to B_i$, i.e. A_i implies B_i) – potentially, such rules may have more complex appearance. Mamdani-type fuzzy inference system (FIS) proposed and evolved by E.H. Mamdani and S. Assilian in 1975 owing to the examination of fuzzy logic controller can be expressed as $B'(y) = \bigcup_{x \in U_1} A'(x) \wedge \widetilde{R}(x, y)$, where relation $\widetilde{R}(x, y)$ is calculated as follows: $\widetilde{R}(x, y) = \bigcup_{i=1}^{m} A_i(x) \wedge B_i(y)$, where A_i and B_i are (type-1) fuzzy sets, $A_i \subset U_1$, $B_i \subset U_2$.

The knowledge base Ω represented as a set of If-Then rules constitutes rather convenient and transparent form to express individual expert conceptions of phenomenon under study as well as perceptions of a group of specialists. On the whole, model Ω is a handy tool to discuss hypotheses (under potential tuning up rules and initially set parameters of fuzzy sets, if needed) and to make final decisions.

The process of representing initial data (e.g. linguistic values) as membership functions is called fuzzification; most of applications require to perform at final stages the opposite translation from fuzzy functional forms to crisp values; the latter act as representatives of corresponding fuzzy sets. This is achieved through defuzzification procedures, and one of commonly utilized method is called Center Of Area (COA). It stipulates calculation of the resultant value res* by way of

$$res^* = \frac{\int_U x \cdot \mu(x)dx}{\int_U \mu(x)dx} \qquad (2)$$

The intersection (*AND*) and union (*OR*) operations that are used in computational schemes with fuzzy sets are expressed as functions called t-norms $T(\cdot)$ and s-norms $S(\cdot)$, accordingly [12]. Different types of $T(\cdot)$ and $S(\cdot)$ are presented and discussed at length in the literature (e.g. [13]) – without loss of generality, in the paper we use standard *min* and *max* operators:

$$\mu_{A \cap B}(x) = T(\mu_A(x), \mu_B(x)) = min(\mu_A(x), \mu_B(x)) \qquad (3)$$

$$\mu_{A \cup B}(x) = S(\mu_A(x), \mu_B(x)) = max(\mu_A(x), \mu_B(x)) \qquad (4)$$

It is worth noting that story points are crisp numbers, because they appear to be the most convenient and easy "units" to compare and interpret by Scrum team members as compared to, for instance, numeric intervals. Thus, crisp numbers are associated with story points, which help to rank features in compliance with efforts required to implement them. As it was mentioned before, the valuable source of information are expert judgments (estimations), and once all such estimations are obtained, they should be aggregated to form conjoint opinion. Such activity can be performed by a dedicated person called *analyst*. With this aim in mind, two methods of aggregation are used in the paper.

The first method of aggregation is applied when all estimations elicited from Scrum team members (experts) are different, with one minimum and one maximum denoting left and rights extremities in the resultant sequence. For example, if it is of a form 10, 25, 46, 34, 30, 47, 28, simple expression allows to calculate the aggregated estimate:

$$e_{\mathrm{agr}} = \left(\sum\nolimits_{\mathrm{i}\neq \mathrm{i}_{\min},\mathrm{i}_{\max}} e_{\mathrm{i}} - e_{\min} - e_{\max}\right)\Big/\left(\mathrm{n}_e - 2\right) \qquad (5)$$

where e_{agr} is the aggregated estimate, $e_{\min}$ and $e_{\max}$ are minimum and maximum values among obtained estimations, respectively, n_e is the total number of values in the sequence; summation goes over all estimations excluding $e_{\min}$ and $e_{\max}$.

The second aggregation method (weighted arithmetic mean) can be used in situation of appearance of recurring experts' estimations as in the case of values 10, 25, 10, 34, 25, 47, 28; such outcomes (with repetitions) are rather practicable, so they should be addressed reasonably enough. If R_e is the most recurring estimate (conditional mean) observed in the numeric sequence, then e_{agr} can be obtained as follows:

$$e_{\mathrm{agr}} = R_e - \left(\sum\nolimits_{\mathrm{i}}\left(\left(e_{\mathrm{i}} - R_e\right)\cdot f_{\mathrm{i}}\right)\right)\Big/\mathrm{n}_e \qquad (6)$$

where f_{i} is the frequency of e_{i} occurrence in the row of estimations provided.

All prepared comments allow to proceed to approach that may assist people who are new to Scrum methodology (or, they are newcomers to a specific Scrum team) and who do not fully understand how they can estimate the amount of work to do on the base of story points. The central idea of such approach relates to a natural course, i.e. story points seem brittle and a bit confusing – fine, try in that case to estimate how much certain part of work will take making good use of terms you are familiar with. The aforesaid definitions simplify the perception of the following material, and they should not be considered as an extra "difficulty" to tackle on top of Scrum methodology itself; «*such overload is a bit too thick!*» – the reader may exclaim. We think, in no way, as long as all necessary (not very complex) calculations can be done by analysts; in other respects, interviewing and grasping the verbal statements expressing the results (what is said) in pretty understandable form are natural and plain day-to-day human activities.

3. Expert opinions and levels of confidence – modified Likert scale and fuzzy approach

Suppose that through talks and consultations with experts, the analyst collected the opinions (estimations) of several experts on certain feature expressing how much work, reasoning from their understanding and perception, they'll have to do to implement this feature, complexity of the work and their level of confidence about each of these estimations. After fuzzification of verbal data obtained and applying fuzzy rules, the aggregated result is converted to story points; the latter can be used at subsequent stages in any project management system.

As it was already mentioned earlier, the expert puts his/her opinion concerning complexity, amount of work as well as degree of confidence in estimation expressed

in linguistic forms (statements) [14]. For example, the expert may say the following: «*I'm quite sure that this feature will be difficult to implement, besides I must do a large amount of work to implement this feature, however, I'm not very sure about it*». From this sentence, we can pick out the following pairs of linguistic terms, namely: 'difficult' → 'quite sure' and 'large' → 'not very sure'. With such estimations in mind (and their formal representation by way of fuzzy sets), we'll be able to proceed to the construction of corresponding fuzzy rules [15].

Table 1. Parameters of trapezoidal membership functions representing values of term-sets

The amount of work (set T(*A*)**)**	**The complexity of work (set** T(*C*)**)**	**The overall effort (set** T(*E*)**)**
value 'very small' (1,1,5,20)	value 'very easy' (1,1,5,20)	value 'tiny' (1,1,5,20)
value 'small' (5,15,30,40)	value 'easy' (5,15,30,40)	value 'little' (5,15,30,40)
value 'medium' (25,40,60,75)	value 'normal' (25,40,60,75)	value 'average' (25,40,60,75)
value 'large' (60,70,85,95)	value 'difficult' (60,70,85,95)	value 'big' (60,70,85,95)
value 'very large' (80,95,100,100)	value 'very difficult' (80,95,100,100)	value 'huge' (80,95,100,100)

It is commonly advised to use the interval [1,100] to represent story points estimations, so we direct our attention to the same extreme points 1 and 100 to define the universe U to specify fuzzy sets [4]. The amount of work to do, the complexity of the work and the degrees of confidence are considered as system's input variables, whereas the overall (combined) effort is taken as an output variable. Thus, the following linguistic variables $L_{\mathrm{v}}^{(\mathrm{i})}$ denoted as A, C and E and their values (labels of linguistic terms) are considered [11]: $L_{\mathrm{v}}^{(1)}{=}A \equiv \text{"amount of work to do"}$,
$L_{\mathrm{v}}^{(2)}{=}C \equiv \text{"complexity of work"}$ (Fig. 1),
$L_{\mathrm{v}}^{(3)}{=}E \equiv \text{"the overall (combined) effort"}$ (Fig. 2), where
$\mathrm{T}(A) = \{\text{'very small','small','medium','large','very large'}\}$,
$\mathrm{T}(C) = \{\text{'very easy','easy','normal','difficult','very difficult'}\}$,
$\mathrm{T}(E) = \{\text{'tiny','little','average','big','huge'}\}$.

Table 2. Correspondence between levels of confidence and their values

Level of confidence (linguistic term)	**Value**
'not sure at all'	0.05

'almost not sure'	0.15
'not very sure'	0.35
'more or less sure'	0.5
'sure'	0.65
'quite sure'	0.8
'definitely sure'	0.95
'extremely sure'	1

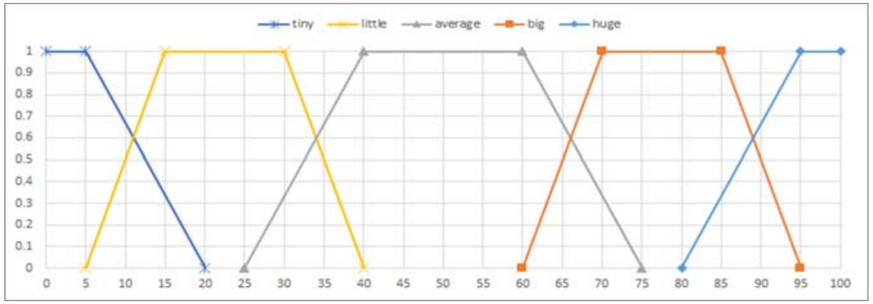

Fig. 1. Linguistic variable C = ″complexity of work″.

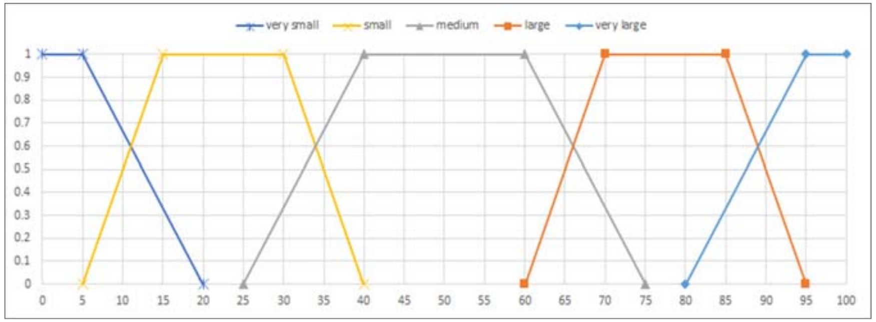

Fig. 2. Linguistic variable E = ″the overall (combined) effort″.

After consultations with experts, the analyst (and his group) defines the parameters of trapezoidal membership functions (1) to represent formally values of term-sets T(*A*), T(*C*) and T(*E*) fuzzy sets as shown in Table 1. For example, if expert says something like «*... this feature is hard to implement, but I must do small amount of work*», we select primary linguistic values 'small' from the set T(*A*) and 'difficult' – from T(*C*). The parameters of membership functions (Table 1) were chosen empirically, although slight alterations of values within certain bounds ($\pm\varepsilon_i$, $i=\overline{1,k}$, k is the number of deliberate assortments of such deviations on all terms of sets T(·)) turn out to be allowable. Such “mobility” of value ranges may bring to the advisability to consider further on type-2 interval fuzzy sets – unlike type-1 sets, they enable to express the uncertainty about the membership grades of elements on the domain considered.

Table 3. The accordance of the amount of work and the complexity of work to overall effort.

A \ *C*	**very easy**	**easy**	**normal**	**difficult**	**very difficult**
very small	tiny	tiny	little	average	average
small	tiny	little	little	average	average
medium	little	little	average	big	big
large	average	average	big	big	huge
very large	average	average	big	huge	huge

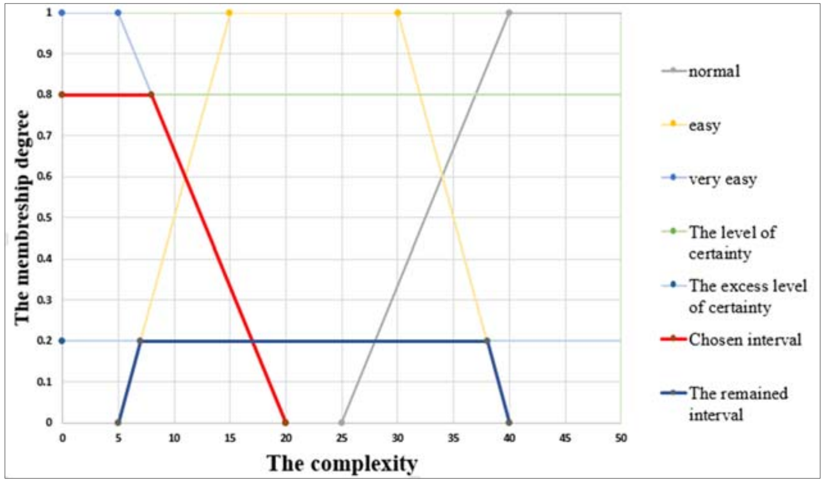

Fig. 3. The distribution of confidence levels (fuzzification stage).

The next step is to relate the level of confidence to fuzzy set being thought about. The ideas and views concerning Likert scale (psychometric response scale suggested by American sociologist Rensis Likert in 1932) allow to come out with relatively simple scheme to use in aforesaid task. Following QingLi, the level of agreement (LA) as an estimate within the range [0,1] can be associated with the membership degree (as an option, terms 'strongly agree', 'agree', 'neither agree, nor disagree', 'disagree' and 'strongly disagree' can be in use) [16]. The sum of LA for all options is equal to 1. In the case considered, the option provided by an expert and the level of agreement is experts’ levels of confidence are shown in Table 2. However, if expert’s level of confidence is not 'extremely sure', we are facing with the excess of LA. Thus, it can be suggested to distribute emergent excess between the nearest neighbors of the option selected by the expert. If there are two nearest neighbors, they both will get half of the excess observed; if there is only one nearest neighbor, it will get the whole amount of excess. In the paper, we use crisp numbers to represent level of confidence' values as the starting point of our approach. These values are based on the results of survey – opinions of approximately 50 people concerning the correspondence between linguistic values (labels) of confidence level and their actual mapped numbers were first elicited and averaged afterwards (see Table 2).

For example, if expert says that his/her level of confidence can be expressed as 'quite sure' (i.e. expert explains that «*... I'm quite sure that…*»), and the feature under consideration is very easy to implement, we choose fuzzy number representing term 'very easy' and define degree of membership as being equal to 0.8 – it is the value of choice. Thus, the excess level of confidence comes to 0.2, and it is handed over to the nearest neighbor of the term 'very easy', which is 'easy'. This distribution of confidence levels is shown graphically in Fig. 3.

Based on the information and knowledge elicited from experts, we may design a set of fuzzy rules (fuzzy rule-base). The amount of work to be done and the complexity of work act as input variables, and their combination result in the value of the overall effort. In general, these rules reflect the perceptions of experts, their feelings and conclusions drawn regarding situation given. For instance, a "typical" question may look as follows: «*How much will it take in the sense of overall effort to accomplish a 'very easy' task that needs just 'medium' amount of work to be done*». The short version of the rule-base is represented in Tab. 3, whereas the full set is provided below (rules Ri, $i=\overline{1,5}$). From the very outset, there were 25 rules (one rule for each combination of the amount of work (*A*) and the complexity of work (*C*)). Later, they were combined on the base of resulting value of overall effort, and only five rules R1,…,R5 were retained.

- rule R1:
IF amount is 'very small' ***AND*** complexity is 'very easy' ***OR***
amount is 'very small' ***AND*** complexity is 'easy' ***OR***
amount is 'small' ***AND*** complexity is 'very easy',
THEN effort is 'tiny'

- rule R2:
IF amount is 'very small' ***AND*** complexity is 'normal' ***OR***
amount is 'small' ***AND*** complexity is 'easy' ***OR***
amount is 'small' ***AND*** complexity is 'normal' ***OR***
amount is 'medium' ***AND*** complexity is 'very easy' ***OR***
amount is 'medium' ***AND*** complexity is 'easy',
THEN effort is 'little'

- rule R3:
IF amount is 'very small' ***AND*** complexity is 'difficult' ***OR***

amount is 'very small' ***AND*** complexity is 'very difficult' ***OR***
amount is 'small' ***AND*** complexity is 'difficult' ***OR***
amount is 'small' ***AND*** complexity is 'very difficult' ***OR***
amount is 'medium' ***AND*** complexity is 'normal' ***OR***
amount is 'large' ***AND*** complexity is 'very easy' ***OR***
amount is 'large' ***AND*** complexity is 'easy' ***OR***
amount is 'very large' ***AND*** complexity is 'very easy' ***OR***
amount is 'very large' ***AND*** complexity is 'easy',
THEN effort is 'average'

- rule R4:
IF amount is 'medium' ***AND*** complexity is 'difficult' ***OR***
amount is 'medium' ***AND*** complexity is 'very difficult' ***OR***
amount is 'large' ***AND*** complexity is 'normal' ***OR***
amount is 'large' ***AND*** complexity is 'difficult' ***OR***
amount is 'very large' ***AND*** complexity is 'normal',
THEN effort is 'big'

- rule R5:
IF amount is 'large' ***AND*** complexity is 'very difficult' ***OR***
amount is 'very large' ***AND*** complexity is 'difficult' ***OR***
amount is 'very large' ***AND*** complexity is 'very difficult',
THEN effort is 'huge'.

Let's consider the following expert's verdict: «*Well, I am quite sure that this {feature} is easy to implement; to tell the truth, I'm also more or less sure that it requires a large amount of work to do*». From this statement, we can extract the following pairs of linguistic terms: 'easy' → 'quite sure' and 'large' → 'more or less sure'. Membership degrees in use are summarized in Tables 4 and 5 (elements of T(*A*) and T(*C*) – five terms in each case):

Table 4. Membership degrees of the complexity C values (example).

The complexity of work (set T(*C*))	**Membership degree**
value 'very easy'	0.1
value 'easy'	0.8
value 'normal'	0.1
value 'difficult'	0
value 'very difficult'	0

Table 5. Membership degrees of the amount of work A values (example).

The amount of work (set T(*A*))	**Membership degree**
value 'very small'	0

value 'small'	0
value 'medium'	0.25
value 'large'	0.5
value 'very large'	0.25

In this case, fuzzy rules R2, R3 and R4 will give non-zero resultant value. As already stated above, Mamdani inference system (FIS) is used in the experiments – it allows to obtain an output in the form of fuzzy set. Rules R2, R3 and R4 "fire", thus ensuring non-zero results; in compliance with (3) and (4), we arrive at the following:

R2: $max\big(min(0.1,0.25),min(0.8,0.25)\big)=0.25$ – membership degree that corresponds to the term 'little' (element of T(E)),

R3: $max=0.5$ – membership degree that corresponds to the term 'average' (element of T(E)),

R4: $max\big(min(0.1,0.5),min(0.1,0.25)\big)=0.1$ (label of the term 'big' as the element of T(E)).

COA (Center Of Area) method (2) is applied to obtain crisp result. According to equation (2), the output value equals to approx. 45 story points as shown in Fig. 4.

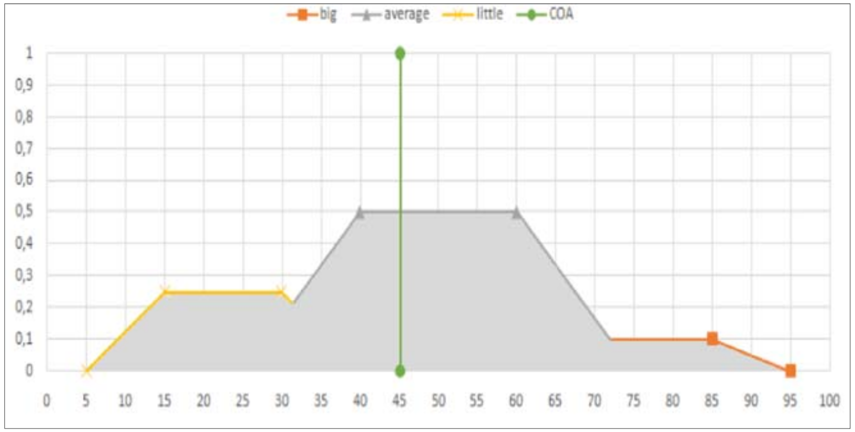

Fig. 4. The result of defuzzification (COA, approx. 45 points).

In Scrum story points estimation' approach the experts often aggregate their opinions using the method of planning poker. It relies on collective judgments (several rounds may become necessary until experts make an agreement) and tries to avoid "pointless haggling over small differences" by compelling to use estimation value from a set of sharply defined distinct values [17]. All participants (they can also be called estimators) secretly write down their estimations in story points on preprepared cards, and then all cards are laid on the table at one time. If all participants select the same value, this value becomes the feature estimation. If not, each expert one after another explains his/her reasons in showing preference for specific value provided, especially when the choice is fixed upon the highest and the lowest estimators in the set. Afterwards, the process is reiterated, i.e. experts vote again, planning poker goes on. It continues until estimators arrive at the agreement.

As regards the aggregation procedure, two approaches mentioned earlier are used. The exact way to calculate the aggregated opinion based on estimations expressed is chosen according to simple rule: if some of them (estimations) are repeated, the equation (6) is used; otherwise, the equation (5) is preferred.

4. Results of experiment – different groups of potential users. Does the proposed method work?

For the sake of completeness, we have asked several groups of people about their views regarding proposed method (its details were discussed with persons concerned in advance). **Group 1** consisted of those people who have worked with story points for a long time. Those delegates who worked with story points before for relatively short-term period formed **group 2**, while those who know what story points are, but have never used them earlier found themselves in the **3rd group**. Finally, people who never even heard of story points fell into **group 4**. As a result, opinions stated below were emphasized (single form of statements are cited for convenience):

(1) **group 1**: «*…I personally consider story points to be the most effective and quite fast way of evaluating features. I make almost no mistakes in estimating features now, and I can adapt myself in new projects in a short time. Your approach is not useful for me now, though I think it might be helpful at the beginning of (my) career*»,

(2) **group 2**: «*…As for me, it took about two months to fully understood the concept of story points, but even now I sometimes make mistakes while estimating features in terms of story points. Today I believe that estimating in story points is more convenient than estimating in hours or some other units. I'm quite experienced member of the currently ongoing project, and I don't need your approach now, though I could still use it, if I have to get the feel of some new project later on*»,

(3) **group 3**: «*…I have heard that story points exist, and that they are used in project estimation, though I have no experience of participating in real projects, where story points were adopted. I think that your approach is better for me right now than story points in their "pure" appearance as I understand it more clearly as compared to story points per se*»,

(4) **group 4**: «*…Oh, I have no idea what are these "story points" are, so obviously, I better prefer to give my opinion on how much work I will have to do to implement the feature, or how difficult this work seems to me*».

Afterwards, we gave people a description of the project (Android App "VR Quest in city" and its features planned for implementation) was introduced to people who took part in the interview session. They were asked to estimate these features both (A) in terms of "plain" story points and (B) using proposed approach.

Table 6. The results of the conducted experiment.

Feature name		**gr. 1**	**gr. 2**	**gr. 3**	**gr. 4**
Create a login form	story points	30	28	40	45
	our approach	35	37	30	28
Find a quest with specific parameters	story points	27	30	55	60
	our approach	29	24	27	30
Save/load a quest	story points	20	25	35	45
	our approach	18	22	20	19
Begin a quest walkthrough	story points	15	18	25	40
	our approach	16	14	15	17
Buy quests in local currency	story points	50	48	75	85
	our approach	52	55	50	45

As shown in Table 6 and Fig. 5-6 (data obtained for groups 1,4 only are visualized), the results of basic story points estimation for the group 1 (participants in this group know how to estimate features in story points), differ not appreciably from the results revealed by proposed approach. It can be treated as initial piece of empirical evidence of the fact that our method is relevant enough and can be used for feature estimation and further elaboration. Moreover, results in both groups 3 and 4 (members of these groups have never used story points before) are substantially different in case of our method as compared with basic story points estimation' approach. This can be attributed to the marked fact that people do not really understand what story points are in the context of non-using them earlier. This is an extra argument in favor of potential utility of the proposed method for those people who are new to Scrum.

Taking story points estimates as landmarks, the Root Mean Square Error is growing steadily from 2.76 for group 1 to 28.26 for group 4 (for groups 3 and 4 the error values are equal to 6.18 and 19.15, correspondingly). The MAD measure, i.e. the size of deviation in units of landmarks from values calculated with the help of proposed approach ((A) and (B) estimates, Table 6), is progressing from 2.4 (group 1) to 27.2 (group 4), while values of 5.8 and 17.6 stand for groups 3 and 4, accordingly. These error values show certain tendency of drawing groups 1 and 2 together along with more perceptible isolation (or, distancing) of «groups 3 and 4» bundle from the practical standpoint of both perception and acceptance of story point-based estimation approach. However, even against a background of such observation, group 3 reveals some positive "detachment" toward group 4. In aggregate, we may conclude that the proposed method has rather tangible effect (in decreasing sequence) on group 1, group 2 and group 3 just "touching" the latter in passing. To the opinion of authors, it can be treated as encouraging sign that is incentive to continue research in this direction.

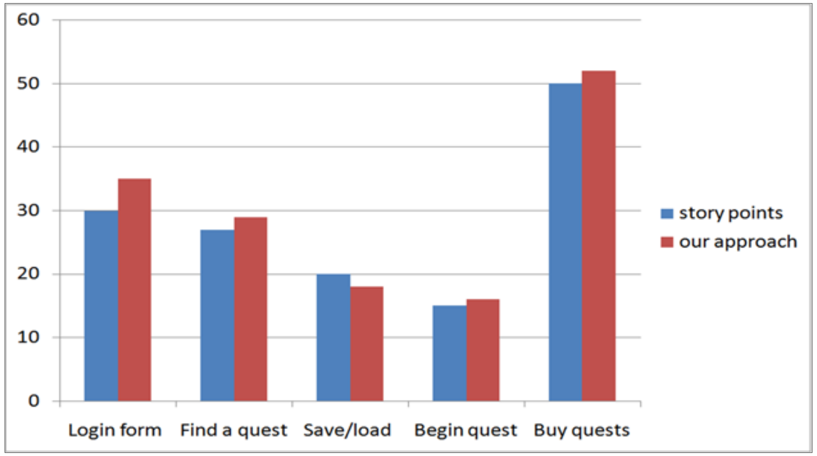

Fig. 5. The results of the conducted experiment as applied to group 1.

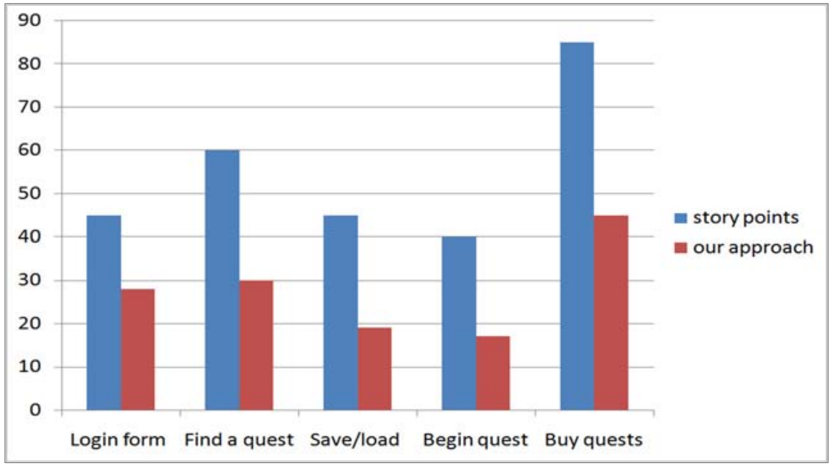

Fig. 6. The results of the conducted experiment as applied to group 4.

5. Conclusion

A novel approach that relates to feature estimation in terms of story points was presented in the paper. The natural idea behind the approach reflects the fact that people may estimate their perception (ideas) concerning the *complexity* of implementation of certain product's feature to be and the *amount of work* to be done to develop this feature. Besides, they can also specify the level of their confidence (or, confidence degree) in evaluation provided. Fuzzy inference scheme lays both solid and transparent groundwork for converting aforesaid input information (data) to the number of story points that can be utilized in the software project management (SPM) at a later stage.

To the opinion of authors, this approach allows people to adapt to Scrum more smoothly, with better understanding of what is implied by story points, grasping the general idea and learning faster their use in practice. The experimental study of the proposed method has shown results approaching the estimations provided by Scrum experts who have been working in real projects and making use of story points for several years. According to survey conducted, such approach can be successfully applied by Scrum newbies, since it is more convenient for people who just make up with story points estimations.

It must be noted that full awareness of strong and weak points of the proposed approach reasoning from one example (project) cannot be realized entirely. Therefore, a sequel of empirical studies and active cooperation with Scrum teams may result in enhancement of the approach. One thing is just to mention that the method seems both promising and handy, but it's quite another matter to make it applicable in practice because of convenience and clearness, at least, as a part of induction stage of the "immersion" to Scrum. Transparent and well perceptible ideas of fuzzy logic are very much to the point here.

Further steps can be associated with intensive studies of more complicated methods of aggregation of the experts' opinions – in particular, they may consider the level (or, weight) of professional qualification of domain experts drawn into project activity. Currently a program's prototype to support (implement) the approach discussed in the paper is under development. The present-day agenda also covers the development of plugin for JIRA tracking system. It is also worth mentioning that certain refinements and changes of the proposed approach can be done at the theoretical level either – some of them are visible enough at present. For instance, the confidence degree values can be represented as intervals, i.e. a form of uncertainty/vagueness expression at the lowest level of comprehension. Such intervals may come about as an effect of possible discord concerning the choice of crisp values shown in the Tab. 2. For the time being, these values may be treated as rough aggregated estimates underlying the computational steps of the discussed approach. Besides, the transition from intervals to type-1 fuzzy sets is also an explicable option to consider. Fuzzy set can be decomposed into a series of nested crisp intervals (so-called α-cuts of a fuzzy set), and this fact can be effectively used in algorithms.

Without confining ourselves to just modeling linguistic terms that stand for confidence levels in use, type-2 fuzzy sets and systems are also regarded as "right" candidates for expansion research efforts in a given problem.

References

[1]. Trendowicz A., 2013. Software Cost Estimation, Benchmarking, and Risk Assessment: The Software Decision-Makers' Guide to Predictable Software Development, Springer-Verlag

[2]. Živadinović J., Medić Z., Maksimović D., et al., 2011. Methods of Effort Estimation in Software Engineering. Proc. Int. Symposium Engineering Management and Competitiveness (EMC), 417–422.

[3]. Briand L.C., Wieczorek I. Resource Estimation in Software Engineering. Int. Software Engineering Research Network, TR ISERN 00-05, web-resource: https://pdfs.semanticscholar.org/943d/a2bb363c06319218ee204622bb10f816490f.pdf (access date 24.02.2017)

[4]. Shivangi S., Umesh K., 2016. Review of Various Software Cost Estimation Techniques. International Journal of Computer Applications, vol. 141, 31–34.

[5]. Colomo-Palacios R. González-Carrasco I., et al., 2012. Resyster: A Hybrid Recommender System for Scrum Team Roles based on Fuzzy and Rough Sets. Int. Journal Appl. Math. Comput. Science, 2012, Vol. 22, No. 4, 801–816.

[6]. Industrial Logic site: Stop Using Story Points, Kerievsky J. (blog), 2012, web-resource: https://www.industriallogic.com/blog/stop-using-story-points/ (access date 24.02.2017)

[7]. Pries K.H., Quigley J., 2010. Scrum Project Management, CRC Press

[8]. Aliev R.A., Aliyev R.R., 2001. Soft Computing and Its Applications, World Scientific

[9]. Zadeh L.A., 1965. Fuzzy Sets, Information and Control, #8, 338–353.

[10]. Bingyi K., Daijun W., Li Y., Deng Y., 2012. A Method of Converting Z-Number to Classical Fuzzy Number. Journal of Information & Computational Science, 9, #3, 703–709.

[11]. Zadeh L.A., 1975. The Concept of a Linguistic Variable and Its Application to Approximate Reasoning - I. Information Sciences, vol. 8, no. 3, 199–249.

[12]. Fuzzy Logic Fundamentals, Pearson Education, 2001, Ch.3, 61–99, web-resource: http://ptgmedia.pearsoncmg.com/images/0135705991/samplechapter/0135705991.pdf (access date 21.03.2017)

[13]. Klir G.J., Bo Yuan., 1995. Fuzzy Sets and Fuzzy Logic: Theory and Applications, 1st ed., Prentice Hall

[14]. Zadeh L.A., 1996. Fuzzy logic = Computing with Words. IEEE Trans. Fuzzy Systems, vol. 4, no. 2, 103–111.

[15]. Zadeh L.A., 1992. Fuzzy Logic and the Calculus of Fuzzy If-Then Rules. Proc. 22nd Intl. Symp. on Multiple-Valued Logic, Los Alamitos, CA: IEEE Computer Society Press, 480–480.

[16]. Quing L., 2013. A Novel Likert Scale Based on Fuzzy Sets Theory. Expert Systems with Applications, vol. 40, #5, 1609–1618.

[17]. Meyer B., 2014. Agile! The Good, the Hype and the Ugly, Springer Int.

Модифицированный метод оценки Story Points в методологии разработки Scrum, основанный на теории нечеткой логики

С.А. Семенкович <sofya-semenkovich@yandex.ru>
О.И. Колеконова <okolekonova@gmail.com>
К.Ю. Дегтярев <kdegtiarev@hse.ru>
Национальный исследовательский Университет
«Высшая Школа Экономики»,
125319, Москва, Кочновский проезд, д. 3, Российская Федерация

Аннотация. Существует несколько известных методов, позволяющих оценить усилия, которые придется потратить на разработку программного обеспечения. В популярной на сегодняшний день методологии гибкой разработки Scrum для этих целей широко используется подход, основанный на story points. Однако, их использование для оценки объема работы может быть затруднительным для тех людей, которые только начинают знакомство с методологией Scrum или впервые попадают в новую Scrum-команду. Описанный в статье подход предлагает использовать оценку трудозатрат на разработку конкретной части программного продукта на основе привычных и понятных для всех фраз естественного языка. Предложенная система нечеткого вывода (модель Мамдани) позволяет преобразовывать мнения людей, выраженные в виде предложений на естественном языке, в число story points – проведенные исследования эмпирически показывают, что те, кто делает первые шаги в методологии Scrum, считают такой подход более удобным и простым, по сравнению с обычным методом оценивания в story points. Также, с целью выяснения, может ли разработанный подход использоваться при работе над реальными проектами, был проведен дополнительный эксперимент, в котором приняли участие четыре группы людей с различными уровнями квалификации в Scrum-разработке. Представителям этих групп было дано задание оценить трудозатраты на разработку отдельных частей некоторого проекта с использованием предложенного подхода и обычных story points единицах. Оценки группы экспертов в области Scrum оказались примерно одинаковы для обоих подходов, в то время как оценки ′новичков′ в методологии сильно отличались при применении двух разных методов. По мнению авторов, предложенный подход может дать возможность более плавного вхождения в методологию Scrum, лучшего понимания природы story points и более быстрой выработке навыков работы с ними на практике. Отдельного внимания заслуживает вопрос изучения разных форм агрегации мнений экспертов, анализ альтернативных подходов к представлению степеней уверенности экспертных оценок и возможная разработка плагина для системы отслеживания ошибок JIRA. Всё это может составить предмет развития данной темы.

Ключевые слова: нечеткая логика; Scrum; story points; экспертные оценки; агрегация мнений; система нечеткого вывода; шкала Лайкерта

DOI: 10.15514/ISPRAS-2017-29(5)-2

Для цитирования: Семенкович С.А., Колеконова О.И., Дегтярев К.Ю. Модифицированный метод оценки Story Points в методологии разработки Scrum, основанный на теории нечеткой логики. Труды ИСП РАН, том 29, вып. 5, 2017 г., стр. 19-38 (на английском языке). DOI: 10.15514/ISPRAS-2017-29(5)-2

Список литературы

[1]. Trendowicz A., 2013. Software Cost Estimation, Benchmarking, and Risk Assessment: The Software Decision-Makers' Guide to Predictable Software Development, Springer-Verlag

[2]. Živadinović J., Medić Z., Maksimović D., et al., 2011. Methods of Effort Estimation in Software Engineering. Proc. Int. Symposium Engineering Management and Competitiveness (EMC), 417–422.

[3]. Briand L.C., Wieczorek I. Resource Estimation in Software Engineering. Int. Software Engineering Research Network, TR ISERN 00-05, web-resource: https://pdfs.semanticscholar.org/943d/a2bb363c06319218ee204622bb10f816490f.pdf (access date 24.02.2017)

[4]. Shivangi S., Umesh K., 2016. Review of Various Software Cost Estimation Techniques. International Journal of Computer Applications, vol. 141, 31–34.

[5]. Colomo-Palacios R. González-Carrasco I., et al., 2012. Resyster: A Hybrid Recommender System for Scrum Team Roles based on Fuzzy and Rough Sets. Int. Journal Appl. Math. Comput. Science, 2012, Vol. 22, No. 4, 801–816.

[6]. Industrial Logic site: Stop Using Story Points, Kerievsky J. (blog), 2012, web-resource: https://www.industriallogic.com/blog/stop-using-story-points/ (access date 24.02.2017)

[7]. Pries K.H., Quigley J., 2010. Scrum Project Management, CRC Press

[8]. Aliev R.A., Aliyev R.R., 2001. Soft Computing and Its Applications, World Scientific

[9]. Zadeh L.A., 1965. Fuzzy Sets, Information and Control, #8, 338–353.

[10]. Bingyi K., Daijun W., Li Y., Deng Y., 2012. A Method of Converting Z-Number to Classical Fuzzy Number. Journal of Information & Computational Science, 9, #3, 703–709.

[11]. Zadeh L.A., 1975. The Concept of a Linguistic Variable and Its Application to Approximate Reasoning - I. Information Sciences, vol. 8, no. 3, 199–249.

[12]. Fuzzy Logic Fundamentals, Pearson Education, 2001, Ch.3, 61–99, web-resource: http://ptgmedia.pearsoncmg.com/images/0135705991/samplechapter/0135705991.pdf (access date 21.03.2017)

[13]. Klir G.J., Bo Yuan., 1995. Fuzzy Sets and Fuzzy Logic: Theory and Applications, 1st ed., Prentice Hall

[14]. Zadeh L.A., 1996. Fuzzy logic = Computing with Words. IEEE Trans. Fuzzy Systems, vol. 4, no. 2, 103–111.

[15]. Zadeh L.A., 1992. Fuzzy Logic and the Calculus of Fuzzy If-Then Rules. Proc. 22nd Intl. Symp. on Multiple-Valued Logic, Los Alamitos, CA: IEEE Computer Society Press, 480–480.

[16]. Quing L., 2013. A Novel Likert Scale Based on Fuzzy Sets Theory. Expert Systems with Applications, vol. 40, #5, 1609–1618.

[17]. Meyer B., 2014. Agile! The Good, the Hype and the Ugly, Springer Int.