

Projekt SNR

lato 2005

1) Wprowadzenie

Biometrię definiuje się jako ogół zautomatyzowanych metod weryfikacji lub identyfikacji tożsamości osoby na podstawie jej cech fizjologicznych (np. wzór tęczówki oka, kształt dłoni) lub behawioralnych (np. sposób chodzenia, podpis odręczny).

Weryfikacja to potwierdzenie tożsamości osoby, która samodzielnie „przedstawia się” systemowi (np. poprzez podanie PINu lub loginu). Identyfikacja ma miejsce wówczas, gdy nie jest znana *a priori* tożsamość osoby, a w celu ustalenia tożsamości osoby przeszukiwana jest baza danych systemu biometrycznego.

Zarówno w przypadku systemu weryfikującego, jak i systemu identyfikującego, użytkownik jest zapamiętywany w systemie poprzez pewną próbkę wzorcową (wzorzec, *ang. template*). W procesie weryfikacji aktualnie pobrana próbka jest porównywana z odpowiednim wzorcem, zależnym od tego za kogo podaje się użytkownik (porównanie „1 do 1”). W procesie identyfikacji aktualnie pobrana próbka porównywana jest z wieloma wzorcami zapamiętanymi w bazie systemu (porównanie „1 do N”).

Cechą charakterystyczną metod biometrycznych jest fakt, iż wzorzec oraz porównywana próbka nigdy nie są identyczne (przeciwnie do tradycyjnej weryfikacji z użyciem hasła). Wprowadza się różne miary podobieństwa służące określeniu czy dana próbka pochodzi czy nie od tej samej osoby od której pochodzi wzorzec. Błąd polegający na stwierdzeniu, iż próbka pochodzi od tej samej osoby co wzorzec, mimo iż faktycznie wzorzec i próbka pochodzą od innych osób, to tzw. błąd fałszywej akceptacji (*ang. false acceptance*). Błąd polegający na stwierdzeniu, że próbka i wzorzec pochodzą od różnych osób, mimo iż faktycznie pochodzą od jednej osoby, to tzw. błąd fałszywego odrzucenia (*ang. false rejection*).

2) Zadanie projektowe

Zadanie projektowe polega na przebadaniu zastosowania perceptronu Rosenblatta, maszyny wektorów podpierających (SVM) oraz sieci neuronowej o ciągłych funkcjach aktywacji do zmodyfikowanego problemu weryfikacji biometrycznej.

Każda z osób wykonujących projekt otrzymuje indywidualnie przygotowany zestaw danych reprezentujący 20-elementowe wektory cech biometrycznych pobranych od 80 różnych osób. Wektory reprezentują odpowiednio przetworzone, połączone dane pochodzące z pomiarów geometrii dłoni i podpisu odręcznego. Zestaw danych jest podzielony na dwa podzbiory: podzbiór wzorców i podzbiór pojedynczych próbek. Kolejność wzorców i próbek zapisanych w odpowiednich plikach jest taka sama, tzn. k -ty wzorzec i k -ta próbka pochodzą od tej samej osoby. W zestawie danych próbki oraz wzorce pochodzące od danej osoby występują tylko raz (tzn. nie istnieją dwie próbki/wzorce pochodzące od tej samej osoby). Dostępne dane należy losowo podzielić na zbiór trenujący (dane od $n = 60$ osób) oraz zbiór testujący (dane od $n = 20$ osób).

Zarówno w zbiorze trenującym jak i zbiorze testującym należy przeprowadzić następujące operacje. Korzystając z powyżej opisanych danych utworzyć n wektorów „zgodnych” Z , powstałych poprzez połączenie wektora wzorca W i wektora próbki P , pochodzących od tych samych osób (nowy, „powiększony” wektor to wektor $Z = \begin{bmatrix} W \\ P \end{bmatrix}$).

Następnie losowo utworzyć $2n$ wektorów “niezgodnych” Z , tzn. takich dla których dla każdego W , P pochodzi od dwóch różnych losowo wybranych osób, innych niż ta, od której pochodzi W .

Tak przygotowane wektory Z będą służyły jako dane wejściowe do zadania klasyfikacji na dwie klasy: wektorów “zgodnych” i wektorów “niezgodnych”. Klasyfikatory powinny zostać zrealizowane przy wykorzystaniu trzech technik:

- a) perceptronu Rosenblatta,
- b) maszyny wektorów wspierających (SVM) o dowolnie wybranym jądrze,
- c) sieci neuronowej o ciągłych, dowolnie wybranych funkcjach aktywacji.

W ramach badań projektowych należy:

- przebadać jakość klasyfikacji dla zbioru trenującego i dla zbioru testującego (w szczególności podając stosunek liczby błędnie zaklasyfikowanych wektorów do wszystkich wektorów)
- zbadać wpływ wartości parametrów na jakość klasyfikatorów
- porównać sposób tworzenia poszczególnych klasyfikatorów (uczenie sieci, wyznaczanie parametrów SVM, ...) zwracając uwagę na nakład obliczeń oraz uzyskaną jakość klasyfikacji

3) Narzędzia

Nie stawia się żadnych ograniczeń dotyczących wykorzystywanych w badaniach narzędzi. Ze względu na dostępność wielu przydatnych funkcji zaleca się wykorzystanie środowiska MatLab (w szczególności Neural Network Toolbox oraz Optimization Toolbox).

4) Ocena projektu

Projekt ma charakter badawczy. Ocenie będzie podlegała wszechstronność i dociekliwość w przeprowadzanych badaniach, a także umiejętność wyciągania wniosków z uzyskanych rezultatów. Jakość implementacji nie podlega ocenie, jakkolwiek źródła powinny zostać dostarczone wraz ze sprawozdaniem końcowym.

5) Informacje dodatkowe

Prowadzący projekt: Łukasz Stasiak, Istasiak@elka.pw.edu.pl, konsultacje: pok. 518A, czwartki w godz. 12.00-15.00 lub w innym terminie po uzgodnieniu (mail).

Termin oddania projektu: zgodnie z informacjami podanymi na wykładzie.