

Mapmaker Akademii Wojny 2018

„Przebieg prac moderatora konkursowego”
by Michael.

@Licencje: Poniższe przemyślenia zostały ułożone w jedną całość tylko i wyłącznie przeze mnie, na podstawie kilkuletniej obserwacji zachowań pojawiających się w trakcie konkursu, a także w oparciu o wiedzę z zakresu analizy i statystyki. Wszelkie próby kopiowania, przerabiania i wykorzystywania tego dokumentu, tylko za zgodą autora.

@Kontakt: mapmaker.heroes3@gmail.com

Spis treści:

Kryteria oceny – wstęp teoretyczny	_____	strona 3	_____
Sposoby realizacji kryteriów	_____	strona 4	_____
Działania podjęte podczas trwania konkursu	_____	strona 6	_____
Dokumentacja mailowa	_____	strona 7	_____
Podsumowanie i propozycje	_____	strona 8	_____

Kryteria oceny – wstęp teoretyczny

Sprawozdanie zacznę od wymienienia kilku ważnych kwestii związanych z oceną map, a mianowicie wspomnę o głównych kryteriach, którymi się posługiwałem. Mimo długości tej analizy uważam, że warto się z nią zapoznać w celu pełniejszego zrozumienia skąd w trakcie konkursu pojawiły się takie, a nie inne decyzje. Weryfikacja ocen w moim odczuciu przebiegała na dwóch płaszczyznach:

- I. Na płaszczyźnie sprzętowej tzn. operując danymi udostępnionymi przez Organizatora, a mianowicie adresami IP recenzentów, oraz adresami IP autorów map umożliwiającymi wyłapanie recenzji nabitych z multikont, map należących do tych samych autorów itd. Organizator zapewnił również anonimowość poszczególnych recenzentów również dla moderacji konkursowej. Dopiero po akceptacji odrzucenia danej recenzji otrzymywaliśmy informację o autorze recenzji.
- II. Oraz na płaszczyźnie statystycznej. Poniżej wymieniam 3 główne kryteria, którymi się kierowałem podczas tego etapu weryfikacji ocen:
 - 1) W każdej statystyce niezależnie od typu pojawiają się wartości skrajne (tzw. fluktuacje statystyczne), które przy badaniu tendencji (w tym wypadku średniej końcowej) należy wykluczyć z obliczeń. Tłumacząc to zjawisko prostym przykładem: Załóżmy, że daną mapę oceniło 10 użytkowników i każdy z nich napisał takiej samej długości i merytoryki recenzje.
8 z nich wystawiło recenzje w przedziale między 3-4, ostatnich dwóch użytkowników oceniło mapę na 1 oraz 5. Żaden z tej dwójki nie miał racji niezależnie od swojej recenzji oraz jej argumentacji. Wynika to z tego, że skoro jeden dostrzegł same wady i żadnych zalet, a drugi same zalety bez żadnych wad to obie recenzje były nieobiektywne, bo mapa zawiera i jedno i drugie. Wynika to również z tego, że pozostałych 8 użytkowników wypowiedziało się bardziej wiarygodnie ponieważ bliżej średniej. To jeden z największych paradoksów statystyki. Każda osoba wierzy że oceniła pewną sytuację obiektywnie, a jednak niektóre z nich zawyżyły bądź zaniżyły końcowe obliczenia w stosunku do innych. Po to jest także obliczanie odchylenia standardowego (żeby wyznaczyć pewną granicę błędu dla subiektywnej oceny recenzenta, powyżej której recenzja jest usuwana, ale o tym za chwilę).
 - 2) Świat byłby zbyt piękny gdyby można wszystko ocenić tylko za pomocą pierwszego kryterium. Wynika to z tego że ma ono pewną istotną wadę, a mianowicie założenie o takiej samej długości recenzji i jej merytorycznym uzasadnieniu. Jeśli będziemy mieli do czynienia z sytuacją identyczną jak w powyższym przykładzie ale z jedną różnicą: 8/10 recenzentów poda ocenę bez dostatecznego uzasadnienia, a osoby z 1 i 5 podadzą długie wypracowania. W tym przypadku pojawia się grono ekspertów (moderatorów konkursowych), którzy weryfikują indywidualnie poprawność merytoryczną wszystkich 10 recenzji. O ile w tym punkcie analiza statystyczna nam bezpośrednio nie pomoże, o tyle przyspieszy odnalezienie map w których wymagane jest dokonanie analizy za pomocą właśnie tego kryterium.
 - 3) Trzecim kryterium (poza fluktuacjami z 1 punktu i merytoryką z 2 punktu) jest porównanie. Jest to kryterium w którym oceniamy względem siebie poszczególne recenzje jako ciąg recenzji wystawionych przez konkretnego autora. Jest to tak naprawdę kryterium częściowo sprawdzające poprawność dwóch poprzednich kryteriów.

Sposoby realizacji kryteriów

W trakcie trwania konkursu sporządziłem tabelę zawierającą informację o każdej mapie wraz z ocenami wystawionymi przez recenzentów, która dostępna jest w załączniku do tego pliku o nazwie: „Podsumowanie.xlsx”.

Dzięki temu, że moderatorzy konkursowi zabierają się do pracy dopiero 2 tygodnie przed zakończeniem konkursu dostajemy pełną gamę ocen, pozwalającą na w miarę rzeczywiste odzwierciedlenie pewnych tendencji w grupie końcowych wyników. Teraz omówię sposoby realizacji powyższych 3 głównych kryteriów:

1) Analiza fluktuacji i średnich.

- Za pierwszy sposób realizacji tego kryterium przyjąłem wyliczanie odchylenia standardowego dla wszystkich map, a następnie jego uśrednienie. Widzimy w tabelce, w kolumnie odchylenie standardowe wyniki dla poszczególnych map przed oraz po działaniach moderacji konkursowej. Im mniejsza jest średnia odchylenia, tym więcej ocen znajduje się bliżej średniej, a co za tym idzie, tym więcej ocen i recenzji jest ze sobą zgodnych. Na czerwono zaznaczone są wyniki, które znajdują się ponad średnią odchylen (w naszym przypadku $> 0,78$). Mapy powyżej takiego odchylenia można traktować jako takie w których pojawiły się naciągane oceny. Mapy w których odchylenie standardowe jest mniejsze od uśrednionego nie powinny zawierać naciąganych recenzji z **punktu widzenia tego kryterium**.
- Za drugi sposób przyjąłem analizę centylową. W pliku dodanym jako załącznik do tego dokumentu o nazwie: „Podsumowanie.xlsx” widzimy 2 typy formatowania. Jedno (kolory poszczególnych ocen) to położenie danej oceny względem całej grupy wyników. Poszczególne centyle określają nam ile wyników było niższych bądź wyższych od 50%. Drugie to strzałki pionowe bądź pochylone, które oznaczają wpływ na docelowe wyniki. Zatem wnioskiem jest, że recenzje, które są koloru czerwonego wraz z pionowymi strzałkami skierowanymi w górę i dół dla danej mapy powinny zostać szczegółowo zweryfikowane (pod kątem zawyżania i zaniżania oceny). Recenzje, w których ocena ma pochyloną strzałkę nie powinny zawierać błędów z **punktu widzenia analizy centylowej**.

2) Analiza merytoryki.

- Zanim dokonamy realizacji tego kryterium warto zauważyć prostą zależność - im więcej dany użytkownik wystawił długich i merytorycznych recenzji tym większa szansa że będą one uczciwe, a co za tym idzie tym mniejsza że będzie chciał naciągać oceny. Oczywiście nie można zakładać, że ten sposób realizacji pozwala na 100% wyłapanie wszystkich naciągaczy niemniej jednak z pewnością pozwala na wyłapanie dużej ich części.
- Innym wnioskiem jest też fakt, że recenzje wystawione pojedynczo czy nawet podwójnie będą budziły duże wątpliwości. Jeszcze więcej będą budziły pojedyncze recenzje dodane przez świeżo zarejestrowanych użytkowników portalu (zwłaszcza takich, których konta powstały w trakcie trwania konkursu). Takim recenzjom należy przyrzeć się z dwukrotnie większą precyzją.

3) Porównanie ocen wystawionych przez konkretnego recenzenta.

- Za pierwszy sposób realizacji tego kryterium przyjąłem średnia ze wszystkich ocen wystawionych przez konkretnego recenzenta. Powinna ona oscylować wokół oceny środkowej czyli 3.

Zatem wnioskiem jest, że recenzenci posiadający średnią między 4-5 oraz między 1-2 będą budzili wątpliwości **z punktu widzenia tego kryterium**.

- Drugim sposobem na sprawdzenie czy dany recenzent nie naciągał ocen, jest porównanie jego oceny względem pozostałych przez niego dodanych przez pryzmat różnic w średnich ocenach między poszczególnymi mapami. Brzmi skomplikowanie, ale istota tego rozwiązania jest bardzo prosta. Wystarczy spojrzeć na poniższy przykład:

	Mapa A	Mapa B
Recenzent R1	Ocena 5	Ocena 2,5
Recenzent R2	Ocena 3	Ocena 1,5

Przykładowo dany recenzent R1 ocenił mapę A na 5, a mapę B na 2,5. Inny recenzent R2 ocenił tą samą mapę A na 3, a mapę B na 1,5. W świetle tego sposobu oboje recenzenci **mimo dodania różnych ocen** zachowali się poprawnie. Wynika to z tego, że oboje dokonali rozpoziomowania między poszczególnymi mapami (Mapa A według obu jest lepsza niż mapa B).

	Mapa A	Mapa B
Recenzent R1	Ocena 5	Ocena 2,5
Recenzent R2	Ocena 3	Ocena 1,5
Recenzent R3	Ocena 2	Ocena 4

W powyższym przykładzie widzimy, że albo dwóch recenzentów R1 i R2 wystawiło nieprawidłowe oceny, albo recenzent R3. Z punktu widzenia statystyki im więcej ocen zachowuje podobną tendencję (w tym przypadku im więcej jest recenzentów postaci R1 oraz R2), tym większa szansa, że pojedyncze odchylenie (tu w postaci recenzenta R3) będzie nieprawidłowe. Tłumacząc powyższy przykład mapa A jest według R1 i R2 lepsza niż B, z taką tendencją nie zgadza się recenzent R3 skąd wynika że gdzieś jest błąd (a w naszym przypadku naciągnięcie oceny).

Działania podjęte podczas trwania konkursu

Tutaj kończy się wstęp teoretyczny do tego skomplikowanego tematu jakim jest weryfikacja poprawności recenzji. Teraz opisze (skrótowo) jakie działania miały miejsce w trakcie trwania konkursu:

- 1) Mechaniczne usuwanie recenzji nabytych z tego samego adresu IP.
- 2) Usuwanie recenzji krótkich i nieuzasadnionych, niepozwalających na stwierdzenie czy autor faktycznie w daną mapę grał, bądź pojedynczych recenzji, które zostały nieodpowiednio uargumentowane z kont świeżo dodanych do portalu. Przykładowo: użytkowników, którzy zarejestrowali się tylko po to żeby raz odwiedzić portal na 10 minut dodać recenzję 5 i nigdy więcej nie wrócić, będzie budził podejrzenia, a jego recenzja będzie poddana podwójnej weryfikacji.
- 3) Kolejnym krokiem była dokładna analiza statystyczna recenzji. W tym miejscu należy zaznaczyć, że dokonaliśmy podziału pracy wraz z drugim moderatorem konkursowym – New Year'em. Ja zająłem się wysłaniem maili do osób, które w jakiś sposób naruszyły powyższą statystykę.
- 4) Dokument „Podsumowanie.xlsx” pokazuje różnicę, między początkowym stanem rzeczy (przed naszą moderacją, czyli stan konkursu jaki byłby gdybyśmy nic nie usuwali) oraz stanem końcowym (po naszej moderacji, usuwaniu/ modyfikowaniu recenzji - czyli stan obecny). Nasze działania spowodowały że zostały usunięte wszystkie nabite recenzje z dwóch różnych multikont (w dokumencie podkreślonych na szaro oraz jasnoniebiesko). Została również usunięta duża część pojedynczych recenzji wystawionych przez świeżo zarejestrowanych użytkowników oraz udało się odnaleźć powiązania autorów z czterema mapami wstawionymi przez te same osoby. Dzięki tym wszystkim zmianom obserwujemy znaczącą poprawę końcowych odchyień standardowych (bo aż o ~10,5% średnio dla każdej mapy).

Dokumentacja mailowa

W celach dokumentacyjnych przeklejam tu treść wszystkich powiadomień mailowych przeze mnie wysłanych:

Od: Michael
Do: Clanger
Wysłany: 2018-09-23, 23:26
Temat: Moderacja ocen konkursowych

Witam,

W trosce o poziom konkursu w imieniu moderacji wysyłam poniższy mail, który jest prośbą o ponowną weryfikację przedstawionych recenzji oraz zmianę ocen końcowych w tych recenzjach.

Zgodnie z wynikami analizy statystycznej, wygląda na to że 4 poniższe recenzje mają:

- > ocenę zaniżoną (*)- Ocena 2 dla "Komnaty Utopionego"
- > ocenę zawyżoną (*)- Ocena 4,5 dla "Zapomniane Wyspy"
- > ocenę zawyżoną (*)- Ocena 3 dla "Tears of the Dragon"
- > ocenę zaniżoną (*)- Ocena 1,5 dla "In comes the flood"

(*)-względem pozostałych Twoich recenzji.

Uzasadnienie:
Mapy takie jak "Tears of the Dragon" czy "Zapomniane Wyspy" są mapami niegrywalnymi w opinii zdecydowanej większości recenzentów. Moderacja z dużym prawdopodobieństwem usunie po zakończeniu konkursu obie te mapy, w przeciwieństwie do map "In comes the flood" czy "Komnat Utopionego". Ocenianie słabych i bardzo słabych map na 4,5; 3 a map dobrych i grywalnych (nie koniecznie wygrywających konkurs) na 1,5;2 uznają za naciąganie i próbę zafałszowania wyników dla tych map.

Wymagane zmiany:

- > Podwyższenie oceny dla "Komnaty Utopionego" do minimum 3.
- > Zmniejszenie oceny dla "Zapomniane Wyspy" do maksimum 3.
- > Zmniejszenie oceny dla "Tears of the Dragon" do maksimum 2.
- > Podwyższenie oceny dla "In comes the flood" do minimum 3.

Wymagane zmiany zostały dobrane na podstawie minimalnych i maksymalnych ocen pozostałych recenzentów, a także w oparciu o merytorykę ich wypowiedzi.

W przypadku braku ustosunkowania się do powyższego maila recenzje dodane przez Autora do tych map zostaną usunięte.

Z poważaniem
Moderator konkursowy: Michael

Od: Michael
Do: kellavend
Wysłany: 2018-09-23, 23:24
Temat: Moderacja ocen konkursowych

Witam,

W trosce o poziom konkursu w imieniu moderacji wysyłam poniższy mail, który jest prośbą o ponowną weryfikację ocen końcowych dla poniższych recenzji:

- > Ocena 4 dla "21000 mil podmorskiej żeglugi"
- > Ocena 4 dla "In comes the flood"
- > Ocena 5 dla "Artificem Commendat Opus"
- > Ocena 5 dla "Damage Inc"
- > Ocena 5 dla "The Offensive"
- > Ocena 4,5 dla "Wyprawa przeciwko Zepsuciu"

Uzasadnienie:
Wszystkie recenzje są dobrze uargumentowane merytorycznie, jednakże mają również bardzo wysoką ocenę końcową przez co nie wiadomo która mapa dla Autora była najlepsza, a która była mniej doskonała. Co prawda nie ma żadnych wymagań odnośnie liczby wystawionych ocen 5 czy 4 jednak mamy 10, a nie 3 stopniową skalę, w której można oddać dokładnie finalną ocenę danej mapy.

Zalecane zmiany:
Prośbą o uszeregowanie tych map według subiektywnej oceny Autora recenzji. Rozpoziomowanie o przykładowe +0,5 -0,5 oceny dla wystawionych recenzji.

Brak uszeregowania bądź zmiany oceny nie będzie skutkowało usunięciem recenzji (ze względu na dużą wartość merytoryczną), a jedynie może budzić wątpliwości innych recenzentów podczas oglądania wyników analizy statystycznej konkursu udostępnionej po zakończeniu konkursu.

Z poważaniem
Moderator konkursowy: Michael

Podsumowanie i propozycje

Podsumowując patrząc po sposobie opisu dodawanych recenzji widzę bardzo pozytywne zmiany, w stosunku do poprzednich edycji. Co prawda było mniej recenzji i map niż w poprzednich edycjach, jednak z drugiej strony znakomita większość recenzentów napisała dłuższe i bardziej merytoryczne recenzje niż w poprzednich latach (pozwalające w czytelny sposób zrozumieć dlaczego została postawiona taka, a nie inna ocena końcowa). Poza tym, tylko jedna mapa nie zakwalifikowała się do końcowego rankingu, a było to głównie spowodowane przez samego Autora sztucznie nabijającego recenzje. Porównując do poprzednich edycji uważam, że wprowadzenie moderatorów konkursowych było dobrym zabiegiem co widzimy po znacznie spokojniejszym przebiegu konkursu niż w poprzednich latach. Uważam również, że wypracowany przez mnie system analizy statystycznej powinien również znaleźć zastosowanie także w kolejnej edycji Mapmakera Akademii Wojny.

Dziękuję czytelniku, że dotarłeś do końca tego artykułu, mam również nadzieję, że motywy którymi się kierowaliśmy podejmując poszczególne decyzje stały się już jasne.

Pozdrawiam i czekam na kolejne mapy nadsyłane w najbliższych latach.

Moderator konkursowy:
Michael