TECHNICAL TRANSACTIONS 5/2017
CZASOPISMO TECHNICZNE 5/2017
CIVIL ENGINEERING
DOI: 10.4467/2353737XCT.17.065.6422

Wojciech Drozd (wdrozd@izwbit.wil.pk.edu.pl)
Institute of Construction and Transportation Engineering and Management, Faculty of Civil Engineering, Cracow University of Technology

Identifying and profiling the patterns of construction accidents using affinity analysis

Identyfikacja i profilowanie wzorców wypadków budowlanych przy pomocy analizy koszykowe

Abstract

The construction site and its elements create circumstances that are conducive to the formation of safety risks during the execution of works. Analysis indicates the critical importance of these factors in the set of characteristics that constitute the causes of accidents in the construction industry. The main substantive tasks in this article include isolating patterns of accidents on the site and identifying the analysed characteristics that are important in defining these patterns. In terms of methodology, the paper presents affinity analysis as the method of analysing data resources. The research was carried out on the basis of data from the register kept by the District Labour Inspectorate in Krakow (2014–2016).

Keywords: construction site, accidents at work, affinity analysis

Streszczenie

Teren budowy i jego elementy stwarzają okoliczności, które sprzyjają powstawaniu zagrożeń bezpieczeństwa pracy w realizacji robót budowlanych. Analizy wskazują na decydujące znaczenie tych czynników w zbiorze cech opisujących przyczyny wypadków w budownictwie. Zasadnicze zadania merytoryczne w artykule obejmują wyodrębnienie wzorców wypadków na terenie budowy i wskazanie na te spośród analizowanych cech, które mają istotne znaczenie w definiowaniu tych wzorców. W zakresie metodologicznym wykorzystano analizę zasobów danych za pomocą analizy koszykowej. Badania zrealizowano na podstawie danych z rejestru prowadzonego przez Okręgowy Inspektorat Pracy w Krakowie (2014–2016).

Słowa kluczowe: teren budowy, wypadki przy pracy, analiza koszykowa

1. Introduction

accidents at work, including those on construction sites, are random occurrences which are difficult or impossible to predict. Therefore, studying them and identifying relationships between the traits which characterise them is not easy [3, 4, 5, 6]. In particular, this applies to studies in which focus on the factors which give rise to negative consequences for human life and health. This article attempts to identify and profile accident patterns, based on data from the register maintained at the District Labour Inspectorate, in Krakow. The analysed events included accidents which occurred between the years 2014 and 2016. The number of observations was sixty-five. The main goal was focused on isolating patterns of construction accidents and identifying those of the analysed characteristics that are important in defining these patterns. The methodology included the analysis of data resources using conceptual grouping in the form of affinity analysis.

There are a variety of data mining techniques which indicate new dependence in the collection of data. One of these techniques is the discovery of association, which is the most popular example of affinity analysis – this is a method which was designed for research of customer preferences (what put in the cart). This article presents a successful attempt to use this method in the field of research patterns of construction accidents.

2. Affinity analysis

affinity analysis [1, 2] is used to find hidden dependencies in a large data set in the form of simple rules - so-called association rules: IF [body] THEN [head]. They are written using the conditions: [body conditions] => [head conditions]. Commas used in writing body conditions or head conditions represent the conjunction 'and'.

When selecting the input data format, it must be remembered that there are many potential rules. For example, in the case of three dichotomous variables ('YES/NO' responses to three questions), we can receive up to 650 rules – this is the number of permutations for the three variables and two possible values. Of course, we are interested only in those rules which often occur in the analysed data, i.e. those which describe frequently occurring patterns. Therefore, to isolate the rules that carry important information, we use three parameters used for the statistical evaluation of the validity of the rules:

- **Support** – the percentage of instances of the rule in the analysed dataset;
- **Confidence** – the percentage of instances of the rule in a set of observations that contain the body (for the rule A=>B, this corresponds to the conditional probability $P[B|A]$);
- **Lift** – determines how the occurrence of the body increases or decreases the occurrence of the head. Values higher (or lower) than 1 indicate that in the set of observations for which the body occurs, the likelihood of the occurrence of the head is higher (or lower) than in the whole dataset in general.

Strictly speaking, criteria for the assessment of the rules found are calculated using the following formulas where 'number of transactions' refers to the number of occurrences of a given pattern, i.e. the number of accidents with the given characteristics:

$$\text{support}(A) = \frac{\textit{cardinality}\ (A)}{\text{the number of transactions in the data set}} \tag{1}$$

$$\text{support}\ (\textit{if}\ A,\ \textit{it}\ C) = \frac{\textit{cardinality}\ (A,C)}{\text{the number of transactions in the data set}} \tag{2}$$

$$\textit{confidence}\ (\textit{if}\ A,\ \textit{it}\ C) = \frac{\text{support}\ (\textit{if}\ A,\ \textit{it}\ C)}{\text{support}\ (A)} \tag{3}$$

$$\textit{lift}\ (\textit{if}\ A,\ \textit{it}\ C) = \frac{\textit{confidence}\ (\textit{if}\ A,\ \textit{it}\ C)}{\text{support}\ (C)} \tag{4}$$

3. Selection of variables

the affinity analysis was performed using the SAL module (sequence, associations and link analysis) in Statistica. Properly prepared data on construction accidents from the register of the District Labour Inspectorate in Krakow were used. The created dataset *'Accidents 2014 – 16'* is shown in Tables 1a to 1c (this is limited to showing only 2 of the 65 observations).

Table 1a. Accidents 2014–2016

No.	Sw	Tw	Wi (years)	Stp (years)	Ldn (years)	Sz	Zwd	Mw
1	L	P	39	8	16	UPO	DIGGER OPER.	BUILDING
2	C	P	40	13	16	DG	CONSTR. WOR.	BUILDING

Source: own study

Table. 1b. Accidents 2014–2016

Pp	Cw	Cmc	Wdo	Cmo	Wyu	Cmu
RZ	PNZ	RM	POR	RM	POR	EI
TYNK	PNN	RU	UW	RU	UNO	PP

Source: own study

Table 1c. Accidents 2014–2016

CAUSE														
BK	BŚI	BŚZ	NCM	NPD	POŚ	PNO	PUZ	SPF	SZ	TN	UKM	WZ	ZN	ZO
						1	1			1		1		1
	1	1	1					1	1					1

Source: own study

Where:

Ldn – The number of days of incapacity

Mw – Place of accident

Pp – Work process (RZ – earthworks, TYNK – plastering, MUR – masonry, PRZEB – reconstruction, MON – assembly, DESK – formwork, DEM – dismantling, DACH – roofing, INST – installation, TRANSP – transport, BET – concreting)
Stp – Seniority of the injured person
Sw – Effects of accident
Tw – Type of accident (individual, group)
Wi – Age of the injured person
Zwd – Occupation of the victim (CARPENTER - JOINER, ROOFER, ELECTRICIAN, TILER - MASON, GEOLOGIST, FORKLIFT TRUCK DRIVER, HOUSE PAINTER, GAS FITTER, PLUMBER, BRICKLAYER - PLASTERER, BACKHOE OPERATOR, TOWER CRANE OPERATOR, GENERAL CONSTRUCTION WORKER, WELDER, STEEL FIXER - CONCRETER)
Cmc – Material factor associated with the activity performed by the injured person at the time of the accident
Cmo – Material factor associated with the deviation
Cmu – Material factor which is the source of injury
DP – Portable ladder
EB – Building element
EI – Element of the system
EN – Power tool
ŁO – Shovel
MB – Building materials
OD – Chippings
PP – Horizontal surface (ground)
RM – Moving machinery and equipment
RU – Scaffolding
SP – Falling object
SZW – Lift shaft
ŚC – Wall
UR – Excavated material
Cw – Activity performed by the injured person at the time of the accident
DEM – Dismantling building elements
KN – Contact with tool
PNN – Work with non-mechanised tool
PNZ – Work with mechanised tool
PODN – Manual handling (lifting, lowering)
SCH – Movement (going downwards, upwards, through)
SM – Storage method
UK – Loss of control over tool
Sz – Employment status of the injured person
DG – Economic activity
UPN – Employment contract for an indefinite period

UPO – Employment contract for a definite period
UZ – Work order contract
Wdo – An event which is a deviation from the normal state
BK – Lack of required qualifications (no medical examination, training)
BŚI – Lack of personal protection equipment (improper use)
BŚZ – Lack of collective protection
NCM – Improper condition of the material factor
NPD – Improper passageways and access paths
PNO – Performing extra work which does not fall within the scope of duties
POŚ – Slipping and stumbling
PUZ – Carrying out work without removal of threats
SPF – Improper psycho-physical condition (alcohol, fatigue)
SZ – Arbitrary behaviour of the employee, (lack of supervision, lack of concentration)
TN – Tolerance of deviations from safety rules by management
UKM – Loss of control over the machine
WZ – Walking or driving into a risk area
ZN – Emergency
ZO – Improper organisation of work
Wyu – Event causing injury
OE – Cut
OMS – Overloading the musculoskeletal system
POR – Electric shock
PRZ – Crushing
UNO – Impact inanimate objects
UPZ – Fall into depression
URU – Getting hit or captured by moving elements or elements in operation
USP – Getting struck by falling objects
UW – Falling from height
WY – Explosion
ZAS – Getting buried in an excavation

4. Accident profiles

the analysis specifies the following areas:

a) **Accident profiles** – a list of events that occur together in the incidents. The proposed method of sorting data can be easily changed and saved, e.g. according to support (size of profile occurrences/number of accidents); see Table 2.

Table 2. Selected accident profiles

No.	Popular kits	The numbers of elements	Cardina-lity	Support %
1	(Wdo==UW, Wyu==UNO, Cmu==PP, SZ)	4	20	30.77
2	(Mw==BUDOWA, Wdo==UW, Wyu==UNO, Cmu==PP)	4	16	24.62
3	(Wdo==UW, Wyu==UNO, Cmu==PP, BŚI)	4	16	24.62
4	(PUZ, Wdo==UW, Wyu==UNO, Cmu==PP)	4	15	23.08
5	(Cw==PNN, Wdo==UW, Wyu==UNO, Cmu==PP)	4	14	21.54
6	(Wdo==UW, Cmu==PP, BŚI, SZ)	4	13	20.00

Source: own study

The values shown in Table 2 indicate that in certain sizes, out of the sixty-five accidents examined, the following accident profiles occurred:

- In 20 out of 65: falling from a height ÷ impact inanimate objects ÷ horizontal surface ÷ arbitrary behaviour of the employee, as the cause. Support for this pattern is 30.77% (20/65),
- In 16 out of 65: construction site ÷ falling from a height ÷ impact inanimate objects ÷ horizontal surface. Support for this pattern is 24.62% (16/65),
- In 16 out of 65: falling from a height ÷ impact inanimate objects ÷ horizontal surface ÷ lack of personal protection equipment. Support for this pattern is 24.62% (16/65),
- In 15 out of 65: carrying out work without removal of threats ÷ falling from a height ÷ impact inanimate objects ÷ horizontal surface. Support for this pattern is 23.08% (15/65),
- In 14 of out 65: work with non-mechanised tool ÷ falling from a height ÷ impact inanimate objects ÷ horizontal surface. Support for this pattern is 21.54% (14/65),
- In 13 out of 65: falling from a height ÷ horizontal surface ÷ lack of personal protection equipment ÷ arbitrary behaviour of the employee. Support for this pattern is 20.00% (13/65).

The frequency of the most common accident profiles is shown in Figure 1.

b) **Association rules** – all rules discovered, meeting the relevant criteria: support of at least 20%, confidence of at least 10%, lift greater than 1 (minimum values have been chosen so as not to include patterns that occur only a few times).

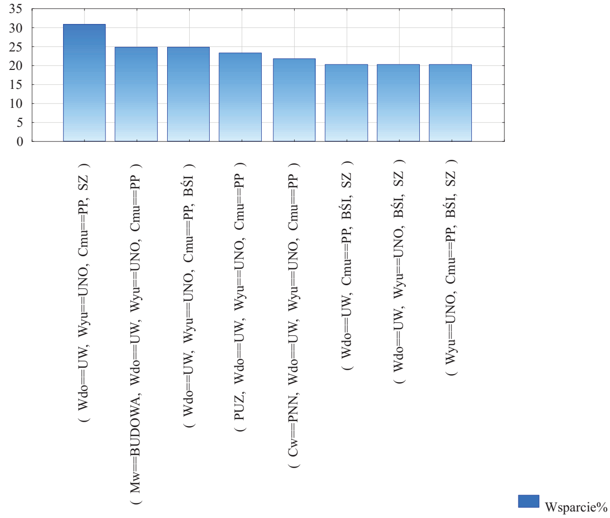

Fig. 1. The frequency of accident profiles
Source: own study

Table 3. Summary association rules for the selected items

<table>
<tr><th>No.</th><th>Predecessor</th><th>==></th><th>Consequent</th><th>Support %</th><th>Trust %</th><th>Increase %</th></tr>
<tr><td>1</td><td>Wdo==UW, Cmu==PP</td><td>==></td><td>BŚI, SZ</td><td>20.00</td><td>50.00</td><td>2.03</td></tr>
<tr><td colspan="7">Support: Accidents in which falling from height is the a deviation from the normal state, the horizontal surface is the material factor causing injury and the cause is the arbitrary behaviour of the employee and the lack of personal protection equipment account for 20.00% of all accidents examined.
Confidence: If falling from height is the a deviation from a normal state and the horizontal surface is the material factor causing injury, the likelihood that the accident occurred as
a result of arbitrary behaviour of the employee and the lack of personal protection equipment is 50%.
Lift: If we know that falling from height is the a deviation from a normal state, the horizontal surface is the material factor causing injury, the likelihood that the accident occurred as a result of arbitrary behaviour of the employee and the lack of personal protection equipment is 2.03 times greater than in the entire dataset (not taking into account the information in the body).</td></tr>
<tr><td>2</td><td>Mw==BUDOWA, Wdo==UW</td><td>==></td><td>Wyu==UNO</td><td>24.62</td><td>100.00</td><td>2.50</td></tr>
<tr><td colspan="7">Support: Accidents occurring during the construction (erection) of a building, in which falling from height is the a deviation from a normal state and the injury is caused by impact inanimate objects, account for 24.62% of all accidents examined.
Confidence: If the accident takes place at the construction site and falling from height is the a deviation from normal state, the likelihood that the injury is caused by impact inanimate objects is 100%.
Lift: If we know that the accident occurred during the construction of a building and falling from height is the a deviation from a normal state, the likelihood that the injury is caused by impact inanimate objects is 2.5 times higher than in the entire dataset (not taking into account information in the body).</td></tr>
<tr><td>3</td><td>Wdo==UW, Cmu==PP, BŚI</td><td>==></td><td>Wyu==UNO</td><td>24.62</td><td>100.00</td><td>2.50</td></tr>
</table>

Support: Accidents in which falling from height is the a deviation from a normal state and the horizontal surface and the lack of personal protection equipment is the material factor causing injury and the causing event is impact inanimate objects, account for 24.62% of all accidents examined. **Confidence:** If accidents in which falling from height is the a deviation from a normal state and the horizontal surface and the lack of personal protection equipment is the material factor causing injury, the likelihood that the causing event is impact inanimate objects is 100%. **Lift:** If we know that falling from height is the a deviation from a normal state and the horizontal surface and the lack of personal protection equipment is the material factor causing injury, the likelihood that the event causing the injury impact inanimate objects is 2.5 times greater than in the entire dataset (not taking into account the information in the body).

Source: own study

c) **Rules diagram** – the following diagram graphically illustrates the designated relationships between the individual categories of the considered variables. Relationships with a high level of support and confidence correspond to larger and darker points (Fig. 2).

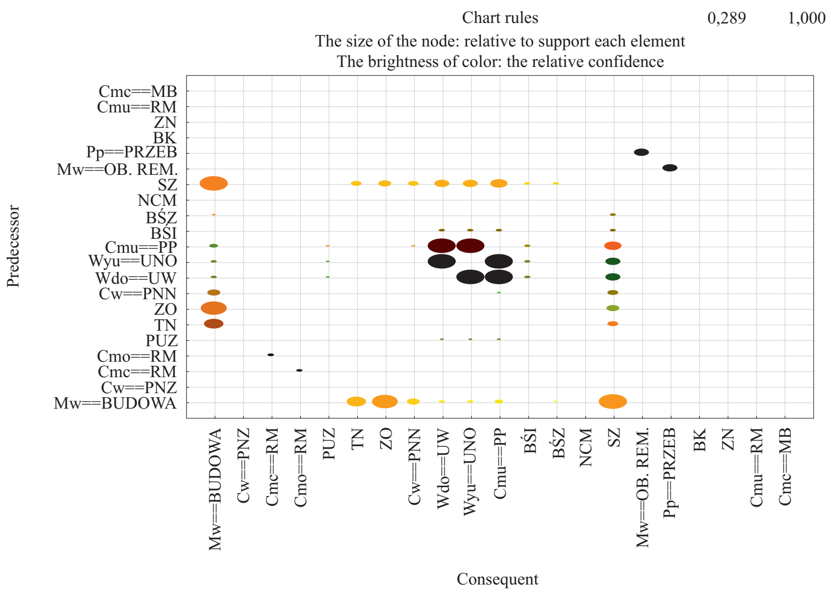

Fig. 2. Chart rules.
Source: own study

The diagram shows particularly strong rules (in terms of statistical criteria). These are presented in Table 4.

Table 4. The strongest association identified in the chart rules

Predecessor	==>	Consequent	Support %	Trust %	Increase %
Cmu==PP	==>	Wdo==UW	40.00	96.30	2.41
Cmu==PP	==>	Wyu==UNO	40.00	96.30	2.41
Wdo==UW	==>	Cmu==PP	40.00	100.00	2.41
Wdo==UW	==>	Wyu==UNO	40.00	100.00	2.50
Wyu==UNO	==>	Cmu==PP	40.00	100.00	2.41
Wyu==UNO	==>	Wdo==UW	40.00	100.00	2.50

Source: own study

5. Summary

the research covered construction site accidents which occurred from 2014-2016 in the Lesser Poland province. The observations registered at the District Labour Inspectorate in Krakow made it possible to conduct an analysis, the aim of which was to isolate patterns of construction accidents and to indicate those among the analysed characteristics that are important in defining these patterns. The methodology included the analysis of data resources using conceptual grouping in the form of affinity analysis. An attempt was made to increase the body of knowledge on the development of scientific methods to assess threats on construction sites and investigate the possibility of their practical application in improving the safety of working conditions. These methods are effective tools for identifying patterns in multidimensional sets which characterise construction accidents; they make it possible to create profiles and association rules according to confidence (the percentage of instances of the rule in a set of observations containing the body). These types of subgroups of related rules exist in the analysis results. The resulting rules essentially apply to the variables themselves, which occur once in the body and once in the head – they describe the same accident profile. These rules should be considered together and should not be separated; they are dominated by falling from height, impact inanimate objects and horizontal surfaces. The strongest relationships are as follows:

Predecessor	==>	**Consequent**
Cmu==PP	==>	Wdo==UW
Cmu==PP	==>	Wyu==UNO
Wdo==UW	==>	Cmu==PP
Wdo==UW	==>	Wyu==UNO
Wyu==UNO	==>	Cmu==PP
Wyu==UNO	==>	Wdo==UW

References

[1] Agrawal R., Imielinski T., Swami A., *Mining association rules between sets of items in large database,* Proceedings of the ACM SIGMOD Conference on Management of Data, 1993, 207–216.

[2] Breiman L., Friedman J., Stone C.J., Olshen R.A., *Classification and Regression Trees,* Wadstworth Statistics/Probability, 1984.

[3] Drozd W., *Regresion analysis of accident absenteeism and variables describing working conditions,* Recent advances in civil engineering: Construction management – Inżynieria Lądowa, Politechnika Krakowska, Kraków 2015, 13–27.

[4] Drozd W., *Charakterystyka terenu budowy w aspekcie zagrożeń bezpieczeństwa pracy,* Czasopismo Inżynierii Lądowej, Środowiska i Architektury, Politechnika Rzeszowska, 01/03/2016, 165–172.

[5] Hoła B., *Modelowanie jakościowe i ilościowe wypadkowości w budownictwie,* Politechnika Wrocławska, Wrocław 2008.

[6] Hoła B., *Bezpieczeństwo pracy w procesach budowlanych,* Politechnika Wrocławka, Wrocław 2016.