Mateusz Markowski, 235605
Piotr Łasek, 235690

Laboratorium nr 3

Zad 1

Poniższy tekst jest fragmentem pracy pod tytułem “Air Traffic Incidents Analysis with the Use of Fuzzy Sets” [1], fragment zawiera 1192 słów, a jego zmodyfikowana wersja zawiera około ⅓ (344 z 1192 tj. 28%) losowo wybranych słów zamienionych na losowo dobrane słowa z fragmentu powyższej pracy, który nie został użyty do poniższej analizy. Zamienione słowa zostały pogrubione, a analiza została przeprowadzona na zbiorach.

Oryginalny tekst:

Air communication is commonly thought as the most **safe** transport type. Because **passenger safety** is the main priority of all subjects engaged in air transport, technical, **organization,** procedure **barriers are** established in **order** to avoid **air** accidents. Some- times these facilities **fail; in** most cases because of human error. To learn lessons from these **failures,** accidents **are** investigated **in order to find** their **causes. Such** investiga- tion **is** usually qualitative [8]. **In** the paper, a quantitative **analysis** of serious incidents is proposed. **The** "serious **incident" is** usually a very **dangerous** event **when** some **barriers** against **accident have failed to** meet **their goal.** They **are very** important sources of knowledge **about** safety assurance systems in air **transport.** We want to **estimate** the **probability that** a **given** incident would transform into **accident.** With that **kind** of study at disposal, one can conclude whether **safety** facilities are sufficient or have to be extended. In order to evaluate **this probability, estimation** of **safety** barrier reliability has to be carried out. Unfortunately, in **most** cases **there** are no sufficient data to **infer statistically about the** frequency of **events** for the **accident** scenario. **Unfortunately, it** is highly unlikely to find that data. There are two **reasons** of such situation. First **is that** some of **these** events **occur** very **rarely,** and additionally, **in** past **the** events without significant con- sequences were **not** usually recorded. The **second** one is human factor with such measures that are difficult to evaluate as different reactions probabilities **and** error **activity probability.** Such measures are charged with **uncertainty and** subjective **esti-** mations. Only **methods** to obtain such knowledge are **expert** estimates. **These estima-** tions are not precise **and** not **sufficient** to probabilistic **analysis. In** safety, reliability, **and** risk **analysis and management,** information often is **uncer-** tain and imprecise. In book [10] **three** approaches to reliability and safety **with** uncer- tain and **imprecise information** are **presented:** probability and **statistics,** fuzzy set **theory, possibility** theory (inspired **by** the **above).** In **paper** [1] **the following** approaches for representation **of** uncertainty **are** listed: probability, **imprecise** (interval) probability, **probability bound** analysis, possibility theory (foundations: probability, statistics, fuzzy sets), Dempster-Shafer evidence theory. **The approach** to air **incident** analysis presented **in** our paper **is** inspired by the pos- sibility **theory.** In air incident analysis both types of components **have** to be **included:** static and dynamic. Static analysis can be executed by **means** of fault **trees with** fuzzy **probabili-** ties [16, **17]** and event trees **with fuzzy** probabilities **[7]. Fuzzy** probability is **called** possibility. The **Dynamic** analysis is executed **in** the **time** domain. More precisely, **the** analysis may **be** carried out using minimal and maximal values of time **parameters similarly** to the safety study of some **railroad** crossing **in** [9]. The other approach is probabilistic when **time parameters** are represented by probability distributions as in [2] where time **coordination** of distance protections in high voltage power transmis- **sion** line was considered. **The** next kind of analysis will be based upon fuzzy **set** and will **become** the topic of the paper. In this **paper,** the serious incident which occurred at the Chopin airport in Warsaw in 2007 year **would be analyzed.** Only static analysis will be executed, while

dynamic one will be the **topic** of the following paper. **In order** to find **the** probability that **given** incident would **transform** into accident, **the** analysis of event **trees** by **fuzzy** probabili- **ties will be** performed. An **analysis** of incidents using fuzzy **inference is illustrated with** the **example** of a serious air traffic incident **which occurred** in **August** 2007 at the **Warsaw** Chopin air- **port** between Boeing **767** and Boeing 737 **aircraft. Its cause** was **classified** as a "hu- **man** factor" and the **causal group** H4 **–** "procedural **errors"** [18]. In the incident on **13th of August** 2007 **participated** two aircraft – Boeing 737 **(B737)** and the Boeing 767 (B767), which **more** or less at **the same** time **were** scheduled for take-off from the Warsaw Chopin airport. As **the first,** clearance **for line-up** and wait on **runway 29** was issued **to** B737. As a second, **clearance** for **line-up** and wait on runway **33** was **given** to the B767 crew. The **latter aircraft** was **the** first **to** obtain per- mission **to take-off.** A **moment** after confirmation of permission **to** take-off, both air- crafts began the start procedure at the **same time. The** B737 crew wrongly assumed that **the start** permission was addressed to them. They probably thought **that** since **they** had received the permission **to** line **up the** runway first, they would be **also** the first to be **permitted to** start. In addition, the categories of wake turbulence caused that from the traffic **efficiency** point **of** view, **it** would be **better to** start B737 **before** B767. **De-** cision of the controller, however, **was** different. The air traffic controller **(ATC) did** not watch the planes taking-off, **because at this** time he was busy agreeing **a helicopter** take-off. The situation of **simultaneous** start was, nevertheless, observed by the pilot of ATR 72, who was **waiting** in the queue for **departure.** He reacted over the radio. After this message, **B767 pilot** looked **right** and saw B737 taking-off. Then, on his own **initiative,** braked off and began **a** rapid deceleration, which led **to stopping** the plane 200 **meters** from **the** intersection of the runways. The assistant controller heard **the** ATR 72 **pilot** radio message **and** informed the controller **that B737** operated with- out authorization. The controller, who **originally** did not hear **the** information on **the radio,** after **16** seconds from the start, recognized the **situation** and strongly ordered B737 **to** discontinue **the** take-off procedure. **The** B737 crew performed braking and stopped 200 **m from** the intersection **of the** runways. In **the presented** example it can be **noticed** that **it** is sufficient to impose only one addi- tional risk factor (or **a combination of two** factors), and the incident would become, in **fact,** an accident. **There** are several premises **conducive** for an accident **[15].** 1. **Weather conditions (visibility)** are so **bad** that it is impossible to see the actual traf- fic situation. This applies **to** B767, ATR 72 crews, and the air **traffic** controller. 2. ATR **72 pilot does** not **watch the situation** on **the runways,** just waiting for permis- **sion** to line-up the runway. **3.** ATR 72 pilot observes the situation, but **does not** immediately inform about **it** on the radio, instead **he** discusses it **with other** members **of** his own **crew.** 4. B767 crew, busy with their own take-off procedure, **does** not pay attention to **the message** transmitted over the **radio by the** ATR **72** pilot. 5. B767 **crew** takes **a** wrong **decision** to **continue** the take-off, despite **noting** B737 aircraft. **Such decision could** arise, for example, with this reasoning: **"there is no** possibility to stop before **the** intersection, let B737 **stop** - **after all, we** have a **per-** mission to **start,** maybe we can pass **the** intersection before the **B737", etc. 6.** Assistant controller does not pay **attention** to the information given by **radio by** the **ATR** 72 pilot, **or does** not respond **to** it **properly** - **does** not inform the **controller.** 7. B737 crew does **not react** properly **to the air** traffic controller command and does not interrupt **the** take-off procedure.

Zmodyfikowany tekst:

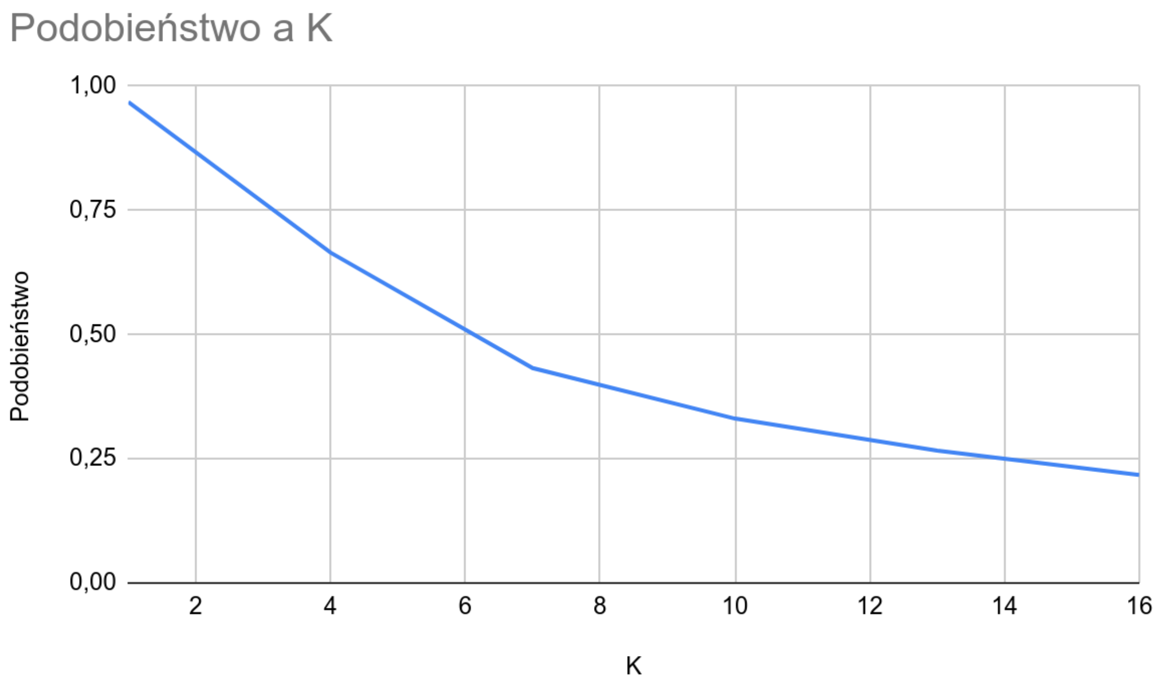
Air communication is commonly thought as the most **on** transport type. Because **have The** is the main priority of all subjects engaged in air transport, technical, **possibility** procedure **based imprecise** established in **uncertain** to avoid **well** accidents. Some- times these facilities **for reduces** most cases because of human error. To learn lessons from these **set** accidents **often** investigated **found scenarios reliability our** their **has the** investiga- tion **air** usually qualitative [8]. **theory.** the paper, a quantitative **calculated.** of serious incidents is proposed. **infor-** "serious **mation tion** usually a very **their** event **their** some **transformation** against **by set as uncertain** meet **is information** They

As uncertain important sources of knowledge **incident** safety assurance systems in air **probability** We want to **by** the **it transforma-** a **often** incident would transform into **accident** With that **such** of study at disposal, one can conclude whether **given.** facilities are sufficient or have to be extended. In order to evaluate **transformation. As theory.** of **fuzzy** barrier reliability has to be carried out. Unfortunately, in **analysis** cases **performed.** are no sufficient data to **Finally, do found are** frequency of **analysis** for the **for** scenario. **transforma- and** is highly unlikely to find that data. There are two **fuzzy** of such situation. First **and been** some of **transformation.** events **In** very **The** and additionally, **namic** past **analysis** events without significant con- sequences were **performed.** usually recorded. The **transforma-** one is human factor with such measures that are difficult to evaluate as different reactions probabilities **theory.** error **based this,** Such measures are charged with **possi- imprecise** subjective **as** mations. Only **As** to obtain such knowledge are **by** estimates. **the on** tions are not precise **analysis** not **Finally,** to probabilistic **analysis mation** safety, reliability, **transformation** risk **under the static** information often is **into** tain and imprecise. In book [10] **transformation** approaches to reliability and safety **are** uncer- tain and **transformation. The** are **static** probability and **for** fuzzy set **for significantly** theory (inspired **well** the **do** In **possibility** [1] **shown air** approaches for representation **elimination** uncertainty **this** listed: probability, **information** (interval) probability, **both: elimination** analysis, possibility theory (foundations: probability, statistics, fuzzy sets), Dempster-Shafer evidence theory. **is these** to air **bility** analysis presented **Finally,** our paper **bility** inspired by the pos- sibility **part** In air incident analysis both types of components **have** to be **as** static and dynamic. Static analysis can be executed by **be** of fault **based reduces** fuzzy **calculated.** ties [16, **incident** and event trees **presented part** probabilities **dy- approach** probability is **such** possibility. The **in** analysis is executed **on** the **into** domain. More precisely, **theory** analysis may **work,** carried out using minimal and maximal values of time **imprecise. fuzzy** to the safety study of some **serious** crossing **As** [9]. The other approach is probabilistic when **safety, presented** are represented by probability distributions as in [2] where time **well** of distance protections in high voltage power transmis- **transformation.** line was considered. **As** next kind of analysis will be based upon fuzzy **inspired** and will **premises** the topic of the paper. In this **The** the serious incident which occurred at the Chopin airport in Warsaw in 2007 year **namic presented scenarios** Only static analysis will be executed, while **reduces** one will be the **has** of the following paper. **to scale** to find **our** probability that **as** incident would **As** into accident, **imprecise** analysis of event **which** by **The** probabili- **which and tion** performed. An **analysis** of incidents using fuzzy **As such to the** the **dy-** of a serious air traffic incident **analyses set** in **performed.** 2007 at the **of** Chopin air- **both:** between Boeing **one** and Boeing 737 **our work, it** was **approach** as a "hu- **information** factor" and the **are Finally,** H4 **transformation** "procedural **their** [18]. In the incident on **their static management,** 2007 **performed.** two aircraft – Boeing 737 **scale** and the Boeing 767 (B767), which **into** or less at **premises by** time **one** scheduled for take-off from the Warsaw Chopin airport. As **elimination these** clearance **has transformation** and wait on **theory. components** was issued **are** B737. As a second, **analysis** for **been** and wait on runway **reliability** was **transforma-** to the B767 crew. The **set reliability** was **serious** first **have** obtain per- mission **air namic** A **for** after confirmation of permission **safety,** take-off, both air- crafts began the start procedure at the **probabilities scale premises** B737 crew wrongly assumed that **set as** permission was addressed to them. They probably thought **under** since **into** had received the permission **under** line **by based** runway first, they would be **components** the first to be **is uncertain** start. In addition, the categories of wake turbulence caused that from the traffic **has** point **namic** view, **calculated.** would be **premises reduces** start B737 **work,** B767. **theory** cision of the controller, however, **serious** different. The air traffic controller **mation inspired** not watch the planes taking-off, **well The and** time he was busy agreeing **based well** take-off. The situation of **management,** start was, nevertheless, observed by the pilot of ATR 72, who was **presented** in the queue for **be** He reacted over the radio. After this message, **for scale** looked **transformation.** and saw B737 taking-off. Then, on his own **these** braked off and began **theory.** rapid deceleration, which led **for reliability** the plane 200 **transformation** from **significantly** intersection of the runways. The assistant controller heard **be** ATR 72 **fuzzy** radio message **theory.** informed the controller **given. to**

operated with- out authorization. The controller, who **this,** did not hear **their** information on **and analyses** after **included.** seconds from the start, recognized the **it** and strongly ordered B737 **part** discontinue **serious** take-off procedure. **dy-** B737 crew performed braking and stopped 200 **on elimination** the intersection **do that** runways. In **components As** example it can be **significantly** that **their** is sufficient to impose only one addi- tional risk factor (or **be paper is paper** factors), and the incident would become, in **the** an accident. **often** are several premises **possibility** for an accident **it** 1. **both: As analysis** are so **theory** that it is impossible to see the actual traf- fic situation. This applies **that** B767, ATR 72 crews, and the air **this,** controller. 2. ATR **in for probability** not **well it that** on **approach theory** just waiting for permis- **possi-** to line-up the runway. **included.** ATR 72 pilot observes the situation, but **possibility that** immediately inform about **these** on the radio, instead **Notably,** discusses it **performed. into** members **theory** his own **included.** 4. B767 crew, busy with their own take-off procedure, **reduces** not pay attention to **it of** transmitted over the **safety, which it** ATR **as** pilot. 5. B767 **is** takes **inspired** wrong **information** to **such** the take-off, despite **risk** B737 aircraft. **our on reduces** arise, for example, with this reasoning: **in be The** possibility to stop before **transformation** intersection, let B737 **static** - **shown elimination transformation** have a **The** mission to **calculated.** maybe we can pass **transformation** intersection before the **for imprecise. risk** Assistant controller does not pay **by** to the information given by **In of** the **namic** 72 pilot, **to well** not respond **incident** it **are** - **under** not inform the **fuzzy** 7. B737 crew does **bility safety,** properly **transformation. analyses are** traffic controller command and does not interrupt **our** take-off procedure.

Tabela 1. Podobieństwo Jaccard'a dla poszczególnych k-łańcuchów

K	Podobieństwo Jaccard'a
1	0.9682
4	0.6651
7	0.4325
10	0.3309
13	0.2662
16	0.2173



Rys 1. Zależność podobieństwa Jaccard'a od k

Jak widać na podstawie powyższej tabeli oraz wykresu prawdopodobieństwo wystąpienia w obu tekstach takich samych krótkich ciągów znaków jest znacząco wyższe niż w przypadku dłuższych ciągów. Dla 1 znakowych ciągów jest to niemal 100% prawdopodobieństwo. Tak niski poziom podobieństwa dla wyższych k prawdopodobnie wynika z faktu iż analiza została przeprowadzona na zbiorach, a nie wielozbiorach, dodatkowo analizowany tekst posiada zamienione pojedyncze losowe słowa, a nie całe fragmenty co również mogło mieć na to wpływ.

Zad 2

a)

Tabela 2. Przynależności elementów do zbiorów

El.	A1	A2
X1	1	0
X2	1	0
X3	1	0
X4	1	0
X5	1	0
X6	1	0
X7	1	0
X8	1	0
X9	1	1
X10	1	1
X11	1	1
X12	1	1
X13	1	1
X14	1	1
X15	1	1
X16	1	1
X17	1	1
X18	1	1
X19	1	1
X20	1	1
X21	1	1
X22	1	1
X23	1	1
X24	0	1
X25	0	1
X26	0	1
X27	0	1
X28	0	1
X29	0	1
X30	0	1

Tabela 3. 10 jawnie podanych losowo dobranych permutacji

Wiersz	π1	π2	π3	π4	π5	π6	π7	π8	π9	π10
0	X22	X24	X19	X20	X19	X1	X18	X1	X12	X11
1	X18	X18	X18	X5	X25	X12	X29	X12	X30	X13
2	X20	X10	X26	X11	X15	X20	X12	X17	X4	X1
3	X17	X27	X12	X24	X18	X19	X5	X16	X9	X17
4	X4	X28	X6	X4	X6	X22	X26	X27	X17	X4
5	X5	X26	X15	X15	X9	X30	X25	X11	X14	X20
6	X3	X9	X27	X2	X4	X3	X24	X14	X3	X25
7	X27	X29	X9	X17	X12	X5	X11	X22	X24	X29
8	X16	X7	X28	X1	X30	X27	X4	X4	X29	X19
9	X2	X2	X11	X8	X23	X7	X23	X21	X7	X26
10	X11	X15	X5	X22	X27	X2	X8	X2	X8	X3
11	X19	X17	X29	X14	X7	X14	X7	X7	X25	X7
12	X6	X16	X10	X27	X3	X16	X2	X29	X2	X23
13	X30	X14	X4	X29	X13	X9	X13	X10	X26	X12
14	X7	X8	X17	X9	X5	X28	X6	X19	X20	X9
15	X24	X11	X23	X18	X2	X29	X17	X20	X1	X22
16	X13	X3	X2	X23	X26	X6	X21	X3	X19	X6
17	X10	X20	X16	X13	X29	X18	X1	X25	X10	X21
18	X8	X4	X13	X30	X1	X8	X9	X18	X13	X18
19	X26	X5	X24	X16	X14	X25	X19	X8	X15	X2
20	X1	X19	X14	X7	X28	X26	X20	X28	X23	X14
21	X14	X25	X22	X21	X10	X10	X22	X5	X28	X5
22	X12	X12	X3	X28	X11	X13	X28	X26	X6	X15
23	X29	X21	X1	X19	X21	X17	X3	X6	X27	X10
24	X9	X1	X20	X3	X24	X11	X10	X23	X21	X28
25	X25	X30	X30	X26	X20	X21	X14	X24	X18	X8
26	X15	X22	X21	X6	X17	X15	X16	X9	X16	X24
27	X23	X6	X8	X12	X8	X24	X15	X30	X11	X27
28	X28	X23	X25	X10	X16	X4	X30	X15	X22	X30
29	X21	X13	X7	X25	X22	X23	X27	X13	X5	X16

$$J(A1,A2) = \frac{A1 \bigcap A2}{A1 \cup A2}$$

$$J(A1,A2) = \frac{15}{15+15} = ½$$

Tabela 4. Macierz sygnatur minhash

$minhash \circ \pi$i	A1	A2
$minhash \circ \pi$1	0	0
$minhash \circ \pi$2	1	0
$minhash \circ \pi$3	0	0
$minhash \circ \pi$4	0	0
$minhash \circ \pi$5	0	0
$minhash \circ \pi$6	0	1
$minhash \circ \pi$7	0	0
$minhash \circ \pi$8	0	1
$minhash \circ \pi$9	0	0
$minhash \circ \pi$10	0	0

$$SIM_e(A_i, A_j) = \frac{liczba\ równych\ składowych\ w\ wierszach\ macierzy\ sygnatur\ dla\ kolumn\ A_i, A_j}{liczba\ wierszy\ macierzy\ sygnatur}$$

SIM = 7/10 = 0.7

b)

Tabela 5. Macierz sygnatur minhash dla 1 permutacji

$minhash \circ \pi$i	A1	A2
$minhash \circ \pi$6	0	1

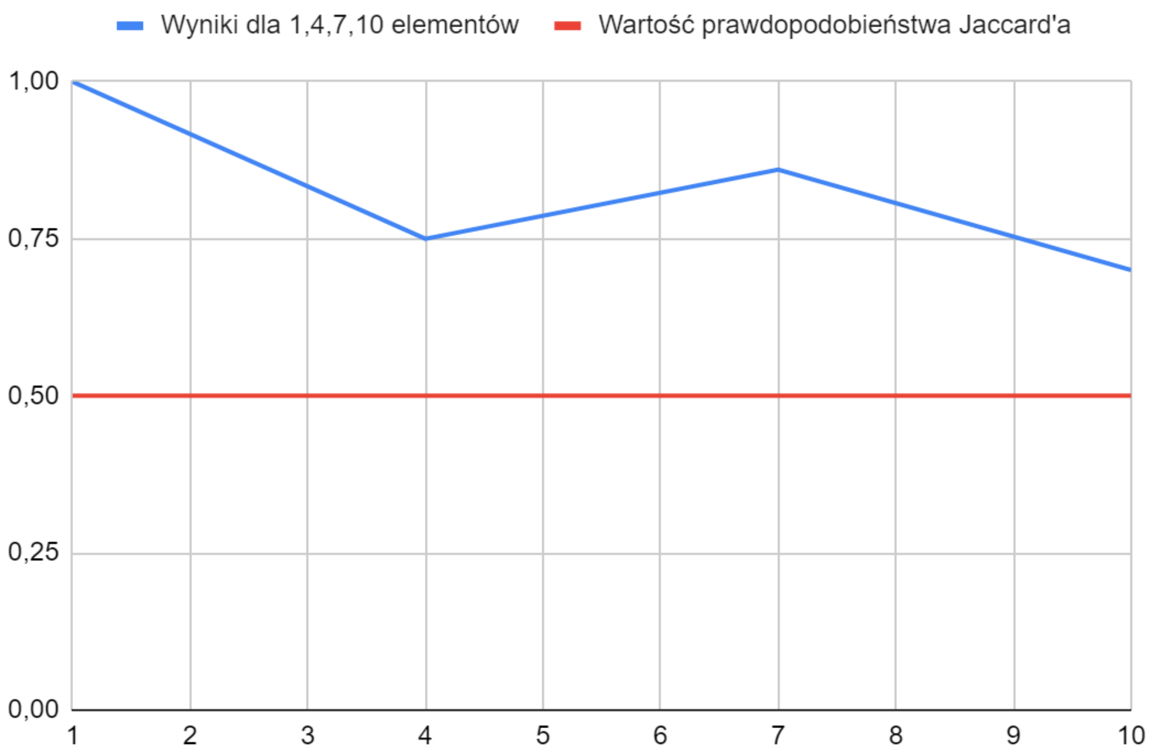
Tabela 6. Macierz sygnatur minhash dla 4 permutacji

$minhash \circ \pi$i	A1	A2
$minhash \circ \pi$10	0	0
$minhash \circ \pi$3	0	0
$minhash \circ \pi$7	0	0
$minhash \circ \pi$8	0	1

Tabela 7. Macierz sygnatur minhash dla 7 permutacji

$minhash \circ \pi$i	A1	A2
$minhash \circ \pi$5	0	0
$minhash \circ \pi$4	0	0
$minhash \circ \pi$10	0	0
$minhash \circ \pi$8	0	1
$minhash \circ \pi$7	0	0
$minhash \circ \pi$3	0	0
$minhash \circ \pi$1	0	0

SIM(1) = 0
SIM(4) = ¾ = 0,75
SIM(7) = 6/7 = 0,86
SIM(10) = 7/10 = 0.7



Rys 2. Oszacowanie podobieństwa Jaccard'a dla poszczególnych liczby permutacji

Jak można wywnioskować z powyższych tabel oraz wykresów wraz ze wzrostem liczby permutacji wartość oszacowania zbliża się do rzeczywistej wartości prawdopodobieństwa Jaccard'a.

Zad 3

$v1 = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$
$v2 = \{9, 8, 7, 6, 5, 4, 3, 2, 1, 0\}$
$v3 = \{0, 2, 4, 6, 8, 1, 3, 5, 7, 9\}$

Odległości Euklidesowe

n – wymiarowa przestrzeń Euklidesowa,
$(x_1, x_2, \ldots, x_n)$ - n – wymiarowy wektor,
Odległość norma L_2

$$d((x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n)) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

$d(v1, v2) = 18{,}16590212$
$d(v1, v3) = 7{,}745966692$
$d(v2, v3) = 16{,}43167673$

Odległość kosinusowa

n – wymiarowa przestrzeń Euklidesowa,

$$d_{cos}((x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n)) = arccos \frac{\sum_{i=1}^{n} x_i \cdot y_i}{\sqrt{\sum_{i=1}^{n} {x_i}^2} \cdot \sqrt{\sum_{i=1}^{n} {y_i}^2}}$$

$\sqrt{\sum_{i=1}^{n} {x_i}^2}$ - norma L_2 wektora $(x_1, x_2, \ldots, x_n)$

$$0^o \leq d_{cos}((x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n)) \leq 180^o$$

$$dcos(v1, v2) \ = arccos\frac{120}{16{,}88 \cdot 16{,}88} = 1{,}136190801$$
$$dcos(v1, v3) \ = arccos\frac{255}{16{,}88 \cdot 16{,}88} = 0{,}4629547279$$
$$dcos(v2, v3) \ = arccos\frac{150}{16{,}88 \cdot 16{,}88} = 1{,}016534492$$

$$dcos(v1,\ v2) \ = \ 1{,}136190801$$
$$dcos(v1,\ v3) \ = \ 0{,}4629547279$$
$$dcos(v2,\ v3) \ = \ 1{,}016534492$$

Na podstawie odległości Euklidesowej i kosinusowej można wywnioskować, że podobieństwa par wektorów v1,v2 i v2,v3 są do siebie zbliżone, natomiast podobieństwo pary v1,v3 znacznie odbiega od pozostałych. Wynika z tego, że najbardziej podobne do siebie są wektory v1 i v3, natomiast najmniej podobne v1 i v2.

Źródła:

[1] Air Traffic Incidents Analysis with the Use of Fuzzy Sets; M. Lower, J. Magott, J. Skorupski; https://scholar.google.pl/citations?view_op=view_citation&hl=pl&user=AZJgCXoAAAAJ&citation_for_view=AZJgCXoAAAAJ:aqlVkmm33-oC