



Decision Support System for Blockage Management in Fire Service

Adam Krasuski¹, Karol Kreński¹

¹ Section of Computer Science, The Main School of Fire Service, Warsaw, Poland,
krasuski@inf.sgsp.edu.pl, krenski@inf.sgsp.edu.pl

Abstract. In this article we present the foundations of a decision support system for blockage management in Fire Service. Blockage refers to the situation when all fire units are out and a new incident occurs. The approach is based on two phases: off-line data preparation and online blockage estimation. The off-line phase consists of methods from data mining and natural language processing and results in semantically coherent information granules. The online phase is about building the probabilistic models that estimate the blockage probability based on these granules. Finally, the selected classifier judges whether a blockage can occur and whether the resources from neighbour fire stations should be asked for assistance.

Keywords: Knowledge Discovery, Domain Knowledge, Granular Modeling, Layered Architectures, Fire Services, Text Data, Heterogeneous Data Sources

1. Introduction

In order to adequately respond to emergencies, a sufficient amount of staff and resources as well as geographic coverage are needed. The placement of most of the Fire&Rescue (F&R) stations in Poland are accurate for wide time perspective. Therefore they are difficult to be relocated, even when they are not ideally situated. To overcome this problem of lack optimally situated stations, manual allocation of resources to stations is performed. Planning for resources and manpower coverage, includes a wide variety of factors, among others: national and European Union regulations, fire threatening life, frequency and spatial distribution of calls and roads network, etc.

The calculations where to build and how to equip the station are annually based. Therefore they are accurate for wide time perspective. However, they are useless in those cases when the number of daily calls increases significantly above the annual average. In such situations, stations run out of resources very fast, which can lead to a *blockage*. Blockage refers to the

situation when all fire engines (or generally units) are out and a new incident occurs. Blockages decrease public safety significantly. They lead to the situation when there are no rescue services to handle the incidents and a dangerous situation expands.

In this context, the blockage management are organizational and operational activities focused on preventing the blockages. One of the most popular methods of managing the blockages is temporal movement of engines from the station which has full resources to the one which has a high probability of blockage at the moment. Real time movement of engines from one station to another requires making fast decisions based on accurately predicted future development. The wrong decisions may result not only in economic losses but also in a large number of casualties. Inappropriate management of engines deploying could also lead to disorganisation of rescue system on higher level and to decrease in the potential of rescue system. All these factors make the problem of engines movement in real time very important. Therefore the computer systems which can support the decision makers are very needed.

This article is focused on the handling of the blockages in the fire stations. The main contribution of the paper is the description of an effective solution method that supports decision making in a dynamic deployment of Fire&Rescue (F&R) engines across stations. The core of our method is estimating the expected return times of the gone fire brigades which, along with the distributions of the emergencies against the hours of the day and the distribution of the emergencies per day, allows the evaluation of the probability of the blockage. By using this method it is also possible to predict the return time of the given engine, and through this to evaluate if it is necessary to move the engine from other stations.

The rest of the paper is organised as follows: section 2 contains a brief review of the techniques of units deployment to highlight a comparison of different methods. In section 3 we discuss briefly the impact of the blockages on public safety. In section 4 we present our method of prediction of the blockages. In section 5 we describe the experiments which we conducted to validate our methodology. The last section of the article concludes and interprets the research results and considers perspectives for the application of our model.

2. Related Work

The problem of deployment of the units across stations refers to the number of general problems such as: short-term car rental logistics prob-

lem (Fink and Reiners, 2006), queuing loss system (Shonick and Jackson, 1973; Ormeci and Burnetas, 2004), determination of the optimal fleet size and vehicle transfer (Li and Tao, 2010) or patrol force deployment optimization (Lee et al., 1979; Lau et al., 2010).

In the fire protection domain the blockage issue was addressed in (Hasofer et al., 2007; Tillander and Keski-Rahkonen, 2008). Those articles seem to be focused on the strategical level and the conclusions from the works may benefit the decision makers resolving the problems of equipping the fire stations with the necessary resources. The probability of the blockage can be decreased simply by adding more units or more equipment, so that there are always reasonable reserves. The idea is based on the following approach:

The average number of simultaneous fires C in the area under the study can be calculated as follows:

$$C = \lambda\tau \quad (1)$$

where λ is the number of the fires per unit time and τ is the average operating time per fire.

The probability of the blockage P_B can be assessed by formula (2) originally described in (Beckmann, 1968) in connection with the telephone exchanges, where blocking also occurs when an incoming telephone call finds all lines busy.

$$P_B = \frac{\frac{C^S}{S!}}{\sum_{r=0}^S \frac{C^r}{r!}} \quad (2)$$

where C is, as before, the average number of the simultaneous fires and S is the number of the available fire units. Once the acceptable level of P_B is defined, it is possible to calculate the minimum number of the fire units for the station.

The second approach proposed by Peace (Peace, 2001) is based on various types of risk. In order to provide Fire Service Emergency cover for a given area, the worst case planning scenario (WCPS) is used. This is the worst case for which cover will be routinely planned. The WCPS is selected on the basis of professional judgement. For example, for an estate of domestic dwellings it might be: "whole house involved in fire with rescue required". The set of WCPSes is then assigned to the area. Next step of the approach is allocating resources to stations' turn-out areas. The procedure for allocating resources to the stations is semi-automatic and consists of manual allocation with support of computer software called *pathfinder*. The software reports the consequences of a given vehicle allocation, while the choice

which vehicles are required and where they should be allocated is left to the judgement of the planner.

However, the above methods are focused on ‘designing’ the fire stations rather than on handling dynamic, everyday situations when the fire engines go to the incidents and create the blockage probability. Better suited data and broader analyses are needed for such scenarios, as described further in this article.

Another approach was invented by Taylor and Huxley (Taylor and Huxley, 1989). The proposed model, called the Police Patrol Scheduling System (PPSS), was produced in order to provide better manpower personnel deployment in terms of workload demand over the hour of the week. The model that was developed is an integer programming problem (Nemhauser and Wolsey, 1988). This model helps calculate the number of officers needed by hour by day.

Due to long computation time for the realistic problem, the PPSS model was extended by Church and coworkers (Church et al., 2001). They developed a heuristic solver for the above model. The solver is based upon a TABU search process (Clover, 1989, 1990) and involves a technique called strategic oscillation. The model narrows the time perspective in which we can predict the number of officers, but it is still not resistant to temporary growths of demands in some areas. Therefore, the real time models which can predict the probability of the blockage at any time interval are still needed.

3. Blockages and their Impact on Public Safety

In this section we present short statistical analysis focused on discovering how dangerous the blockages are for public safety. We try to determine what types of losses (if any) blockages generate. We define some numerical values for evaluating the harmfulness of the blockages for the public safety. We analyse these aspects in two different regions: urban with dense coverage of fire stations and small towns with one fire station and few volunteer fire stations located in the surroundings.

It is difficult to evaluate the strict impact of blockage on the public safety. In our work we try to approximate it by a few numerical measures. That is: the delay in arrival of fire engines at the fire ground if the blockage occurs, the number of fatalities and casualties in blockage calls (BC) related to non-blockage calls (NBC), the number of medium and large fires in BC in comparison to the NBC. Throughout these parameters we try to evaluate

the harmfulness of the blockages. We use information stored in Polish Incident Safety Reporting System – EWID. Unfortunately, the EWID database lacks credible information about the financial losses of the fires, therefore we do not include such data in the analysis. Table 1 summarizes the information about the blockages within first area – urban region (Warsaw City).

Table 1
The parameters describing the difference between non-blockage calls (NBC) and the blockage calls (BC) – Warsaw City

C1	C2	C3	C4	C5	C6	C7	C8	C9
FS 1	-2.11	0	3.45%	3.00%	29	65	1.28%	0.00%
FS 2	-2.17	0	0.00%	3.03%	3	236	0.28%	0.00%
FS 3	0.82	0	0.00%	2.13%	2	72	0.21%	0.00%
FS 4	-0.26	0	0.00%	1.75%	3	62	0.40%	0.00%
FS 5	—	—	—	1.17%	0	0	0.18%	—
FS 6	-3.68	0	0.00%	1.26%	1	27	0.70%	0.00%
FS 7	—	—	—	1.92%	0	0	1.02%	—
FS 8	-1.21	0	0.00%	1.22%	1	70	0.72%	0.00%

Symbols used in the table: C1 – Fire Station, C2 – the delay between arrival time in NBC and BC [min], C3 – the number of fatalities in BC, C4 – the average number of casualties for one BC, C5 – the average number of casualties for one NBC, C6 – the number of fires for BC, C7 – the number of local threats for BC, C8 – the average number of medium fires for one NBC, C9 – the average number of medium fires for BC.

Table 1 shows that within the urban regions blockages do not decrease the public safety in general. There is no significant delay in arrival time at the fire ground, they do not generate greater number of casualties or fatalities and the fires in blockages do not expand to medium fires. Therefore, statistically in the urban regions with dense coverage of fire stations the blockages are not an issue. However, table 1 also shows that in most cases for the analysed region, blockages are generated throughout the local threats. After deeper investigation it turned out that these are the incidents generated by local storms, in the form of broken trees, flooded basements or icicles. There is also an exception. The fire station *FS1* had 29 fires served in a blockage mode. In this case, we observe the growth of the number of casualties in comparison to average in non-blockage calls. However, due to the small number of casualties in general this indicator cannot be treated as a strong measure of impact of blockages on public safety.

In the second case we present the same parameters of blockages evaluations for small towns. Table 2 summarises the information.

Table 2

The parameters describing the difference between non-blockage calls and the blockages calls – small towns

C1	C2	C3	C4	C5	C6	C7	C8
FS 1	0.79	0	0	11	34	5.1%	0.0%
FS 2	-3.27	0	0	6	26	8.4%	0.0%
FS 3	6.80	0	0	8	60	3.6%	33.3%
FS 4	3.34	0	0	12	17	7.7%	20.0%
FS 5	-1.24	0	0	24	53	7.5%	26.3%
FS 6	-0.73	0	0	9	31	6.0%	0.0%
FS 7	-1.74	0	0	17	42	7.5%	6.3%
FS 8	2.27	0	0	12	41	6.3%	20.0%
FS 9	-0.28	0	0	27	30	7.3%	17.4%
FS 10	-0.11	0	0	59	92	4.1%	3.5%
FS 11	0.00	0	0	11	67	4.5%	0.0%
FS 12	9.47	0	0	10	54	8.7%	25.0%
FS 13	3.22	0	0	7	34	3.1%	0.0%
FS 14	1.32	0	0	11	62	5.3%	22.2%
FS 15	-1.90	0	0	8	17	6.1%	14.3%
FS 16	-0.07	0	0	55	38	18.6%	35.0%
FS 17	2.52	0	0	210	110	7.5%	11.1%

Symbols used in the table: C1 – Fire Station, C2 – the delay between arrival time in NBC and BC [min], C3 – the number of fatalities in BC, C4 – the average number of casualties for one BC, C5 – the number of fires for BC, C6 – the number of local threats for BC, C7 – the average number of medium fires for one NBC, C8 – the average number of medium fires for BC

Table 2 shows that for the regions with small density of fire stations blockages do have an impact on public safety. This is not visible in the form of number of casualties or fatalities (the reason is probably the small number of such cases in general) but we can observe cases with significant increase in the number of medium fires in blockages in comparison to the non-blockage calls. Moreover, contrary to the Warsaw region, there are many fires served in blockage mode and the analysis is statistically significant. In most cases the greater number of medium fires is correlated with the delay in arrival at fire ground. It means that fire engines come too late and fires are then fully developed. The financial losses resulting from the fires could confirm and evaluate the cost of blockages, unfortunately there is no such information in EWID system, as we mentioned before. However, the difference between small and medium fires can illustrate the scale of the problem. This assures us that the problem of the detection of blockages is an important link in ensuring the safety of citizens. Therefore we endeavor to build models to predict the blockages.

4. Description of the Method

Basically, our approach to the problem of blockage in the fire units is as follows: a large collection of detailed data from Incident Data Reporting System (IDRS) is selected. Then the attributes describing the incidents are quantized, combined or generalized creating different views of the data. These views allow to create granules of similar incidents which is technically done by clustering.

A respective operating time distribution is attached to each of the granules. The aim is to obtain the probability distribution of time needed to handle each distinct (in granule sense) emergency situation. Once the probability distribution is found, it becomes possible to estimate in real time when the gone engines are expected to return to the fire station. This can be further extended to monitor the blockage probability when the fire station gets short on the engines' reserves.

In most countries all incidents involving Fire Service are collected in some sort of Incident Data Reporting System (IDRS). The following are examples of IDRS in various countries: ONTIKA in Finland (Rahikainen and Keski-Rahkonen, 2004), IMS in the UK (Department, 2008), NFIRS in the USA (Administration, 2002) or EWID in Poland (Abacus, 2001). For years those IDRS collected large amounts of data. Polish IDRS – EWID is a collection of approximately 7 million incidents. This research is based on 0.26 million reports (Masovia province), as only this amount of data was available to the authors.

Basically, the approach is following: large collection of data in EWID is first split into clusters of similar accidents. For each of the clusters the distribution of operating times is calculated. The aim is to obtain the probability distribution of time needed to handle each distinct (in clustering sense) emergency situation. Once the probability distribution is found, it becomes possible to estimate in real time when the gone engines are expected to return to the fire station. This can be further extended to monitor the blockage probability when the fire station gets short on the engines' reserves.

4.1. Granule Generation

EWID is composed of two distinct parts: a) the structured part of predefined database attributes (information table) and b) the natural language (NL) part. The approach was to have both parts clustered separately and then have logical *AND* applied to the clusters in order to increase the clustering quality.

The data from the IDRS were first cleaned and preprocessed using the standard and dedicated methods as described in (Krasuski et al., 2013). The number of attributes that described the incidents was reduced on the basis of expert judgement.

The attributes values were quantized, combined or generalized. For example, different time windows are useful for observing how different threats, which Fire Service tries to deal with, are distributed for the given area. This particular problem may be examined on hourly basis as well as season basis and provides both useful and distinct results.

The structured part was standardized. In order to improve the quality of the clustering, the NL part was lemmatized. In computational linguistics, lemmatization is a process of determining the lemma for a given word, i.e. all the inflexed forms are flattened to their basic forms. This is particularly the issue in 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 (Morfologik, 2013). The dataset was then divided into three groups: fire incidents, local threats and false alarms. Then clustering was performed only on the fire and local threats incidents and the false alarms were excluded.

We chose clustering method which requires that the target number of resulting clusters is defined in advance and the initial clustering was focused at finding this target number. PAM (Van der Laan et al., 2003) was chosen as the clustering algorithm and a sample of 10000 incidents was evaluated. The accuracy of the resulting clusters was determined using Silhouette width $S(i)$ (Rousseeuw, 1987) and Calinski Harabasz index $CH(k)$ (Cali/ński and Harabasz, 1974). The number of the clusters in the experiment was varied from 50 to 700 and the $S(i)$ and $CH(k)$ were calculated. Figure 1 depicts the scores of Average Silhouette width.

The Silhouette width grows with the increase of the number of the clusters. Around the number of 300 clusters, the $S(i)$ stabilizes at 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 the 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 the clusters, to 1 denoting very good quality. Therefore, the value 0.1 achieved in the experiment is considerably low. In the next step, full Silhouette analysis for $k = 300$ clusters was performed to find the clusters with a better quality than just average 0.1 . According to this analysis there is a set of the clusters which have a fairly high value of $S(i)$ and a set of the clusters for which the value of $s(i)$ is below 0 . The negative values represent the incidents which

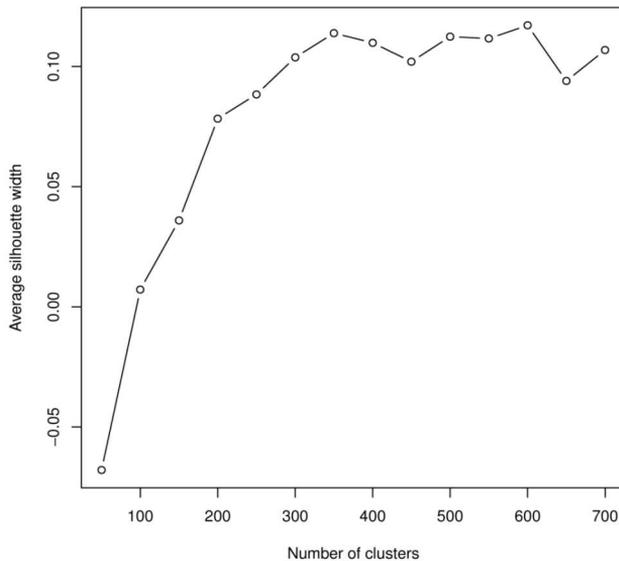


Figure 1. Silhouette width against the number of the clusters (high values are better)

rarely occur and are significantly different from other cases. They weaken the $S(i)$ because there are no similar incidents inside the cluster. Our further analyses were performed on quality clusters only, where $s(i) > 0.3$ with the cardinality of at least 40 (based on an arbitrary judgement).

After determining the target number of the clusters, the full database was clustered. In order to handle large size of the database, CLARA (Kaufman et al., 1990) was used as the clustering method, as this method is designed to overcome memory management issues.

The next stage was the clustering of the NL part of EWID using *Latent Semantic Analysis* (LSA) (Deerwester et al., 1990; Landauer et al., 1998). The basic idea of LSA is to create the 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. Unfortunately, the available resources didn't meet the demands for LSA and this task was not completed on the whole corpora. Therefore LSA was only used within each cluster to check the cohesion of the clusters and if it was needed to divide the clusters into subclusters. Each of the description of the incidents within the given cluster was represented as a vector in LSA space. Next the angles among vectors were compared.

4.2. Name Generation for the Obtained Clusters

Assigning the names to the resulting clusters is a well known issue. An easy approach was taken – three most representative terms were chosen for each cluster (from lemmatized NL part of saved emergencies) and they were joined together by an underscore producing the label. The representativeness of the terms was estimated by the TF-IDF index (Jones, 1993). TF-IDF index is calculated as follows:

$$tf-idf_{i,j} = tf_{i,j} \times idf_i \quad (3)$$

where $tf_{i,j}$ is a term frequency and idf_i is an inverse document frequency. Term frequency is calculated according to the formula (4)

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (4)$$

where $n_{i,j}$ is the number of occurrences of the considered term (t_i) in the description of the incident d_j , and the denominator is the size of the description of the incident $|d_j|$.

Inverse document frequency is calculated according to the formula (5)

$$idf_i = \log \frac{|D|}{|\{j : t_i \in d_j\}|} \quad (5)$$

where $|D|$ – the cardinality of D , or the total number of incidents in the given cluster, $|\{j : t_i \in d_j\}|$ – the number of description of the incidents where the term t_i appears.

Names of the clusters generated in this manner may not be an exact indication of what the clusters really contain. However, the names were meant to be used as replacement for just numbering the clusters in order to provide some hints about the data. Names of the clusters are very important for the visualisation of the engines deployment and interaction between computer system and Control Room Staff (domain experts).

4.3. Regressions of Operating Time Distributions

Having defined the sets of incidents, it becomes possible to fit the distribution of the operating time for each single set. The methodology was as follows: skewness and kurtosis were first calculated for each of the clusters. Skewness and kurtosis are the means of features extraction from the data density/histogram in order to define the family of adequate functions. By inspecting all the results, the group of potential functions was defined: *log-normal*, *gamma*, *weibull*, *chi-squared*, *beta*, *f*. For each of the clusters an iteration of the above six functions was started which

was aimed at finding the best-fitted function. Maximum-Likelihood Estimation (MLE) was used as the criterion of how well the functions are fitted. As a result the distribution of the operating time of each cluster was obtained. Table 3 illustrates the results of the distribution fitting of 16 sampled clusters. Any single result turned out to always favor one of just three functions: log-normal, gamma, weibull, so only these are recorded in the table.

Table 3
Calculated operating time distributions for the clusters

	Cluster name	<i>DT</i>	<i>P1</i>	<i>P2</i>	<i>N</i>	<i>s_i</i>
1	forest(adj.)_forest(sub.)_brushwood	L	4.50	0.54	550	0.74
2	grass_railway_embankment	L	3.53	0.61	767	0.43
3	embankment_railway_grass	L	3.84	0.66	61	0.45
4	grass_gazebo_railway	L	4.10	0.70	501	0.59
5	basement_cellar_chamber	W	1.85	80.77	178	0.48
6	forest(sub.)_brushwood_forest(adj.)	L	4.29	0.78	122	0.74
7	gazebo_roof_way	W	6.53	129.14	118	0.51
8	flat_iron_local	G	2.55	0.06	448	0.46
9	elevator_chute_flat	L	3.51	0.50	63	0.49
10	grass_wasteland_dry	G	3.29	0.04	308	0.84
11	newsstand_shop_pavilion	L	3.82	0.54	284	0.46
12	lavatory_water-closet_toi-toi	L	3.17	0.78	465	0.44
13	cowshed_barn_straw	W	2.95	218.15	110	0.45
14	coaldust_hydrogen_auto-ignition	G	2.25	0.03	41	0.60
15	trailer_gazebo_tunnel	G	3.01	0.03	74	0.41
16	hall_palisade_mezzanine	L	3.89	0.59	45	0.42

Symbols used in the table: *DT* – distribution type (L – log-normal, G – gamma, W – weibull), *P1* – first parameter for the distribution (μ – for log-normal, k for gamma and weibull), *P2* – second parameter for the distribution (σ for log-normal, θ for gamma and λ for weibull), *N* – number of incidents inside given cluster, *s_i* – Silhouette width for given cluster.

4.4. The Blockage

The blockage occurs when all engines are out and a new emergency arises. It is possible to estimate the probability of such risk in real time and perhaps counteract by requesting engines from other fire stations. The proposed model is the composition of three independent elements: a) probability that none of the engines return within critical time interval, b) probability that a new emergency will occur in a given time of the day, c) probability that a given number of emergencies for a day will occur.

4.4 a) Probability that None of the Engines Return Within Time Interval

The best description of the approach will be an imaginary example scenario: There are three fire engines in the fire station. At 00:00 engine-A is destined to handle emergency-A. Then at 00:20 engine-B is sent to emergency-B and finally engine-C is allocated to emergency-C at 01:40 leaving the fire station without any reserves. The question arises whether it is likely for any of the engines to return before a new emergency arises, commencing the blockage.

By providing the characteristics of A, B, C emergencies (which can be done right away or by gradual collection of facts from the field given by the officer in charge) it is possible to relate A, B, C to the respective clusters by similarity of the features. All clusters have their operating time distribution probabilities defined and by choosing the right clusters, the expected A, B, C distributions can be obtained. As a result engines A, B, C should be returning according to the distributions depicted on Figure 2.

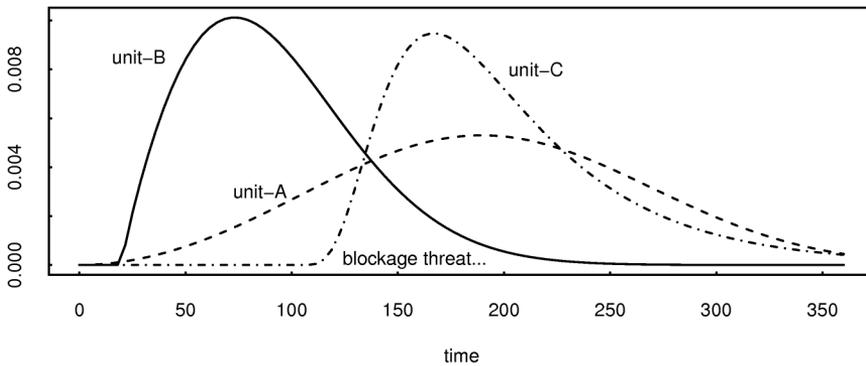


Figure 2. All engines gone creating blockage

The blockage probability commences when the last of the engines, engine-C leaves the fire station. The other end of the blockage interval needs to be calculated by taking into account the decreasing probability that none of the engines return in any next moment in time. Formula (6) was used and the resulting curve is depicted by Figure 3(a). The plot starts from the value 100, as it relates to the time 01:40, at which the last engine left the fire station.

$$P_N(t) = (1 - F_A(t))(1 - F_B(t))(1 - F_C(t)) \tag{6}$$

where $P_N(t)$ is the probability that none of the engines returns before time t and the $F(t)$ are the Cumulative Distribution Functions for the events that the respective engines return.

4.4 b) Probability that a New Emergency Within a Given Time of the Day Will Occur

Based on EWID data, the histogram of how often emergencies occur in a 24-hours period was obtained (Figure 3(b)). The sine function $F_D(t) = I - A\sin(\omega t + \psi)$ was chosen for the model. The regression resulted in obtaining the following function:

$$F_D(t) = 0.042 - 0.031\sin(0.273t - 1.56) \tag{7}$$

The probability $P_D(t)$ that a new emergency will occur in a given time of the day can be calculated as follows:

$$P_D(t) = \int_{t_0}^t F_D(t) \tag{8}$$

4.4 c) Probability that a Number of Emergencies Per Day Will Occur

Another factor to be taken into account is the number of emergencies per day for a given station. The log-normal distribution fitted best (according to MLE) to EWID data (Figure 3(c)). This probability should be regarded as a chance for occurring *n or more emergencies* rather than a fixed number of *n emergencies* and is expressed by formula (9). It is *n* that allows to judge how likely it is to have a new emergency for the given fire station/area: for provincial fire stations the threat of the blockage is considerably less likely than for a high traffic, big city fire stations, because the first ones typically deal with fewer emergencies per day.

$$P_E(n) = 1 - F_E(n) \tag{9}$$

where $P_E(n)$ is the probability that *n* or more emergencies per day will occur and the $F_E(n)$ is the respective Cumulative Distribution Function.

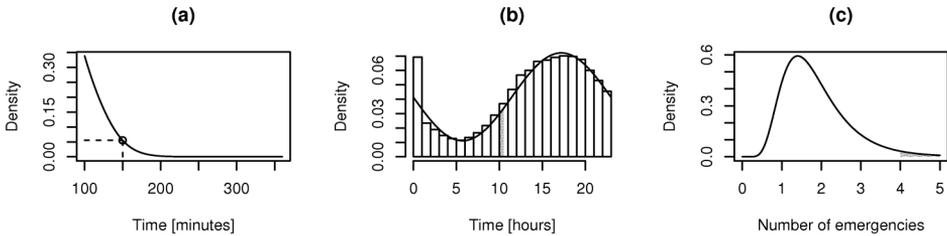


Figure 3. Blockage probability as the composition of three elements

By combining the three independent elements, the final formula for the blockage can be obtained Formula (10).

$$P_B = P_N(t) \times P_D(t) \times P_E(n) \tag{10}$$

5. Experimental Verification

In this section we evaluate the model of blockages prediction empirically. The data for the experiment were selected from IDRS and they represent the real situations, which occurred in the Warsaw Fire Station No. 1. We analysed only those cases when all fire engines were away from the station, involved in F&R operations. Out of the 615 selected cases, 122 finished with blockages and 493 were non-blockage cases. The main goal of the experiment was to predict the actual blockage.

The incidents in the set were labeled “B” for blockage and “N” for non-blockage. The set was divided into two subsets: the training set included 95 blockages and 368 non-blockages, and the test set included 27 blockages and 125 non-blockages.

The first step of the experiment was to determine which value of P_B (formula (10)) separates best between blockages and non-blockages.

The calibration was performed on the training set. Figure 4a) outlines the density distributions functions of P_B for blockages and non-blockages. According to the picture, there is no evident point which separates the two curves. Therefore ROC curve is used to determine the border value (Figure 4b). Based on the AUC function we looked for the point which is characterised by the greatest growth of True Positives (TP) with smallest growth of True Negatives (TN). As the result of this analysis we chose the value of $P_B = 0,0138$ which is a trade-off between the benefit and the cost.

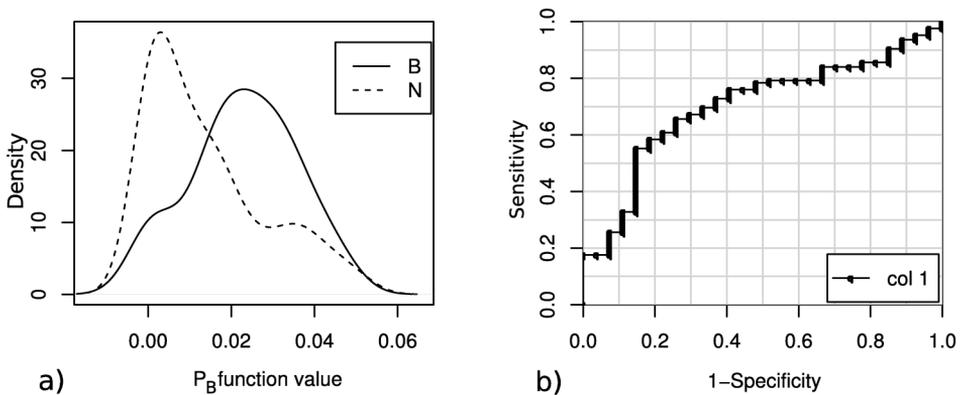


Figure 4. The density distributions of (B)blockage and (N)non-blockage and ROC curve for the training set

Following are the results of validating the model against the test set.

Table 4

Confusion matrix of the prediction model

	Real Positive	Real Negative
Classified Positive	22 (TP)	52 (TN)
Classified Negative	5 (FP)	73 (FN)

Calculated measures:

$$Accuracy = \frac{tp + fn}{tp + tn + fp + fn} = 0.625 \quad (11)$$

$$Recall = \frac{tp}{tp + fp} = 0.814 \quad (12)$$

First measure (Accuracy) evaluates how robust model is in predicting both the situation where there was a real blockage as well as where there was not. However, for us (due to potential losses) more important are the situations where there was a blockage and the model did not predict it. The second measure (recall) reflects the second case. In this case 81% of blockages were accurately predicted.

6. Conclusions

The foundations of decision support system for the blockage probability was presented in the article. The evaluation was done in the probability domain, by combining three elements: a) the probability that none of the engine will return within the given time interval (when all engines are out), b) the probability that at the given part of the day a new emergency will occur and c) the probability that a number of emergencies will occur on a given day. As it is usually the case with such models, there is a possibility to extend the model by including additional parameters (e.g. weather conditions, commanders experience (Krasuski et al., 2012)), but chosen three factors seem reasonable in authors' opinion and were considerably easily available from EWID database.

The proposed method is considerably easy to implement on computer systems. Although the clustering needs to be periodically renewed and this is a computation intensive task, this needs to be done just occasionally, perhaps once a month. However, the actual calculating of the blockage probability (the supposed everyday usage) is trivial and can be done within

seconds. It means that the algorithm can be implemented on workstations at the control room to support decision making when it comes to managing the blockage or relocating the engines. It might be more purposeful to have such a system running at the district level, where the operators are capable of complementing the missing resources from other stations. Additionally, the probability of the blockage can be calculated for the time of the travel of such a complementing engine – perhaps this probability will be low enough for the complementation to be unnecessary.

R E F E R E N C E S

- Abacus (2001). Ewidencja zdarzeń – EWID99. Technical report, Abacus, http://www.ewid.pl/?set=rozw_ewid&gr=roz. [23.04.2007].
- Administration, U. S. F. (2002). National Fire Incident Reporting System. Quick reference guide. Technical report, National Fire Data Center.
- Beckmann, P. (1968). *An introduction to elementary queuing theory and telephone traffic*. The Golem Press, Boulder, Colo.
- Caliński, T. and Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-Theory and Methods*, 3(1):1–27.
- Church, R., Sorensen, P., and Corrigan, W. (2001). Manpower deployment in emergency services. *Fire technology*, 37(3):219–234.
- Clover, F. (1990). Tabu search – part 1. *ORSA Journal on Computing*, 1(2):190–206.
- Clover, F. (1990). Tabu search – part 2. *ORSA Journal on Computing*, 2(1):4–32.
- Deerwester, S., Dumais, S., Furnas, G., Landauer, T., and Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American society for information science*, 41(6):391–407.
- Department, I. T. (2008). Incident Management Systems (IMS). User Guide. Technical report, London Fire Brigade.
- Fink, A. and Reiners, T. (2006). Modeling and solving the short-term car rental logistics problem. *Transportation Research Part E: Logistics and Transportation Review*, 42(4):272–292.
- Hasofer, A., Beck, V., and Bennetts, I. (2007). *Risk analysis in building fire safety engineering*. Butterworth-Heinemann.
- Jones, K. (1993). A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, 28(1):11–21.
- Kaufman, L., Rousseeuw, P., and Corporation, E. (1990). *Finding groups in data: an introduction to cluster analysis*, volume 39. Wiley Online Library.
- Krasuski, A., Kreński, K., and Lazowy, S. (2012). A Method for Estimating the Efficiency of Commanding in the State Fire Service of Poland. *Fire Technology*, 48(4):795–805.

- Krasuski, A., Ślęzak, D., Kreński, K., and Łazowy, S. (2013). Granular Knowledge Discovery Framework. *New Trends in Databases and Information Systems*, 185:109–118.
- Landauer, T., Foltz, P., and Laham, D. (1998). An introduction to latent semantic analysis. *Discourse processes*, 25(2):259–284.
- Lau, H., Ho, G., Zhao, Y., and Hon, W. (2010). Optimizing patrol force deployment using a genetic algorithm. *Expert Systems With Applications*, 37(12):8148–8154.
- Lee, S., Franz, L., and Wynne, A. (1979). Optimizing state patrol manpower allocation. *Journal of the Operational Research Society*, pages 885–896.
- Li, Z. and Tao, F. (2010). On determining optimal fleet size and vehicle transfer policy for a car rental company. *Computers & operations research*, 37(2):341–350.
- Morfologik, P. (2013). Morfologik – About the project. <http://morfologik.blogspot.com/2006/05/about-project.html>.
- Nemhauser, G. and Wolsey, L. (1988). *Integer and combinatorial optimization*, volume 18. Wiley New York.
- Ormeci, E. and Burnetas, A. (2004). Admission control with batch arrivals. *Operations Research Letters*, 32(5):448–454.
- Peace, D. M. S. (2001). Planning new standards of fire service emergency cover for the United Kingdom. *Fire technology*, 37(3):279–290.
- Rahikainen, J. and Keski-Rahkonen, O. (2004). Statistical determination of ignition frequency of structural fires in different premises in Finland. *Fire technology*, 40(4):335–353.
- Rousseeuw, P. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65.
- Shonick, W. and Jackson, J. (1973). An improved stochastic model for occupancy-related random variables in general-acute hospitals. *Operations Research*, pages 952–965.
- Taylor, P. and Huxley, S. (1989). A break from tradition for the San Francisco police: Patrol officer scheduling using an optimization-based decision support system. *Interfaces*, pages 4–24.
- Tillander, K. and Keski-Rahkonen, O. (2008). The influence of fire department intervention to the fire safety of a building assessed using fire risk analysis. In *Proceedings of the 3rd International Conference on Performance-Based Codes and Fire Safety Design Methods*, pages 247–256.
- Van der Laan, M., Pollard, K., and Bryan, J. (2003). A new partitioning around medoids algorithm. *Journal of Statistical Computation and Simulation*, 73(8):575–584.