AUTOMATYKA • 2006 • Tom 10 • Zeszyt 3

Krzysztof Przybyszewski*

Zastosowanie zbiorów rozmytych do oceny testów (problemów zamkniętych)

1. Wprowadzenie

Moduły ewaluacyjne odgrywają zasadniczą rolę w systemach zdalnego nauczania (SZN) [6]. Umożliwiają symulację obecności nauczyciela w trybie asynchronicznym oraz indywidualizację ścieżki nauczania studenta [4]. Pożądane byłoby, aby procesy decyzyjne zachodzące w tych modułach były jak najbardziej podobne do procesów podejmowania decyzji przez nauczyciela w trakcie oceniania studenta.

Nawet w przypadku nauczania tradycyjnego proces ewaluacji oparty jest na działaniu systemu ekspertowego: nauczyciel posługuje się własną wiedzą i ustalonymi regułami, oceniając postępy każdego ze studentów oraz proponując dalsze partie materiału czy też sposoby wzbogacania jego umiejętności [3, 2]. Tak jak w większości systemów ekspertowych, ostateczna informacja wygenerowana w procesie decyzyjnym jako ocena jest nieprecyzyjna (w sensie logiki matematycznej) i ma postać określenia słownego lub liczby, która jest reprezentantem przedziału liczb. Naturalne wydaje się więc zastosowanie w procesie oceniania zbiorów i liczb rozmytych.

W SZN najczęściej stosowane są przy ewaluacji tradycyjne metody: dokonuje jej nauczyciel lub ich grupa. Nie ma tutaj znaczenia, na jakim poziomie SZN jest dokonywana ta ocena: czy jest to ewaluacja na poziomie kursu przedmiotowego, semestru, czy też dotyczy całego okresu nauki. Próby automatyzacji procesu dotyczą jedynie ewaluacji opartej na ocenie problemów zamkniętych, a i w tym przypadku decyzję o ocenie dotychczasowych osiągnięć i dalszej ścieżce kształcenia podejmuje się na podstawie tablic decyzyjnych z ostrymi warunkami, co wydaje się niezbyt zgodne z obrazem procesu zachodzącego podczas ewaluacji dokonywanej przez nauczyciela. Przykładami takich systemów ewaluacyjnych są egzaminy specjalizacyjne oferowane przez firmy CISCO lub Microsoft.

2. Liczby rozmyte jako reprezentacja skali ocen

Proponujemy następującą metodę oceniania opartą na zastosowaniu liczb i zbiorów rozmytych.

* Instytut Nauczania na Odległość, Wyższa Szkoła Humanistyczno-Ekonomiczna w Łodzi

Niech $A_a \subseteq \mathbf{R}$ będzie liczbą rozmytą [7] określoną przez trzy parametry: m_L, a, m_P, dla której funkcja przynależności ma następującą postać:

$$\mu(x) = \begin{cases} 0 & \text{dla } x < 0 \\ \dfrac{x}{a - m_L} & \text{dla } 0 \le x < a - m_L \quad \text{oraz} \quad (a - m_L) \neq 0 \\ 1 & \text{dla } a - m_L \le x \le a + m_P \\ \dfrac{5 - x}{5 - a - m_P} & \text{dla } a + m_P < x \le 5 \quad \text{oraz} \quad (a - m_P) \neq 5 \\ 0 & \text{dla } x > 5 \end{cases} \tag{1}$$

Liczbę A_a możemy nazwać *trapezoidalną liczbą rozmytą*. Parametr a nazywamy *centrum* liczby, natomiast parametry m_L i m_P nazywamy odpowiednio: *lewostronną* i *prawostronną szerokością* liczby. Liczbę A_a możemy zapisać w następujący sposób: $A_a = (m_L; a; m_P)$.

Dla trapezoidalnej liczby rozmytej definiuje się także *punkt środkowy* tej liczby $\tilde{a}$, jako środek przedziału, dla którego funkcja przynależności przyjmuje wartość 1. O trójkątnej liczbie rozmytej postaci:

$$\mu(x) = \begin{cases} 0 & \text{dla } x < 0 \\ \dfrac{x}{\tilde{a}} & \text{dla } 0 \le x < \tilde{a} \\ 1 & \text{dla } x = \tilde{a} \\ \dfrac{5 - x}{5 - \tilde{a}} & \text{dla } \tilde{a} < x \le 5 \\ 0 & \text{dla } x > 5 \end{cases} \tag{2}$$

mówimy, że jest generowana przez liczbę A_a i zapisujemy ją symbolem $\widetilde{A_a}$.

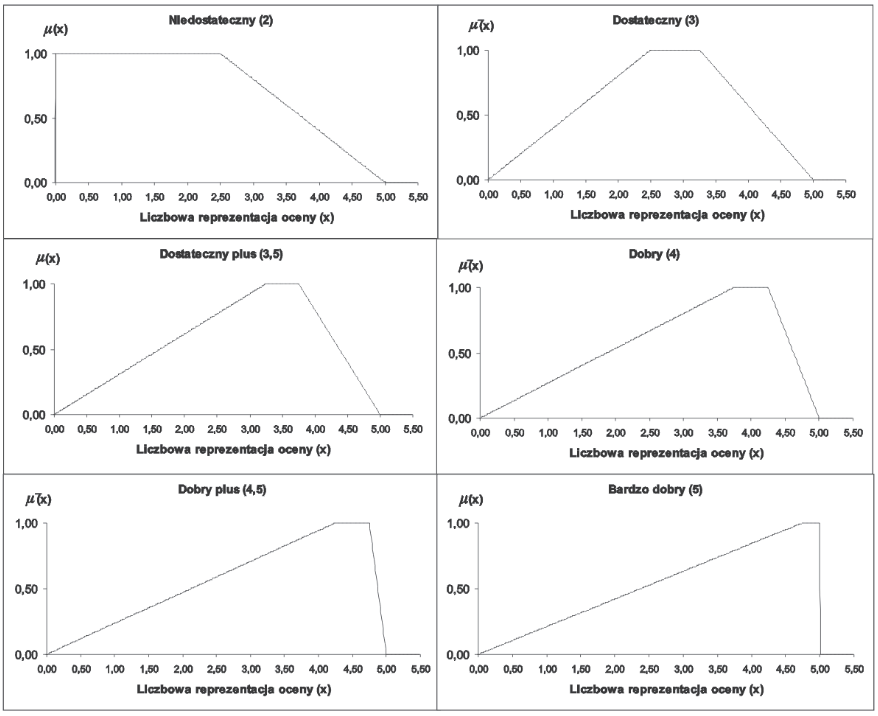
Zbiór liczb rozmytych:

$SM = \{(2;2;0,5),(0,5;3;0,25),(0,25;3,5;0,25),(0,25;4;0,25),(0,25;4,5;0,25),(0,25;5;0)\}$

jest obrazem skali ocen stosowanej w polskich szkołach wyższych. W takim przypadku skala ocen jest zbiorem zmiennych lingwistycznych SM_{ling}, następującej postaci:

$\mathrm{SM}_{\mathrm{ling}}$={niedostateczny, dostateczny, dostateczny plus, dobry, dobry plus, bardzo dobry}.

Interpretację graficzną liczb rozmytych ze skali ocen stosowanej w polskich szkołach wyższych przedstawiono na rysunku 1.



Rys. 1. Graficzna reprezentacja skali ocen stosowanej w polskich uczelniach

Jedynym warunkiem, jaki muszą spełniać liczby rozmyte reprezentujące skalę ocen, jest warunek (nazwany *warunkiem pełnego wypełnienia skali*), aby były one składnikami dekompozycji trapezoidalnej liczby rozmytej (2,5;2,5;2,5). Dla skali *SM* można to zapisać, posługując się zasadami operacji dokonywanych na zbiorach rozmytych [7, 5] w następującej postaci

$$(2,5;2,5;2,5) = (2;2;0,5) \cup (0,5;3;0,25) \cup (0,25;3,5;0,25) \cup$$
$$\cup (0,25;4;0,25) \cup (0,25;4,5;0,25) \cup (0,25;5;0).$$

Ogólny zapis warunku dla dowolnej skali ocen wyrażonej przez liczby rozmyte będzie miał postać

$$(2,5;2,5;2,5) = \bigcup_i (m_{Li}; a_i; m_{Pi}) \tag{3}$$

gdzie indeks i przyjmuje wszystkie dostępne wartości w danej skali ocen.

Niejednokrotnie stosowane są skale ocen inne niż przedstawiona.

Przykładem może być powszechnie stosowana w Europie Zachodniej skala oparta na sześciu ocenach (SM_{EU})

$$SM_{EUling} = \{F, E, D, C, B, A\}.$$

Obrazem tej skali ocen może być następujący zbiór trapezoidalnych liczb rozmytych:

$$SM_{EU} = \{(0;0;0,5),(0,5;1;0,5),(0,5;2;0,5),(0,5;3;0,5),(0,5;4;0,5),(0,5;5;0)\},$$

$$SM_{EU} = \{A_0, A_1, A_2, A_3, A_4, A_5\}.$$

Łatwo sprawdzić, że skala ta spełnia warunek pełnego wypełnienia skali (3). Skala SO_{EU} charakteryzuje się przede wszystkim równymi szerokościami połówkowymi liczb rozmytych, które są elementami skali (z wyłączeniem liczb skrajnych: A_0 i A_5) oraz równoważnością centrum i punktu środkowego każdej liczby (także z wyłączeniem liczb skrajnych)

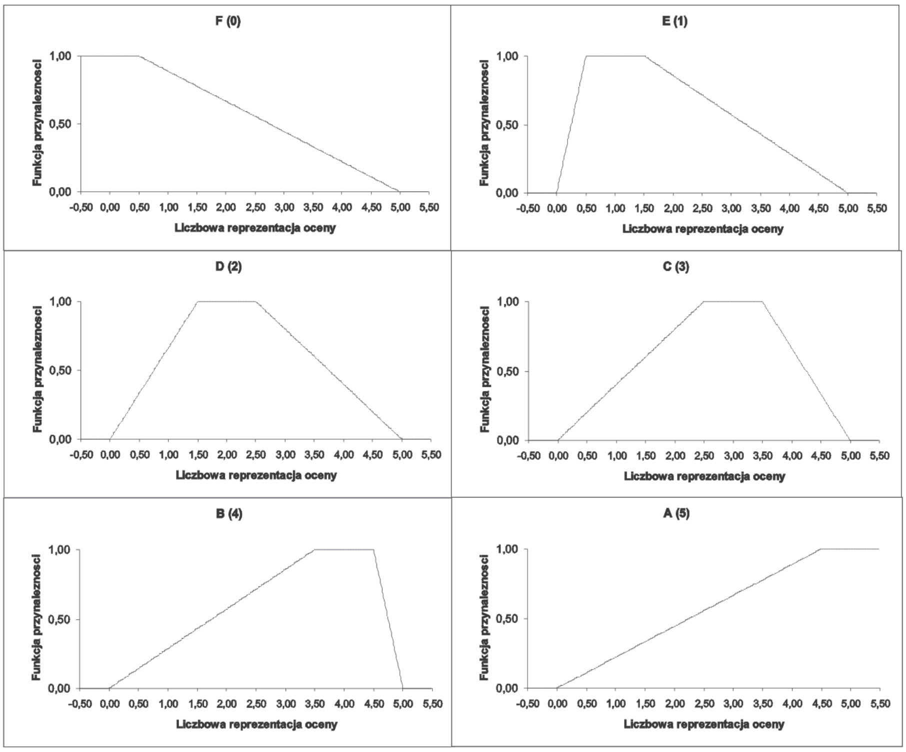
$$\underset{0<i<5}{\forall} a_i = \widetilde{a_i} \wedge m_{Li} = m_{Pi}.$$

Skalę tę można łatwo rozszerzyć w ten sposób, aby liczby skrajne także spełniały warunki symetrii: wystarczy w przypadku liczby rozmytej A_0 przyporządkować jej lewostronnej szerokości m_{L0} wartość różną od zera i równą wartości lewostronnej szerokości każdej z pozostałych liczb (w tym przypadku $A_0 = (0,5;\ 0;\ 0,5)$) oraz dokonać takiej samej operacji dla prawostronnej szerokości liczby A_5 ($A_5 = (0,5;\ 5;\ 0,5)$). Ten paradoksalny zabieg (wyprowadzenie liczb rozmytych poza zakres objęty skalą reprezentowaną przez zmienne lingwistyczne) pozwala wprowadzić symetrię i znacznie ułatwia operacje na liczbach rozmytych dokonywane na przykład przy ewaluacji problemów zamkniętych.

Taka sytuacja (zarówno dla skali rozszerzonej, jak i zwykłej) wydaje się charakterystyczna dla prawidłowo skonstruowanej skali ocen, a skalę o takich właściwościach można nazwać *skalą zrównoważoną*. Interpretację graficzną tej skali ocen (rozszerzonej) przedstawiono na rysunku 2.

Możliwe jest także przekształcenie skali ocen stosowanej w polskich uczelniach wyższych *SM* w rozszerzoną skalę zrównoważoną SM_{roz}. Wymaga to wprowadzenia ocen: *niedostateczny zerowy, niedostateczny zerowy plus, niedostateczny jeden, niedostateczny jeden plus, niedostateczny plus* oraz zmiany szerokości lewostronnej i prawostronnej oceny *niedostatecznej*, a także szerokości prawostronnej oceny *bardzo dobrej*. Zbiór liczb rozmytych reprezentujących tę skalę ocen będzie następujący

$$\begin{aligned} SM_{roz} = \{&(0,25;0;0,25),(0,25;0,5;0,25),(0,25;1;0,25),(0,25;1,5;0,25),(0,25;2;0,25),\\ &(0,25;2,5;0,25),(0,25;3;0,25),(0,25;3,5;0,25),(0,25;4;0,25),(0,25;4,5;0,25),\\ &(0,25;5;0,25)\}.\end{aligned}$$



Rys. 2. Interpretacja graficzna liczb rozmytych
budujących rozszerzoną skalę SM_{EU}

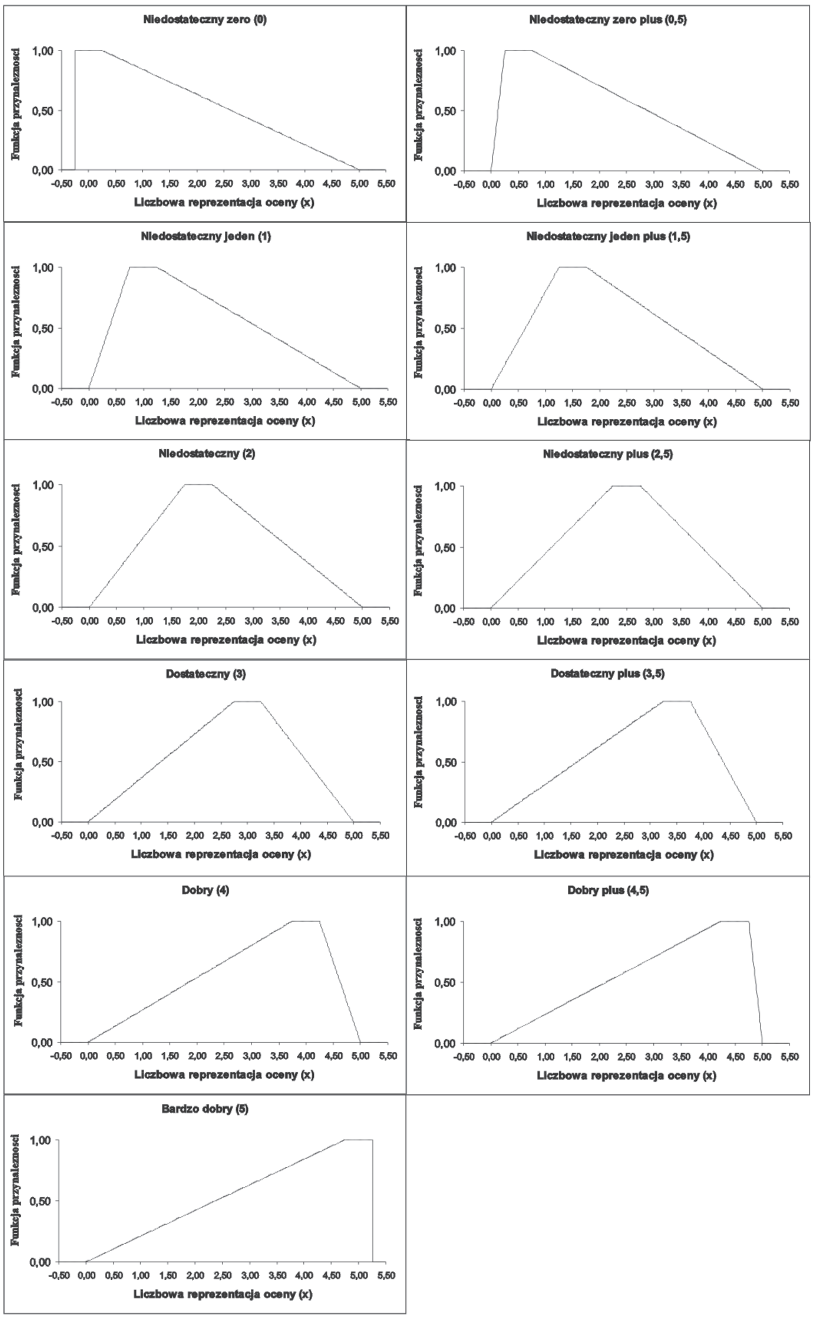
Interpretację graficzną skali SM_{roz} przedstawiono na rysunku 3.

Innym przykładem skali ocen może być skala procentowa przyjmowana przy ocenianiu poprzez określenie reguł za pomocą względnych przedziałów procentowych ilości zdobytych punktów. Przykładem jest skala ocen przyjęta w Polskim Uniwersytecie Wirtualnym (PUW).

Zestawienie przedziałów procentowych, odpowiadających im ocen i liczb rozmytych przedstawiono w tabeli 2.

W skali ocen przyjętej w PUW występuje ciekawa anomalia dla oceny *dostateczny plus*: centrum liczby nie należy do przedziału liczbowego, dla którego funkcja przynależności przyjmuje wartość 1: $3{,}5 \notin [3, 75; 4, 0)$.

Wskazuje to na niefortunny dobór przedziałów procentowych, ale jednocześnie świadczy o elastyczności zaproponowanej metody definiowania skali ocen wykorzystującej liczby rozmyte. Dla wszystkich liczb rozmytych reprezentujących tę skalę, centrum liczby nie jest równe jej punktowi środkowemu Skala przyjęta w PUW nie jest skalą zrównoważoną.



Rys. 3. Interpretacja graficzna liczb rozmytych reprezentujących rozszerzoną skalę stosowaną w polskich uczelniach wyższych

Tabela 2
Skala ocen przyjęta w PUW i odpowiadający jej zbiór liczb rozmytych

Przedział procentowy [%]	Ocena	Liczba rozmyta
[0, 60)	Niedostateczny	(2;2;1)
[60, 75)	Dostateczny	(0,3,0,75)
[75, 80)	Dostateczny plus	(-0,25;3,5;1)
[80, 90)	Dobry	(0;4;0,5)
[90, 95)	Dobry plus	(0;4,5;0,25)
[95, 100]	Bardzo dobry	(0,25;5;0)

3. Ocena wynikowa i średnia z ocen

W bardzo wielu przypadkach mamy do czynienia z koniecznością wystawienia oceny wynikowej (końcowej) na podstawie ocen cząstkowych. Taka sytuacja występuje w przypadku oceny wystawianej z egzaminu dyplomowego lub oceny semestralnej z danego przedmiotu, a nawet w przypadku oceniania sprawdzianów zawierających więcej niż jeden problem.

W takim przypadku proponujemy przyjęcie dwuetapowego algorytmu oceniania [5].

Pierwszy etap polega na wyznaczeniu liczby rozmytej reprezentującej ocenę średnią. Dokonuje się tego poprzez obliczenie średniej arytmetycznej [1] wszystkich trapezoidalnych liczb rozmytych reprezentujących oceny cząstkowe, zgodnie z zasadą rozszerzenia operacji matematycznych ze zbiorów nierozmytych [7]. Jeśli $PM = \left\{A_P^1, A_P^2, \ldots, A_P^N\right\}$ jest zbiorem wszystkich reprezentantów ocen cząstkowych, to liczba rozmyta $\overline{A_P}$ określona zależnością

$$\overline{A_P} = \frac{1}{N} A_P^1 \oplus \frac{1}{N} A_P^2 \oplus \ldots \oplus \frac{1}{N} A_P^N \tag{4}$$

jest reprezentantem oceny średniej. Współczynniki $1/N$ pełnią rolę wag przypisywanych poszczególnym ocenom cząstkowym. Możliwe jest przypisanie innych wartości wag poszczególnym ocenom cząstkowym, z zastrzeżeniem, że ich suma musi być równa 1.

$$\overline{A_P} = w_1 \cdot A_P^1 \oplus w_2 \cdot A_P^2 \oplus \ldots \oplus w_N \cdot A_P^N,$$

przy czym

$$\sum_{i=1}^{N} w_i = 1.$$

W drugim etapie proponujemy przyjęcie jednej z dwóch strategii wyznaczania oceny końcowej:

1. Jeśli ocena średnia jest brana pod uwagę jako ocena cząstkowa, w dalszym procesie oceniania, to pozostawiamy ją bez zmian.
2. Jeśli na jej podstawie wyznaczamy ocenę koncową reprezentowaną przez trapezoidalną liczbę rozmytą A_{FM}, to liczbę tę można otrzymać wybierając ze zbioru liczb rozmytych reprezentujących skalę ocen SM, tę ocenę, dla której przecięcie z trójkątną liczbą rozmytą generowaną przez punkt środkowy oceny średniej (A_P) jest rozmytym zbiorem znormalizowanym

$$A_{FM} = A_{SM}^{i} : \left(A_{SM}^{i} \in SM\right) \wedge \left(h\left(\widetilde{A_P} \cap A_{SM}^{i}\right) = 1\right) \tag{5}$$

Taki sposób wyznaczania oceny końcowej wymaga zapamiętania zarówno oceny średniej (A_P), jak i oceny końcowej (A_{FM}). Jest to bardzo korzystne w przypadku zastosowania prezentowanych algorytmów w systemach ekspertowych modułów ewaluacji w SZN. Wydaje się też wymagane w przypadku porównywania ocen wystawionych według różnych skal, na przykład ocen tego samego przedmiotu wystawionych w dwóch różnych uczelniach. Taki algorytm wyznaczania ocen końcowych wydaje się zgodny także z tradycyjnym sposobem oceniania przez nauczyciela.

4. Wykorzystanie liczb rozmytych przy ocenie testów (problemów zamkniętych)

Zaproponowaną metodę wykorzystano do oceny zestawu pytań testowych (problemów zamkniętych) przeprowadzanego na zakończenie jednosemestralnego kursu podstaw fizyki na jednym z kierunków technicznych studiów I stopnia w WSHE. Zestaw składał się z 10 problemów zamkniętych różnego rodzaju. Były to pytania testowe podstawiania, przyporządkowania, uzupełnień oraz wielokrotnego wyboru. Problemy zostały tak skomponowane, że za odpowiedź na każde pytanie możliwe było uzyskanie od 0 (odpowiedź całkowicie błędna) do 3 punktów (odpowiedź całkowicie poprawna).

Klasyczna metoda oceny testu polegała na zsumowaniu ilości uzyskanych punktów i zastosowaniu zebranych w tabeli reguł wyboru do wystawienia oceny końcowej (tab. 3).

Tabela 3

Tabela zawierająca reguły wyznaczania oceny końcowej dla zestawu pytań testowych z podstaw fizyki

Liczba punktów	Ocena
[0, 10]	Niedostateczny
(11, 20]	Dostateczny
(20, 22]	Dostateczny plus
(22, 26]	Dobry
(26, 27]	Dobry plus
(27, 30]	Bardzo dobry

Ocena zestawu pytań zgodna z metodą wykorzystującą operacje na liczbach rozmytych przebiegała w trzech etapach:

1. Wyznaczenie oceny odpowiedzi na każde pytanie. Dla potrzeb tego etapu konieczne było skonstruowanie skali zrównoważonej ocen pojedynczych pytań. Przyjęto skalę SM_1 następującej postaci

$$SM_1 = \{(0,5;0;0,835), (0,835;1,67;0,835), (0,835;3,34;0,835), (0,825;5;0,5)\} = \{A_0, A_1, A_2, A_3\}$$

   Skala SM_1 stanowi reprezentację wszystkich możliwych ocen pojedynczego zadania.
2. Wyznaczenie oceny średniej $\overline{A_{p1}}$ dla całego zestawu (zgodnie z równaniem (4)).
3. Wyznaczenie oceny końcowej A_{FM1} dla całego zestawu zgodnie z równaniem (5), gdzie jako skalę odniesienia przyjęto rozszerzoną, zrównoważoną skalę ocen przyjmowana w polskich uczelniach wyższych SM_{roz}, przy czym ocenę *niedostateczny plus* (0,25;2,5;0,25) przyjęto jako ocenę zaliczającą.

Okazało się, że przyjęta skala ocen dla pojedynczych pytań testowych wyraźnie zawęża przedział oceny bardzo dobrej (możliwa tylko do wystawienia w przypadku odpowiadającemu zdobyciu sumarycznie 30 punktów) i podwyższa dolną granicę dla oceny dostatecznej do 13 punktów. Ponieważ nie było to zgodne z intencjami zespołu opracowującego zestawy testowe, zastosowano inną, niezrównoważoną skalę ocen pojedynczego pytania testowego SM_2, posiadającą następującą reprezentację w liczbach rozmytych

$$SM_2 = \{(0,5;0;1,25), (1,25;2,5;0,75), (0,75;4;0,5), (0,5;5;0,5)\} = \{A_0, A_1, A_2, A_3\}.$$

Przy tak przyjętej skali ocen pojedynczego pytania testowego uzyskano wystarczającą zgodność przedziałów ocen wyznaczanych metodą klasyczną i z zastosowaniem liczb rozmytych.

Każdą z przedstawionych metod zastosowano do oceny 118 rzeczywistych zestawów testowych wybranych losowo z puli 354 zestawów.

Wyniki oceniania przedstawiono w tabeli 4.

Tabela 4

Tabela wyników oceniania zestawu testów na podstawie tabeli reguł oraz z zastosowaniem liczb rozmytych dla dwóch skal ocen: SM_1 i SM_2

Ocena	Liczba prac		
	Metoda klasyczna	SM_1	SM_2
Bardzo dobra	4	1	3
Dobra plus	3	6	11
Dobra	18	18	30
Dostateczna plus	18	30	28
Dostateczna	75	26	22
Niedostateczna plus		20	24
Niedostateczna	–	17	2

Analizując dane z tabeli 4, łatwo można zauważyć, że ocena dokonywana z wykorzystaniem liczb rozmytych jest oceną korzystniejszą dla studenta (szczególnie w przypadku zastosowania skali SM_2). Dokładniejsza analiza wyników (niestety nie jest możliwe zamieszczenie ich wszystkich przy ograniczonej objętości tekstu) wskazuje na zróżnicowanie ocen w obrębie grup studentów, którzy zdobyli taką samą liczbę punktów. Zdarzają się przypadki przesunięcia do grupy z niższą oceną oraz przypadki przesunięcia do grupy z wyższą oceną. Zależy to od wyników oceny rozwiązań poszczególnych zadań. Takie zróżnicowanie, w przypadku oceny dokonywanej z wykorzystaniem tabeli reguł wyboru, nie występuje.

Przykłady takich przesunięć pokazano w tabeli 5 porównując oceny wynikowe otrzymane z wykorzystaniem reguł wyboru (metoda klasyczna) oraz z zastosowaniem liczb rozmytych i skali SM_2, dla przypadków uzyskania 28 i 18 punktów sumarycznych.

Tabela 5
Przykład zróżnicowania ocen w przypadku zastosowania metody oceny zestawu testowego z wykorzystaniem liczb rozmytych dla studentów, którzy zdobyli tę samą sumaryczną liczbę punktów

Liczba punktów	Ocena wystawiana metodą klasyczną	Liczba zadań rozwiązanych za odpowiednią liczbę punktów				$A_{FM1} \subset SM_2$	Ocena wystawiana z zastosowaniem skali SM_2
		0	1	2	3		
28	Bardzo dobry	0	1	0	9	4,725	Dobry plus
		0	0	2	8	4,775	Bardzo dobry
18	Dostateczny	2	1	4	3	3,350	Dostateczny plus
		3	1	1	5	3,225	Dostateczny
		2	3	0	5	3,250	Dostateczny plus
		2	2	2	4	3,300	Dostateczny plus
		1	4	1	4	3,325	Dostateczny plus

5. Wnioski

Można zastosować trapezoidalne liczby rozmyte do reprezentacji przyjętej skali ocen (SM). Przedstawiona metoda reprezentacji pozwala ocenić jakość skali. Wydaje się, że poprawna skala ocen powinna być skalą zrównoważoną.

Wykorzystując zasadę rozszerzenie operacji matematycznych ze zbiorów nierozmytych, można obliczyć ocenę średnią $(\overline{A_P})$ na podstawie ocen cząstkowych, nawet uwzględniając różne od równowagowych współczynniki wag (w_i) przyjmowane dla poszczególnych reprezentantów ocen cząstkowych.

Na podstawie reprezentującej ocenę średnią liczby rozmytej, można dokonać wyznaczenia oceny końcowej (A_{FM}), wykorzystując trójkątną liczbę rozmytą generowaną przez ocenę średnią $(\overline{A_P})$ oraz właściwości przecięcia zbiorów rozmytych.

Proponowana metoda wymaga zapamiętania do dalszych procesów ewaluacji, liczb rozmytych reprezentujących zarówno ocenę średnią, jak i ocenę końcową. Jest to bardzo korzystne w przypadku zastosowania prezentowanych algorytmów w systemach ekspertowych modułów ewaluacji w SZN. Wydaje się też wymagane w przypadku porównywania ocen wystawionych według różnych skal, na przykład: ocen tego samego przedmiotu wystawionych w dwóch różnych uczelniach. Taki algorytm wyznaczania ocen końcowych także wydaje się zgodny z tradycyjnym sposobem oceniania przez nauczyciela.

Działanie proponowanych algorytmów sprawdzono, wyznaczając ocenę końcową dla zestawu testów różnego rodzaju. Wyniki porównano z ocenami wystawionymi przez nauczycieli dla 118 przypadków testu przeprowadzonego w rzeczywistości. Zaproponowana metoda pozwala zróżnicować oceny w grupach studentów, którzy uzyskali taką samą liczbę punktów sumarycznych z całego zestawu, co wydaje się korzystne z punktu widzenia nauczyciela (pozwala precyzyjniej określić poziom przyswojenia wiedzy i zdobytych umiejętności).

Dalsze prace prowadzone w przedstawionym obszarze badań skupiać się będą wokół tematów związanych z zastosowaniem liczb i zbiorów rozmytych w ewaluacji dla innych modułów SZN oraz konstrukcją reguł systemu ekspertowego opartych na zbiorach rozmytych.

Literatura

[1] Debuis D., Prade H.: *Operations on fuzzy numbers*. Intern. Journal System Science 9, 1978, 613–626

[2] Grandbastien M.: *Teaching expertise is at the core of ITS Research.* International Journal of Artificial Intelligence in Education 10, 1999, 335–349

[3] Niemierko B.: *Między oceną szkolną i dydaktyką. Bliżej dydaktyki*. WSiP Warszawa, 1997

[4] Przybyszewski K.:*Tutoriale i trenażery umiejętności w nauczaniu zdalnym*. Automatyka 3 (9), 2005, 799–809

[5] Przybyszewski K.: *A new evaluation metod for e-learning systems.* [w:] L. Rutkowski *et al.* (Eds.): ICAISC 2006, LNAI 4029, Springer-Verlag Berlin Heidelberg, 2006, 1209–1216

[6] Przybyszewski K., Cader A., Filutowicz Z.: *Zarządzanie informacją w interaktywnych systemach nauczania.* Zeszyty Naukowe WSHE 4 (9), 2000, 90–102

[7] Rutkowski L.: *Metody i techniki sztucznej inteligencji*. Warszawa, PWN 2005