

Ustawka 2020

Zob. [ustawka2020.pl](#)

W najgorętszym okresie sesji egzaminacyjnej – w połowie czerwca – piętnaście drużyn z różnych polskich uczelni walczyło w hackathonie zorganizowanym przez IBM Polska przy współpracy Wydziału Matematyki, Informatyki i Mechaniki UW. Cel był pozornie bardzo prosty – przez dziesięć kolejnych dni przewidzieć trzy liczby: liczbę zachorowań, wyzdrowień i zgonów na Covid-19, podawane nazajutrz przez Ministerstwo Zdrowia. O miejscu drużyny w końcowym rankingu decydowała suma kwadratów błędów wszystkich prognoz.

Zwycięzcy zgodzili się podzielić z Czytelnikami *Delty* tajemnicami swoich sukcesów. Poniżej zamieszczamy ich bezpośrednie krótkie relacje, spisane podczas wakacji.

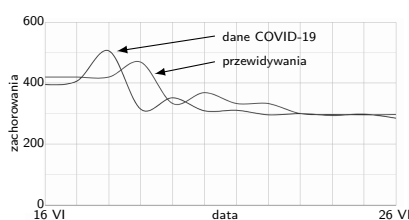
Starsza siostra rozsądku, intuicja

I miejsce

Tomasz BAJKACZ*, Kamil CHOIŃSKI*, Fabian JEZIOREK*

* Wydział Automatyki Elektroniki i Informatyki, Politechnika Śląska

Prognozując wartości na następny dzień, uwzględnialiśmy takie parametry, jak: dane z ostatnich dwóch tygodni, dzień tygodnia (bo np. w weekendy wykonywano mniej testów), bieżące wiadomości, dane meteorologiczne. Codziennie wieczorem zbieraliśmy się i wspólnie analizowaliśmy te dane, i na tej podstawie przygotowywaliśmy, polegając na naszej intuicji, ostateczne typy na następny dzień. Dominujący wpływ na nasze prognozy miały wartości z bieżącego dnia, co możemy zauważyć na wykresie obok.



Początkowo mieliśmy plan, aby stworzyć algorytm dostosowujący krzywą do wykresu, jednak byliśmy wtedy w trakcie sesji i uznaliśmy, że równie dobrze możemy dokonać naszych przewidywań bez wykorzystania modelu. Stwierdziliśmy, że chaotyczna zmienność liczby zachorowań i wyzdrowień uniemożliwia stworzenie działającego modelu. Uznaliśmy też, że nie warto przykładać większej wagi do prognozy liczby zgonów, ponieważ nie miała ona znaczącego wpływu na ostateczny wynik.

Czynnikiem, który istotnie wpłynął na nasze przewidywania, była wiadomość o wyzdrowieniach górników – zwiększyliśmy wtedy sporo wartość prognozy wyzdrowień, co okazało się celnym strzałem. Od tego momentu z pomyślnym skutkiem typowaliśmy większe liczby wyzdrowień od zachorowań.

Szlifując aż do ostatniego dnia

II miejsce

Aleksandra MUSIAŁ**, Michał MAŚLANKA**, Bartosz MAZIARKA**

** Wydział Elektrotechniki i Informatyki, Politechnika Rzeszowska im. Ignacego Łukasiewicza

Nasze podejście do tematu było stosunkowo luźne – bardzo podstawowe przewidywanie szeregów czasowych metodą podwójnego wygładzania wykładniczego (model Holta–Wintersa), dostępną w pakiecie R:

$$s_t = \alpha y_t + (1 - \alpha)(s_{t-1} + b_{t-1}),$$

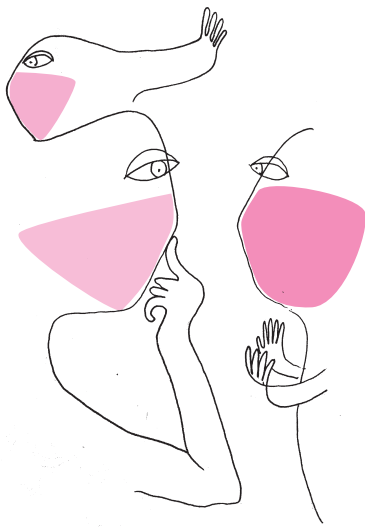
$$b_t = \beta(s_t - s_{t-1}) + (1 - \beta)b_{t-1},$$

gdzie y_t jest wartością zmiennej prognozowanej w chwili t , a b_t – wygładzonym przyrostem trendu. Współczynniki α i β znajduje się eksperymentalnie. Mając te wartości, prognozę na chwilę $t + m$ oblicza się ze wzoru

$$\hat{y}_{t+m} = s_t + mb_t.$$

Szeregi nie miały wyraźnego trendu (był on bardzo łagodnie wzrostowy), więc współczynnik β grał małą rolę w prognozach; współczynnik α był dużo większy. Z szeregów usunęliśmy dane sprzed umownej daty stabilizacji sytuacji, czyli sprzed 2 kwietnia. Przed tą datą wirus rozwijał się i mogłoby to wpłynąć na prognozy.

Mając wstępną prognozę z modelu Holta–Wintersa, patrzyliśmy na sytuację również z perspektywy zdrowego rozsądku, biorąc pod uwagę takie czynniki, jak:



liczbę hospitalizowanych osób, liczbę osób pod respiratorami, dzień tygodnia (np. spadki wyzdrowień w niedzielę), a także czynniki środowiskowe, jak wyjazdy na wakacje. Korekty zwykle się zdarzały, ale najczęściej nie większe niż 10% dla zachorowań i wyzdrowień i 20% dla śmierci.

Na początku hackathonu liczba zachorowań była stabilna, dlatego też co do liczby zachorowań można było śmiało oprzeć się na prognozach z modelu. Przy wyzdrowieniach należało wziąć pod uwagę falę zachorowań na Śląsku, która rozpoczęła się 11 dni od dnia początku zawodów – mieliśmy świadomość, że zarówno śmierci, jak i wyzdrowienia mogą nagle znacząco wzrosnąć, dlatego wyzdrowienia zwiększane były o około 40% w stosunku do prognozowanej wartości. Dość cynicznie stwierdziliśmy, że śmierci mają mały wpływ na błąd końcowy, więc przyjęliśmy, że wartość ta będzie oscylowała w okolicach 10 (czasem dodamy 1, czasem odejmiemy 1, zależnie od przecucia).

Przez kilka dni metoda sprawdzała się... aż nadszedł dzień sądu, czyli 18 czerwca. Wtedy zachorowania gwałtownie spadły, podczas gdy wyzdrowienia mocno wzrosły, nie mówiąc już o niezwykle wysokiej liczbie zgonów. To był moment, w którym należało zrewidować strategię. Stanęliśmy przed problemem – czy spadek zachorowań jest anomalią, czy jednak wartości na niskim poziomie będą się utrzymywały? Chociaż dane historyczne sugerowały, że tak nie będzie, postanowiliśmy przez dłuższy czas prognozować niższe wartości zachorowań, co bardzo się opłaciło.

Wzrost wyzdrowień dało się przewidzieć, ale nie spodziewaliśmy się, że będzie aż tak wysoki. Wartości ponad 380 zdarzyły się w historii tylko 3 razy, i to przy kilkudniowych seriach zachorowań rzędu 450. Zaskoczyło nas to zarówno 19 czerwca, jak i później. Po przebłysku geniuszu w niedzielę 22 czerwca (obniżenie wartości wyzdrowień poniżej 400 z uwagi na niedzielę) spodziewaliśmy się, że wyzdrowienia przestaną rosnąć i wrócą do poziomu około 350, ponieważ większość poszkodowanych w śląskiej fali wróciła już do zdrowia. Przez kilka dni upieraliśmy się przy tej strategii, rezygnując dopiero ostatniego dnia, kiedy to udało nam się postawić najtrafniejszą prognozę w zawodach: o błędzie rzędu 2‰!

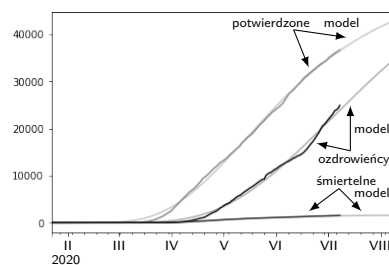
Od innowacji do pandemii

III miejsce

Paweł GOCAL***, Przemysław SARNACKI***, Rafał SZCZEPANIK***

*** Wydział Elektroniki, Wojskowa Akademia Techniczna

E. Le Nagard i A. Steyer, *La prévision des ventes d'un nouveau produit de télécommunication: probit ou théorie des avalanches?*, Rech. Appl. En Mark., t. 10, nr 1, s. 57-68, 1995.



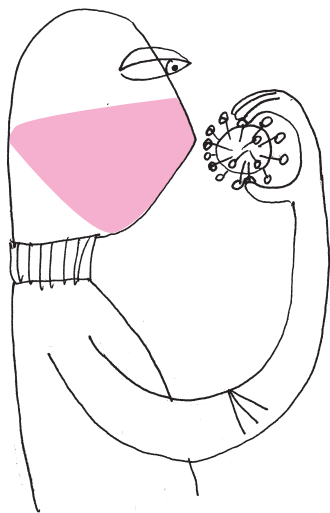
Dopasowanie modelu do danych historycznych

Do stworzenia modelu przewidującego liczbę zakażonych, zdrowych i zmarłych wykorzystaliśmy dane o osobach dotychczas już zarażonych, uwzględniając liczbę zakażeń na każdy poprzedni dzień. Pozwoliło to na określenie trendu. Ponieważ wirus przenosi się drogą kropelkową, najczęściej przy rozmowie, kichaniu i kaszlu, założono, że może się rozprzestrzeniać podobnie jak innowacje. Użyliśmy zaproponowanego przez francuskich naukowców modelu rozprzestrzeniania się innowacji w telekomunikacji (źródło obok):

$$n_t = N \cdot (1 - e^{-at})^\alpha,$$

gdzie n_t oznacza prognozowaną wartość w chwili t , natomiast N , a oraz α to współczynniki modelu. Ich wartości wyznaczaliśmy tak, by zminimalizować błąd średniokwadratowy dopasowania do danych historycznych. Model zaimplementowaliśmy w Pythonie przy użyciu bibliotek NumPy, Pandas i SciPy. Z powodu sytuacji epidemiologicznej współpracowaliśmy zdalnie, korzystając z platform Microsoft Teams, Google Colab i Github.

Zasadniczym problemem w przewidywaniu liczby zachorowań i wyzdrowień była niejednakowa liczba wykonywanych dziennie testów na obecność wirusa. Zauważyliśmy, że wyniki w niedzielę i poniedziałek były mniejsze od pozostałych dni tygodnia – z powodu wykonania mniejszej liczby testów w okresie weekendowym. Gdyby liczba testów wykonywanych dziennie była stała, modele prognozujące dawałyby dużo lepsze wyniki.



Dlatego potraktowaliśmy model jako punkt wyjścia i uzupełnialiśmy go o sumę dwóch poprawek, wyznaczanych na podstawie dnia tygodnia i liczby testów. Przykładowo, jeśli testów było mało, to poprawka wynosiła -10% , jeśli dużo, to $+10\%$, a w przeciwnym wypadku $+5\%$. Analogicznie reagowaliśmy w drugim przypadku.

Choć obiecaliśmy sobie, że nie będziemy modyfikować modelu ani jego wyników w trakcie hackathonu, to ze względu na niespodziewane pojawienie się nowego ogniska choroby na Śląsku wzbogaciliśmy zestaw poprawek o dodatkowy parametr z tym związany, co poprawiło jakość prognoz.

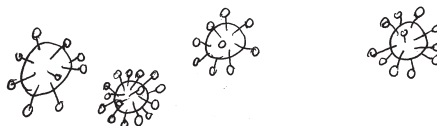
Komentarz kibica

W imieniu Komisji Konkursowej hackathonu, której miałem zaszczyt przewodniczyć, gratuluję zwyciężskim zespołom odwagi zmierzenia się z obu wyzwaniem: prognozowaniem tego, co nieprzewidywalne i... pisaniem do *Delty*. Bardzo byłem ciekaw ich rozwiązań i, zapewne jak wielu Czytelników rubryki, podświadomie oczekiwałem, że najlepsze przewidywania będą efektem użycia spektakularnego modelu matematycznego, być może wymagającego nietrywialnej implementacji. Tymczasem okazało się, że najlepsza prognoza była najzwyczajniej w świecie wynikiem działania trzech, równoległe działających, wszechstronnych sieci neuronowych, trenowanych przez ostatnie 20+ lat: to znaczy mózgowi zwycięskiej trójki. Pracujących w szczególnym trybie, który nazywamy na różne sposoby: *przeczućmi, intuicją, zdrowym rozsądkiem*.

Czy naprawdę powinniśmy się temu dziwić? Przewidywane zjawisko było bardzo skomplikowane, m.in. przez wpływ najróżniejszych zewnętrznych szumów: nieznaną i zmienną liczbę testów, sprawozdawczości, nowych ognisk choroby itp. Nie było czasu na długotrwałe analizy i studia epidemiologiczne; z oczywistych powodów nie można było też zapytać specjalistów... W takiej sytuacji chyba każdy miałby silne przecucie, że postawienie na intuicję jest zgodne ze zdrowym rozsądkiem.

Aby jednak nie nadawać intuicji znaczenia większego, niż zasługuje, warto zauważyć, że całkowicie odmienna strategia – bezrefleksyjna i nad wyraz leniwa (na którą wszakże nikt się nie zdecydował): *jutro będzie tak samo jak dziś*, dawałaby... drugie miejsce w naszym konkursie.

Piotr KRZYŻANOWSKI



Logika implikacji

Aleksy SCHUBERT*

*Wydział Matematyki, Informatyki i Mechaniki, Uniwersytet Warszawski

†Dlaczego w ogóle zajmować się czymś bez sensu? Na przykład dlatego, że czasem ktoś poda nieprawidłowe rozwiązanie zadania, ale stwierdzenie błędu wymaga sporego wysiłku.

Na początek zajmijmy się czymś zupełnie bez sensu[†]. Na przykład zdaniem:

(A) *Jeśli na Merkurym w tej chwili znajduje się człowiek, to na Merkurym znajdują się ślady człowieka.*

To zdanie jednak bylibyśmy gotowi uznać za „bardziej prawdziwe” niż wypowiedź:

(B) *Jeśli na Merkurym w tej chwili znajduje się człowiek, to autor tego artykułu jest Czyngis-chanem.*

Tymczasem na gruncie logiki klasycznej zdania te są równie prawdziwe. W logice tej jeśli pierwszy człon implikacji (np. *na Merkurym w tej chwili znajduje się człowiek*) jest fałszywy albo jeśli drugi człon implikacji (np. *autor tego artykułu jest Czyngis-chanem*) jest prawdziwy, to całe zdanie jest prawdziwe.

Wielu uczniów i studentów burzy się, gdy poda im się taką interpretację wynikania. Jednak w systemie nastawionym wyłącznie na określanie, czy coś jest prawdą, czy fałszem, a takim systemem jest logika klasyczna, wyjścia specjalnie nie mamy. Gdy uprzemy się, że implikacja jest spójnikiem, jak *i* czy *lub*, oraz że