



A Method for Estimating the Efficiency of Commanding in the State Fire Service of Poland

Adam Krasuski, Karol Kreński and Stanisław Łazowy, The Main School of Fire Service, Warsaw, Poland*

Received: 23 May 2011/**Accepted:** 1 October 2011

Abstract. A rough method for estimating commanding efficiency of Fire&Rescue officers is presented in the article. The idea is to evaluate the officers in charge and their actions based on the information collected in the incident data reporting system. The criteria for the commanding evaluation is the distribution of durations of the actions conducted by consecutive commanders.

Keywords: Commanding efficiency, Fire statistics, Operating time distributions, Clustering, Fire service commanding estimation

1. Introduction

In most countries, all incidents involving Fire Service are collected in some sort of incident data reporting system (IDRS). The following are examples of IRDS in various countries: ONTIKA in Finland [1], IMS in UK [2], NFIRS in USA [3] or EWID in Poland [4]. For years those IDRS have collected large amounts of data. For example, the Polish IRDS–EWID till now has collected approximately 6 million incidents.

Since the IDRS is a broad source of information about full range of actions, it is reasonable to benefit from it as much as possible. Such a large IDRS database allows the building of the decision support system (DSS) based on case-based reasoning (CBR) paradigm [5] and this is being researched by the authors [6].

CBR is an Artificial Intelligence technique that simulates the processes that are used by humans for problems solving [5, 7]. CBR systems retrieve solutions which were used in the past to solve similar problems. These solutions need to be then adapted to the new circumstances which makes it a self-learning process.

There is a full spectrum of solutions, from the best to worst, in the IDRS. There is a need to select only valid solutions from the IRDS database as a data layer of a reliable DSS.

In this article the heuristic, which evaluates the efficiency of the solutions collected in EWID is described. This method evaluates the efficiency of the commanders at the fire ground. The rescue methods of the best commanders will further be taken for consideration as a purposed solutions to the DSS system.

* Correspondence should be addressed to: Adam Krasuski, E-mail: krasus@inf.sgsp.edu.pl

2. Data Features and Pre-Processing

In most IDRS the data collected in the database are divided into two sections—the structured (database fields) and unstructured (description in natural language (NL)). The structured section contains, among other, the information about the type and size of rescue action, the number of rescue units and the quantity and type of resources used during the action. The unstructured part of the database, which is filled with NL descriptions, may contain valuable information which is not present in the structured part, e.g. the rescue methods that were used to manage the rescue action.

The pre-processing was first applied to the data. False or incomplete data were removed or repaired, they were merged in one table (information table), normalized and lemmatized. The repair of the data was achieved by using the data from the other fields or metadata where possible.

The proposed heuristic of the commanders' efficiency estimation requires that incidents are clustered. This is needed to make the commanders' comparison reliable, as they have to be compared in similar action types. The full description of clustering is described in Section 3.1. The data were normalized, so that all the attributes from information table were of the same range and weight. Formula 1 depicts the normalization equation.

$$x_{norm} = \frac{x - \bar{x}}{D} \quad (1)$$

where x is a selected value; \bar{x} is the mean of all the values; D is the average absolute deviation calculated according to formula 2.

$$D = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}| \quad (2)$$

Normalization was applied to the structured part of the database only. The NL part of EWID was lemmatized in order to simplify the clustering process of that part of the database.

In computational linguistics, lemmatization is a process of determining the lemma for a given word, i.e. all the inflexed forms are flattened to the basic form. This is particularly an issue with the Polish language, which is very rich in inflexed forms. Lemmatization was performed in Morfologik, an open source Polish morphological analyzer based on ispell dictionary [8].

3. Description of the Method

The proposed method focuses on selecting the best commanding by choosing the commander with best performance. The database information about commanders and their commanding were compared. In order to achieve the unbiased results the scope of the commanding must be set—the commanders have to operate under similar conditions, preferably the circumstances should be identical.

It is difficult to evaluate whether two incidents are identical, especially on the basis of a closed set of attributes in EWID database. It can only be stated, that two incidents are similar to the extent of the considered set of attributes which describe the incident. Therefore, it must be noted that the presented method is only heuristic, which only approximates the commanding efficiency. Any conclusion that can be made is limited to the given commander in given circumstances, not about the performance of the given commander in general.

3.1. Grouping of the Similar Incidents

According to the data structure, the EWID database is composed of two separate parts: (a) the structured part (information table) and (b) the NL part lemmatized in the data pre-processing. Authors have decided that clustering of both parts should be done separately and then logical AND should be applied to the clusters to increase their quality. Unfortunately, the clustering of NL part was not successful and only the structured part was used to define the final clusters.

The dataset was first divided into three groups: fire incidents, local threats and false alarms. Then clustering was performed only on the fire incidents and the other two were excluded.

Clustering is an automatic process which requires that the target number of resulting clusters is defined in advance and the initial clustering aimed at finding this target number. PAM [9] was chosen as the clustering algorithm and a sample of 10000 incidents was evaluated. The accuracy of resulting clusters was determined by using Silhouette width [10] and Calinski Harabasz index [11].

Silhouette width $s(i)$ is defined as follows (formula 3):

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \tag{3}$$

where i is an observation (in this case incident); $a(i)$ is average dissimilarity between i and all other points of the cluster to which i belongs (if i is the only observation in its cluster then $s(i) = 0$ without further calculations); $b(i)$ is the dissimilarity between i and the nearest cluster to which it does not belong.

Calinski and Harabasz index $CH(k)$ estimates the quality of well isolated and coherent clusters. $CH(k)$ is calculated according to the formula 4:

$$CH(k) = \frac{B(k)(k - 1)}{W(k)(n - k)} \tag{4}$$

where k denotes the number of clusters and $B(k)$ and $W(k)$ denote the between (separation) and within (cohesion) cluster sums of squares of the partition, respectively. Separation is measured by the formula 5:

$$BSS = \sum_i \sum_{x \in C_i} (x - m_i)^2 \tag{5}$$

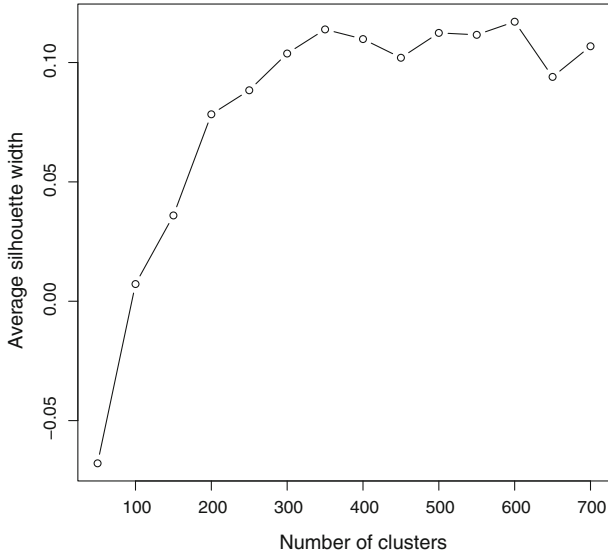


Figure 1. Silhouette width against the number of clusters (high values are better).

$$WSS = \sum_i |C_i|(m - m_i)^2 \quad (6)$$

where $|C_i|$ is the size of cluster i . An optimal number of clusters is then defined as a value of k that maximizes $CH(k)$. In essence, the higher the value of this index, the greater the separation between clusters.

The number of clusters in the experiment was varied from 50 to 700 and the $S(i)$ and $CH(k)$ were calculated. Figure 1 depicts the result of the experiment for Silhouette width.

The Silhouette width grows with the increase of the number of clusters. Around the number of 300 clusters, the $S(i)$ stabilizes around the level 0.1. The reason is that one-object-clusters start to appear ($S(i)$ for this case is equal 0). Therefore, it was reasonable to set the target number of clusters at the value of 300, which is the minimum value where $S(i)$ reaches 0.1.

Silhouette width can vary from -1 , which means very poor quality of clusters, to 1 denoting very good quality. Therefore the value 0.1 achieved in the experiment is rather low. In the next step, full Silhouette analysis for $k = 300$ clusters was performed to check the reason for the low value of the index. Figure 2 depicts the result of the experiment.

According to the Figure 2, there is a set of clusters which have a fairly high value of $S(i)$ and the set of clusters for which the value of $s(i)$ is below 0. The negative values represent the incidents which rarely occur and are clearly different from other cases. They weaken the $S(i)$ because there are no similar incidents in

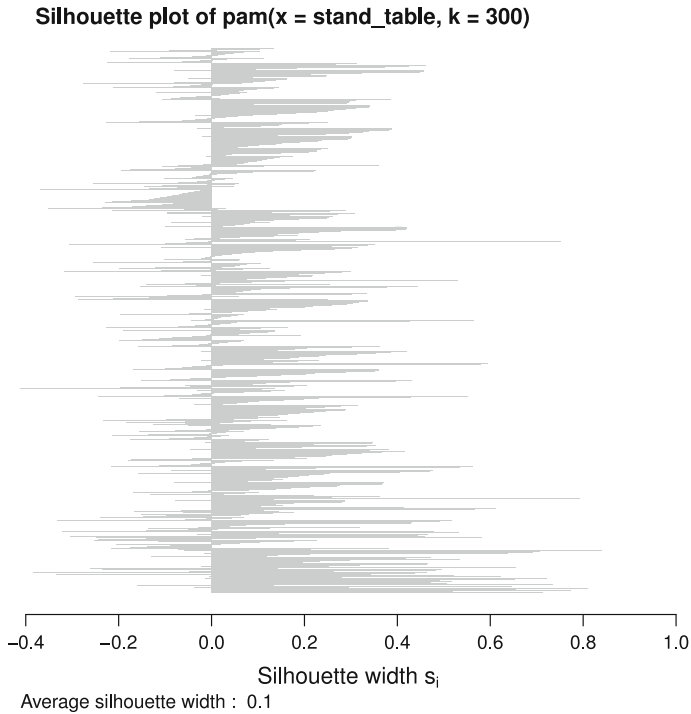


Figure 2. Silhouette width for k = 300.

the data set. Therefore, the further analyses were performed on clusters where $s(i) > 0.4$.

After determining the target number of clusters, the clustering process was performed on full database. In order to handle the large size of the database, CLARA [12] clustering method was used. CLARA is designed to deal with the memory management problems, by not moving all the data into the core memory at once.

The next stage was clustering the NL part of EWID using algorithm called *latent semantic analysis* (LSA) [13, 14]. The basic idea of LSA is to first create concepts for the given text corpus and then assign each single word from a document to a concept. The result is that documents can be expressed in Latent Semantic Space which: (a) is considerably compact and (b) allows for finding indirect similarities between documents.

However, there were difficulties in clustering the NL part as the computer resources were not sufficient to conduct the LSA clustering, which turned out to be very demanding for computer power. The clustering of NL part was not completed and the final clusters resulted from just the structured part of the database.

Figure 3. Comparison of C1 to C5 commanders efficiency as probability distribution of action times [min]: (a) the trashcans fires, (b) the medium fires of the apartment buildings, (c) the small fires of the apartment buildings. Symbols in the tables: \bar{x} is the mean, σ is the standard deviation.

3.2. Regression of the Operating Time Distribution

After the clusters of similar cases are created it is possible to compare the quality of commanding of different commanders. The chosen criteria for comparing the commanders is their *operating time*. Operating time is measured from the arrival of the fire crew at the scene until the end of the action.

The efficiency of commanding of different commanders in different types of the fires can be now evaluated. Examples of types of fires are the small/medium/large apartment buildings fires, the forest fires, the trashcan fires, etc. Each cluster was further clustered by the successive commanders. For each of such subsets the distribution of operating time was calculated and the adequate model was fitted—the log-normal was chosen as the one universal model for all the distributions.

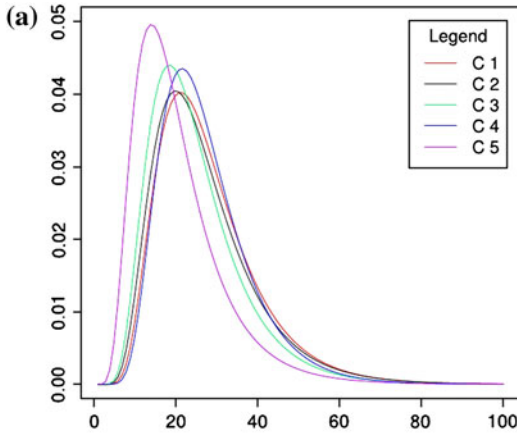
The parameter determining the appropriateness of the model was *P-value* from Shapiro-Wilk test [15, 16]. The Shapiro-Wilk test was performed on the natural logarithm of the operating time values. There was not a single occurrence of *P-value* lower than 0.05, which justifies the usage of log-normal to the acceptable degree.

After the regression, the distributions of operating times of different commanders were compared (Figure 3).

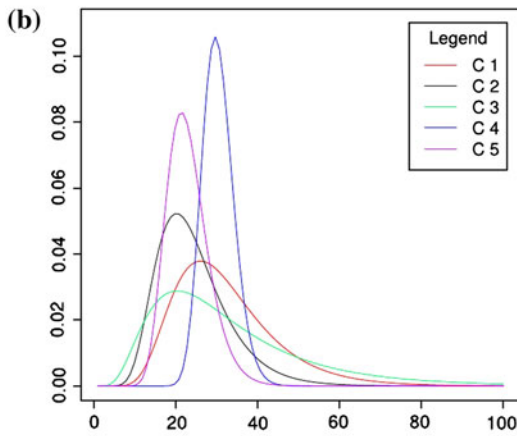
4. Discussion

The final interpretation of the results may be based on the two key parameters of log-normal distribution: the *mean* (\bar{x}) and the *standard deviation* (σ). The low value of the standard deviation implies that the commander uses a closed set of methods during this type of actions. His methods are quite repetitious, which may result in his confidence in the commanding—this is how the commander gains and develops his experience.

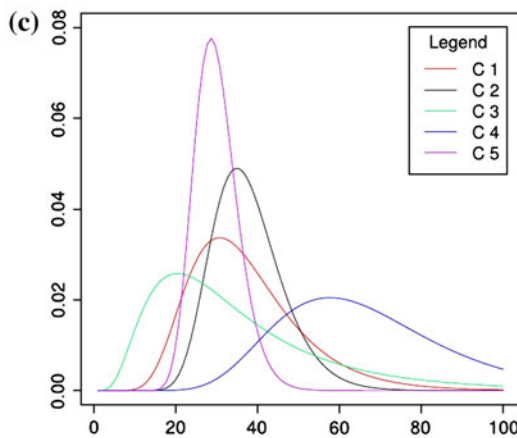
The low value of the standard deviation could be an estimator of the commander's experience for the given type of fire. However, judging just by this one parameter can be misleading—there could be commanders with high repeatability, but their solutions may be not efficient in terms of time. Therefore, the mean of the distribution should be also considered when evaluating the commanders' efficiency—the smaller it's value, the better (smaller average time of handling this type of fire). It must be noted that there is no formula defined, weighting the mean against the standard deviation in order to point the best performing commander. In order to obtain such a formula, a thorough experts' analysis would most likely be required. However, the discussion and the conclusions below is based on the authors rough judgement.



	\bar{x}	σ
C1	3.24	0.43
C2	3.20	0.45
C3	3.11	0.44
C4	3.23	0.39
C5	2.90	0.50



	\bar{x}	σ
C1	3.40	0.38
C2	3.13	0.35
C3	3.35	0.57
C4	3.41	0.13
C5	3.11	0.22



	\bar{x}	σ
C1	3.56	0.36
C2	3.61	0.23
C3	3.35	0.62
C4	4.16	0.32
C5	3.39	0.18

In the case of the log-normal distribution, around 95% of the distribution belongs to the interval $\exp(\bar{x} \pm 2\sigma)$ (two standard deviations from the mean). Comparative example would be the performance of the commanders C5 against C4 in the trashcan fires (Figure 3a). The values for C5 are: $\exp(2.90 \pm 2 \times 0.50)$ versus the values of C4: $\exp(3.23 \pm 2 \times 0.39)$. The obtained intervals are C5: [6.7, 49.4] versus C4: [11.6, 55.1]. This means that the commander C5 handles 95% of his actions within 6.7 min to 49.4 min, while it takes 11.6 min to 55.1 min for the commander C4.

It is interesting to note, that despite the fact, that C5's standard deviation is higher ($\sigma_{C5} = 0.50$ against $\sigma_{C4} = 0.39$), the overall consistency of C5 is better: the range of C5 interval is $49.4 - 6.7 = 42.7$ against C4's range $55.1 - 11.6 = 43.5$.

Next is the discussion of the analysed three types of the fires:

Case 1: the trashcan fires (Figure 3a). The difficulty of handling the trashcans fires is low, therefore the differences among distributions of operating times for different commanders is not significant. It is therefore possible to point that the commander C5 has the best performance in this type of the fires. The standard deviation of C5 is high, which means he is the least consistent of the five commanders, however it is visible from the Figure 3 that he is way ahead of the others in terms of time.

Case 2: the medium fires of the apartment buildings (Figure 3b). This type of the fires demands more experience from the commanders, to handle it successfully. Therefore the differences among distributions, which are depicted on the figure are more evident. As the figure shows, commanders C4 and C5 are quite experienced. They know how to manage this type of the fires and they use the repetitious, successful procedures as indicated by the small value of standard deviation. Commander C5 performs better than C4—they are both comparably repetitious, but C5 has shorter average time.

Case 3: the small fires of the apartment buildings (Figure 3c). Again, commander C5 proves to be the most efficient. He has the lowest value of standard deviation and relatively low value of mean. Although, commander C3 has the lowest value of mean, he lacks repeatability as shown by high value of the standard deviation—the methods are not as reliable, as those of C5.

The weight of the standard deviation value against the mean value is the authors interpretation, but seems to be reasonable for the analysed data as expressed by the Figure 4: large confidence intervals of officer C3 indicate that occasionally the commander performs poorly which decreases his reliability. In terms of specific numbers, it can be stated with 95% confidence, that C3 will handle a fire within 9 min to 96 min interval while C5 will handle a fire within 21 min to 43 min.

5. Conclusions and Future Work

In the article the method of estimation of efficiency of commanding was presented. This method is quite simple, fast and reliable as an estimator of the officers' performance: the shorter the duration of the action, the better the commanding officer. The article proves that EWID, as a large collection of data,

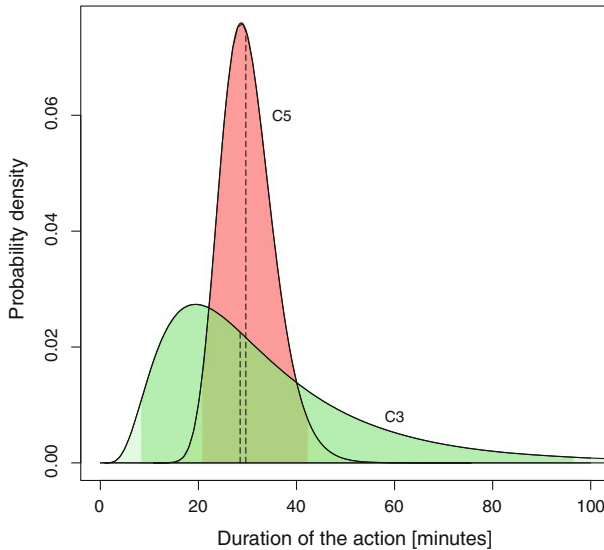


Figure 4. Comparison of 95% confidence intervals of the commanders C3 and C5 (the small fires of apartment buildings). The dotted lines are the means.

allows for more advanced and interesting statistics of this sort, as opposed to just basic statistics currently made. No approaches for evaluation of the commanders' performance were known to the authors prior to this research. This implicates that the results of this research cannot be verified against other methods, as the method proposed here, is the first attempt to compare the commanders. The authors do not indicate any direct usage of the discovered knowledge, however finding the role models for educational purposes for firefighters community can be an example application.

The method has also its shortcomings:

First, the applied clustering process is a very time and resources consuming task. Clustering of large dataset requires a large amount of RAM memory. This is especially the case during the LSA stage. There is a risk that the research could not be repeated on the full EWID database. The database authors were using, consists of 0.3 million records, as opposed to 6 million records of the full EWID database.

The next issue is the use of operating time as an estimator—this one can be biased. For example, commanders' methods may be indeed very fast but generate economic losses at the same time. This can be fixed by enhancing the model by taking the economic losses under account. Unfortunately these data are not directly available from EWID database.

It must be noted, that not all of the commanders can be evaluated by using this method. The statistical mechanisms used applied, requires that there have to be at least 5 entries of a given commander in the validated cluster. Clusters exists where no single commander satisfies this criterion. Officers with 5-entries-experience

shouldn't be considered anyway, as those officers should not be considered experienced—the limitation of the statistics is just the reflection of the limitation of the reality.

The proposed model may be further enhanced by processing more parameters. The subject for this analysis was the city of Warsaw, where Fire&Rescue Units use considerably similar equipment. If the commanders from different urban agglomerations were to be compared, the model should take into account that some of them could gain advantage by the using of a superior equipment. The information about the equipment possessed by the Fire&Rescue Units are accessible from EWID database. Also, the season / part of the day could enhance the model—they can be easily calculated from EWID data, but for this research only hours (from 0 to 23) were clustered.

EWID contains some more potentially important information, which is not easily accessible. For last few years the commanders have been providing the information about the weather in the reports. The problem is that this data are written in the unstructured part of the database in an undefined form. It would require an additional NL processing to access this information.

There are factors, however, which could affect the results, but are not available in EWID. The information about city infrastructure, e.g. distributions of various types of buildings in the given area, could most likely be obtained from the city authorities.

What was mentioned above is just a few example parameters, which have a potential to improve the model. The authors' intention was not to include in this research all these or even more parameters, as this would require a much broader research. The idea was to show the approach and demonstrate that, even this simplified implementation is capable of comparing the commanders.

Open Access

This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Rahikainen J, Keski-Rahkonen O (2004) Statistical determination of ignition frequency of structural fires in different premises in Finland. *Fire Technol* 40(4):335–353
2. IT Training Department (2008) Incident management systems (IMS). User guide. Technical report, Fire Brigade, London
3. United States Fire Administration (2002) National fire incident reporting system. Quick reference guide. Technical report, National Fire Data Center
4. Collective Work (2001) Ewidencja zdarzen—EWID99. Technical report, Abacus. http://www.ewid.pl/?set=rozw_ewid&gr=roz. Accessed date 23 April 2007

5. Aamodt A, Plaza W (1994) Case-based reasoning. *IOComm Artif Intell Commun* 7(1):39–59
6. Krasuski A, Maciak T, Krenski K (2008) Decision support system for fire service based on distributed database and case-based reasoning. *Stud Logic Gramm Rhetor* 11(24):213–221
7. Shiu SCK, Pal SK (2004) Case-base reasoning: concepts, features and soft computing. *Appl Intell* 21(3):233–238
8. Morfologik—About the project. <http://morfologik.blogspot.com/2006/05/about-project.html>
9. Van der Laan M, Pollard K, Bryan J (2003) A new partitioning around medoids algorithm. *J Stat Comput Simul* 73(8):575–584
10. Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 20:53–65
11. Caliński T, Harabasz J (1974) A dendrite method for cluster analysis. *Commun Stat Theory Methods* 3(1):1–27
12. Kaufman L, Rousseeuw PJ, Ebooks Corporation (1990) Finding groups in data: an introduction to cluster analysis, vol 39. Wiley, New York
13. Deerwester S, Dumais ST, Furnas GW, Landauer TK, Harshman R (1990) Indexing by latent semantic analysis. *J Am Soc Inf Sci* 41(6):391–407
14. Landauer TK, Foltz PW, Laham D (1998) An introduction to latent semantic analysis. *Discourse Process* 25(2):259–284
15. Royston JP (1982) An extension of Shapiro and Wilk's W test for normality to large samples. *J R Stat Soc C Appl Stat* 31(2):115–124
16. Royston JP (1982) Algorithm AS 181: The W test for normality. *J R Stat Soc C Appl Stat* 31(2):176–180