

Łukasz Wachstiel

PORÓWNANIE METOD ROZMYTEGO I PROBABILISTYCZNEGO MODELOWANIA ZJAWISKA NA PRZYKŁADZIE OCENY RYZYKA USŁUG INFORMATYCZNYCH

Wprowadzenie

Artykuł konfrontuje ze sobą dwa podejścia do oceny ryzyka w zarządzaniu usługami informatycznymi. Pierwsze z nich polega na zastosowaniu funkcji prawdopodobieństwa jako miernika gwarancji tworzonej przez daną usługę wartości. W drugim zastosowano metodę rozmytych przedziałów gwarancji, oceniając wielkość ryzyka za pomocą wartości funkcji przynależności. Celem tej konfrontacji jest zweryfikowanie następującej hipotezy: rozmyte modelowanie ryzyka jest efektywniejszą metodą od modelowania probabilistycznego przy praktycznej ocenie ryzyka usług informatycznych.

Rozważania teoretyczne zostały tu podparte analizą empiryczną z wykorzystaniem modelu oceny ryzyka opisanego przez Autora w [8], a którego krótka charakterystyka została zamieszczona w rozdziale 2.

1. Definicja ryzyka

Analiza ryzyka jest jednym z najważniejszych procesów, który należy przeprowadzać na każdym etapie cyklu życia usługi informatycznej. Począwszy od definiowania strategii, aż po proces ciągłego doskonalenia, decydenci dokonują nieustannej analizy ryzyka w mniej lub bardziej świadomy sposób. Patrząc jednak z perspektywy klienta biznesowego lub bezpośrednio od strony użytkownika krańcowego, usługa musi generować (tworzyć) wartość.

W celu wyjaśnienia pojęcia wartości usługi posłużono się znaną m.in. z metodologii ITIL® definicją, która za warunek konieczny i wystarczający wytworzenia wartości uznaje koniunkcję dwóch czynników:

- użyteczności, rozumianej jako dodatkowa korzyść dostarczana klientowi poprzez wprowadzenie usługi w celu realizacji jego potrzeb biznesowych,
- gwarancji, czyli zapewnienia funkcjonowania usługi na założonym poziomie [3].

Korzystając z powyższej definicji, można wprowadzić termin ryzyka, które będzie określane jako prawdopodobieństwo wytworzenia wartości usługi informatycznej.

Powyższe sformułowania nie dają jeszcze pełnego obrazu poszukiwanej, rzeczywistej wartości ryzyka. Można jednak dalej wnioskować, że ryzyko to inaczej zapewnienie odpowiedniej gwarancji usługi, zakładając, że każda usługa niesie ze sobą pewną określoną użyteczność. Problematyczne pozostaje pojawiające się już kilkakrotnie pojęcie gwarancji, które nie zostało dokładnie sprecyzowane.

Sięgając ponownie do biblioteki dobrych praktyk zarządzania usługami informatycznymi [3], gwarancję usługi można utożsamiać z czterema czynnikami, które wspólnie zabezpieczają jej działanie na zdefiniowanym poziomie. Są to: dostępność (availability), pojemność (capability), ciągłość (continuity), bezpieczeństwo (security).

W artykule [8] Autor podjął próbę klasyfikacji czynników ryzyka usług informatycznych ze względu na wymienione składowe gwarancji. Ze względu na duży poziom ogólności prowadzonych rozważań (nie ograniczono się do żadnej, konkretnie opisanej i sparametryzowanej usługi) wprowadzono szersze pojęcie grup ryzyka, z których każda zawierała najważniejsze czynniki mogące wpływać ujemnie na: dostępność, pojemność, ciągłość lub bezpieczeństwo usługi. Ponadto pokazano, iż każda z czterech grup posiada swojego reprezentanta, czyli element (czynnik ryzyka) mający największy wpływ na zapewnienie odpowiedniej gwarancji usługi. Wynikiem przeprowadzonych badań było stworzenie modelu oceny ryzyka usług informatycznych, którego uogólnioną postać przedstawiono w następnym punkcie.

2. Uogólniony model oceny ryzyka

Przyjęto, że okres działania usługi T można podzielić na skończoną liczbę p -parami rozłącznych podokresów w postaci $[t_i, t_{i+1}] \subset T$, gdzie $t_i < t_{i+1}$ oraz

$\bigcup_{i=0}^p P_i = T$. Założono dodatkowo, że usługa posiada n użytkowników $U = \{u_1, u_2, \dots, u_n\}$ oraz l zasobów $Z = \{z_1, z_2, \dots, z_l\}$. Pomocne będzie również zdefiniowanie grup użytkowników $G = \{g_1, g_2, \dots, g_m\}$, gdzie $m \leq n^*$.

* Należy zwrócić uwagę, że słaba nierówność zakłada możliwość istnienia grup jednoelementowych.

Analiza grup czynników ryzyka doprowadziła nie tyle do wyłonienia ich reprezentantów, ile przede wszystkim do znalezienia krytycznego elementu gwarancji usługi – zasobów. Nazwano nim parametr usługi, którego zabezpieczenie ma największy wpływ na prawdopodobieństwo wytworzenia wartości końcowej. Dzięki temu można sprowadzić wielowymiarową analizę ryzyka do obserwacji stopnia wykorzystania zasobów usługi w różnych podokresach jej działania*.

Konkretyzując rozważania, można powiedzieć, że stopień wykorzystania zasobu $z \in Z$ w podokresie $P_i = [t_i, t_{i+1}] \subset T$ określa następujący zbiór rozmyty:

$$\{(t, z(t)) : t \in P_i, z(t) \in [0,1]\}, \quad (1)$$

$$z(t) = \min\left(\sum_{i=1}^n u_i(t), 1\right), \quad t \in P_i, \quad (2)$$

gdzie:

$$u_i : P_i \rightarrow [0,1], i = 1, \dots, n \quad (3)$$

to procent wykorzystania zasobu z przez użytkownika u_i w podokresie P_i .

Następnie określa się stopień wykorzystania danego zasobu w całym okresie dostępności usługi jako średnią arytmetyczną z maksymalnych wydajności zasobów w poszczególnych podokresach z uwzględnieniem ich długości (ozn. $|P_i| = t_{i+1} - t_i$):

$$stw(z) = \frac{1}{|T|} \cdot \sum_{\substack{i=1, \dots, p, \\ t \in P_i}} \max(z(t) \cdot |P_i|), \quad (4)$$

gdzie $|T|$ – długość całego okresu.

Na podstawie powyższych ustaleń ryzyko tworzenia wartości usługi informatycznej będzie definiowane jako gwarancja zapewnienia wydajności jej zasobów mierzonej maksymalną wartością stopni wykorzystania wszystkich zasobów obliczanych za pomocą wzoru (4).

W kolejnej części niniejszego artykułu zastosowano dwa odmienne podejścia do modelowania opisanego ryzyka w celu zweryfikowania postawionej we wprowadzeniu hipotezy.

* Stopień wykorzystania zasobu będzie również nazywany wydajnością zasobu.

3. Modelowanie ryzyka – różne podejścia

W klasycznym podejściu do oceny ryzyka zakłada się, że ma się do dyspozycji pewną funkcję prawdopodobieństwa, dzięki której wprost można policzyć wartość wystąpienia sytuacji niepożądaney. Już w pierwszej połowie XX wieku właśnie tak interpretował ryzyko Knight [2], rozróżniając je od niepewności – czynnika niemierzalnego lub niepoliczalnego. Trudno jednak takie podejście stosować, opisując zjawiska otaczającej nas rzeczywistości. Najczęściej ludzie w mowie codziennej posługują się następującymi zwrotami: „małe ryzyko”, „duże ryzyko”, „niewielkie ryzyko”, które bardziej niosą ze sobą pewną informację jakościową – subiektywną niż ilościową – obiektywną. Warto jednak czasami się zastanowić, czy otrzymana z pewnym przybliżeniem, a mówiąc bardziej poprawnie matematycznie – z pewnym stopniem wiarygodności informacja nie jest tak samo wartościowa, jak ta uzyskana metodami probabilistycznymi.

Idealny do konfrontacji tych dwóch alternatywnych podejść wydaje się być problem ryzyka, który już z samej definicji jest czymś bardzo subiektywnym i trudnym w rzetelnej ocenie. Stosowane do jego analizy różne miary prawdopodobieństwa dają często pozorny efekt dokładności, gdyż dane wejściowe takiego modelu bywają nierzadko standaryzowane (normalizowane), aby spełniały konkretne założenia.

Założono, że podstawową miarą ryzyka w analizowanym modelu jest opisana w rozdziale 2 gwarancja zasobów usługi, która zostanie wyznaczona na dwa sposoby.

W pierwszym z nich przyjęto, że wydajności poszczególnych zasobów są wyznaczone za pomocą funkcji gęstości rozkładu Gaussa, a następnie sprowadzane do postaci liczb rozmytych*. Wszystkie operacje weryfikowano jednocześnie na danych empirycznych dla uniwersyteckiej usługi poczty elektronicznej, której parametry umieszczono w tabeli 1.

Tabela 1

Ryzyko gwarancji usługi poczty elektronicznej

Podokresy dostępności	z_1	z_2	z_3	Ryzyko gwarancji
1	2	3	4	5
0-1	0,002	0,027	0,123	0,123
1-2	0,002	0,038	0,156	0,156
2-3	0,002	0,05	0,127	0,127
3-4	0,002	0,01	0,123	0,123

* Problem wspólnej reprezentacji danych rozważono w pracy [4]. Autorzy dowodzili, iż bardziej właściwe jest podejście, w którym zamienia się funkcje gęstości na przedziały rozmyte (a nie odwrotnie), m.in. dlatego, że nie wprowadza się pozornego złudzenia dokładności prowadzonych obliczeń.

cd. tabeli 1

1	2	3	4	5
4-5	0,002	0,002	0,167	0,167
5-6	0,003	0,004	0,189	0,189
6-7	0,002	0,057	0,182	0,182
7-8	0,017	0,079	0,155	0,155
8-9	0,031	0,231	0,183	0,231
9-10	0,033	0,32	0,137	0,32
10-11	0,042	0,367	0,189	0,367
11-12	0,069	0,412	0,199	0,412
12-13	0,105	0,519	0,199	0,519
13-14	0,145	0,51	0,199	0,51
14-15	0,201	0,421	0,199	0,421
15-16	0,189	0,412	0,198	0,412
16-17	0,078	0,49	0,125	0,49
17-18	0,023	0,321	0,134	0,321
18-19	0,011	0,284	0,167	0,284
19-20	0,007	0,164	0,102	0,164
20-21	0,008	0,079	0,105	0,105
21-22	0,004	0,082	0,162	0,162
22-23	0,003	0,045	0,198	0,198
23-0	0,004	0,033	0,134	0,134
Średnia (μ)	0,0410417	0,206542	0,1605	0,206542
Odchylenie st. (σ)	0,0602643	0,185785	0,033052	

Okres dostępności usługi podzielono na 24 podokresy o równej długości (godziny). Zasoby ujęto w trzech głównych grupach*: z_1 – infrastruktura aplikacji, m.in. narzędzia obsługi serwera poczty, oprogramowanie antywirusowe, narzędzia diagnostyczne; z_2 – infrastruktura sieciowa, w której analizowano przede wszystkim ruch sieciowy downstream i upstream; z_3 – infrastruktura sprzętowa, czyli głównie wydajność serwerów i ich podzespołów.

Średnie arytmetyczne (μ) oraz odchylenia standardowe (σ) uzyskane dla poszczególnych wartości wykorzystania zasobów posłużyły do wyznaczenia przedziałów ufności funkcji gęstości f rozkładu normalnego $\sim N(\mu, \sigma)$. W kolejnym etapie zamieniono przedziały ufności na α -przekroje liczb rozmytych [1], stosując następujące przekształcenia:

* Miary, jakimi się posłużono przy badaniach wydajności zasobów, można znaleźć w [6] i [7].

$$\int_{\mu - \sigma(3 - k \frac{3}{\alpha})}^{\mu + \sigma(3 - k \frac{3}{\alpha})} f(x) dx = \beta \Leftrightarrow 2\Phi(3 - k \frac{3}{\alpha}) - 1 = \beta, \quad (5)$$

$$\Phi \sim N(0, 1) \quad (6)$$

gdzie Φ – dystrybuanta rozkładu normalnego.

$$z_k = \left[\mu - \sigma(3 - k \frac{3}{\alpha}), \mu + \sigma(3 - k \frac{3}{\alpha}) \right], \quad (7)$$

$$p\left(\mu - \sigma(3 - k \frac{3}{\alpha})\right) = p\left(\mu + \sigma(3 - k \frac{3}{\alpha})\right) = 1 - \beta, \quad (8)$$

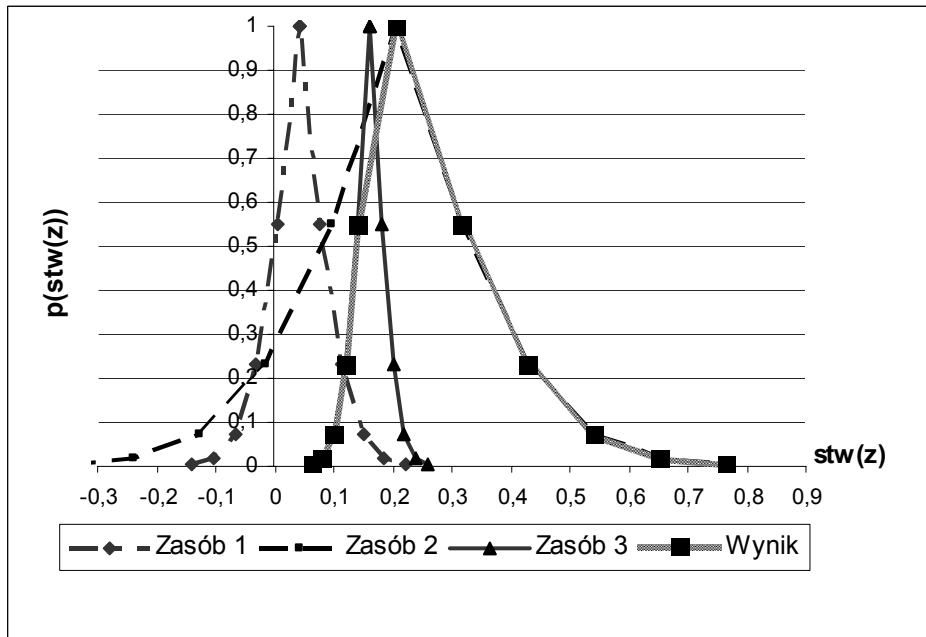
$$k \in N \cup \{0\}, \quad k \leq \alpha. \quad (9)$$

Wyniki transformacji funkcji gęstości stopnia wykorzystania poszczególnych zasobów przedstawiono na rysunku 1. Dodatkowo, korzystając z operacji maksimum na przedziałach rozmytych*, zamieszczono graficzną reprezentację funkcji przynależności dla gwarancji zasobów usługi w czasie T .

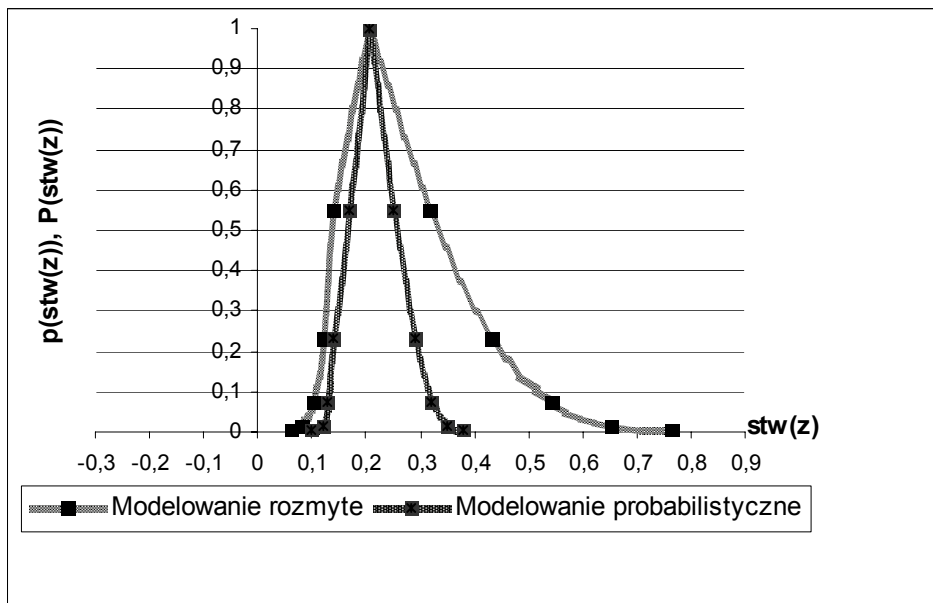
Dzięki przeprowadzonej zamianie funkcji gęstości na przedziały rozmyte przygotowano odpowiednio dane do porównania z drugą metodą modelowania ryzyka, w której stopnie wykorzystania zasobów są generowane pseudolosowo z użyciem funkcji gęstości rozkładu normalnego**, a następnie porównywane z rzeczywistymi wartościami. Otrzymane w ten sposób wydajności poszczególnych zasobów podstawiano do wzoru (4). Wyniki modelowania rozmytego oraz symulacyjnego przedstawiono na rysunku 2.

* Więcej o operacjach na przedziałach liczb rozmytych można przeczytać w pracach [1; 5]. W tym miejscu użyto jednej z prostszych operacji maksimum, która wybiera maksymalne wartości liczb poszczególnych przedziałów i tworzy jeden przedział.

** Zastosowano do tego celu generator liczb wchodzący w skład pakietu *Mathematica*® 4.0.



Rys. 1. Transformacja funkcji gęstości do przedziałów rozmytych



Rys. 2. Porównanie modelowania rozmytego i probabilistycznego

Podsumowanie

Rezultaty przeprowadzonych symulacji (rysunek 2) wskazują na pewne podobieństwa modelowania za pomocą dwóch opisywanych metod. Widać, że poziom wykorzystania zasobów usługi o największym stopniu możliwości wystąpienia jest równy poziomowi o największym prawdopodobieństwie. Różnice pojawiają się dla wydajności o dużym stopniu niepewności. W przypadku modelowania symulacyjnego są „odrzucone” skrajne wartości wydajności, tzn. prawdopodobieństwa ich wystąpienia są bliskie zeru. Nie odzwierciedla to jednak rzeczywistej sytuacji modelowania ryzyka, w której najczęściej poszukuje się słabych punktów gwarancji usługi, chcąc je wyeliminować. Modelowanie symulacyjne rzadko uśrednia pojawiające się wartości ryzyk, wprowadzając jedynie iluzoryczną poprawność otrzymywanych wyników, która nie uwzględnia czynników mogących mieć kluczowy wpływ na funkcjonowanie usługi.

Jednocześnie można zauważyć, że niskie wartości funkcji przynależności będą zawsze implikowały niskie prawdopodobieństwa, a co za tym idzie – zbiór leżący pod wykresem funkcji prawdopodobieństwa będzie się zawierał w zbiorze rozmytym opisywanym funkcją przynależności.

Trudność zastosowania podejścia symulacyjnego przejawia się również w opisywaniu parametrów modelu za pomocą funkcji gęstości zmiennych losowych, co nie zawsze jest łatwym zadaniem. Wymaga to przeprowadzenia dużej ilości obserwacji (wykonywania iteracji na próbach o dużej liczności), dzięki którym będzie można wybrać najdokładniejszy rozkład prawdopodobieństwa. Istnieje jeszcze wiele innych, „technicznych” problemów, jak chociażby wykorzystywane generatory liczb pseudolosowych, o których więcej można przeczytać w opracowaniu [9].

Podsumowując, modelowanie z użyciem metod probabilistycznych nie dostarcza więcej informacji o poszukiwanej wartości ryzyka niż modelowanie rozmyte. Dodatkowo w przypadku modelowania probabilistycznego występuje niepożądane przy ocenie ryzyka zjawisko polegające na uśrednianiu (standaryzacji) rezultatów o skrajnie małym lub dużym prawdopodobieństwie wystąpienia. Sytuacja taka nie występuje podczas użycia drugiej metody, co wynika bezpośrednio z własności zbiorów rozmytych, które dostarczają dokładnej informacji o stopniu przynależności poszczególnych elementów. Weryfikując postawioną na początku artykułu hipotezę, stwierdzono ostatecznie, iż stosunek otrzymanych przy ocenie ryzyka obiektywnych wyników w odniesieniu do poziomu złożoności wykorzystanej aparatury badawczej jest większy w przypadku modelowania rozmytego, co w konsekwencji potwierdza wstępne przypuszczenie o przewadze tej metody w założonej kategorii efektywności.

Literatura

1. Drewniak J.: *Podstawy teorii zbiorów rozmytych: skrypt przeznaczony dla studentów IV i V roku matematyki*. Uniwersytet Śląski, Katowice 1984.
2. Knight F.: *Risk, Uncertainty and Profit*. London 1933.
3. Office of Government Commerce (OGC), *ITIL® Service Service Strategy*. Wydawnictwo TSO (The Stationary Office), Wielka Brytania 2007.
4. Róg P., Sewastianow P.: *Metoda rozmyto-przedziałowa a metoda Monte-Carlo w symulacji procesów produkcyjnych: porównanie*. XIV Górską Szkoła PTI, Szczyrk 2002.
5. Tyrła R.: *Comparison of Fuzzy Numbers Ranking Methods*. Prace Naukowe AJD, Seria Matematyka Scientific Issues of Jan Długosz University in Częstochowa, Mathematics, Czestochowa 2008.
6. Simmonds A.: *Wprowadzenie do transmisji danych*. WKŁ, Warszawa 2009.
7. Światowiak J.: *Microsoft Windows Server 2003/2008. Bezpieczeństwo środowiska z wykorzystaniem Forefront Security*. Helion, Gliwice 2010.
8. Wachstiel Ł.: *Identyfikacja czynników ryzyka w zarządzaniu usługami informatycznymi*. Materiały Krakowskiej Konferencji Młodych Uczonych, Grupa Naukowa Pro Futuro, Kraków 2011, s. 1091-1100.
9. Wieczorkowski R., Zieliński R.: *Komputerowe generatory liczb losowych*. WNT, Warszawa 1997.

COMPARISON OF FUZZY AND PROBABILISTIC MODELING WITH USING IT SERVICES RISK EVALUATION EXAMPLE

Summary

In the article, two approaches to IT services risk assessment has been described and compared. The first one is based on fuzzy methods in the opposite to the second one which uses probability. The following hypothesis has been proposed and proved by showing a practical example: fuzzy modeling is more effective than probabilistic in IT services risk assessment.