

Piotr CZEKALSKI
Politechnika Śląska, Instytut Informatyki

PROGNOZOWANIE ZAPOTRZEBOWANIA NA USŁUGI MEDYCZNE W KASACH CHORYCH

Streszczenie. W artykule przedstawiono wybrane zagadnienia z dziedziny ilościowego prognozowania zapotrzebowania na usługi medyczne w oparciu o metody statystyczne. Bazując na wyznaczeniu trendów w zapotrzebowaniu na usługi medyczne omówiono problematykę wykorzystania przedłużenia analitycznego funkcji do prognozowania oraz zastosowanie numerycznego algorytmu aproksymacji średniokwadratowej jako sposobu wyznaczenia trendu na podstawie szeregu czasowego.

Słowa kluczowe: prognozowanie, ochrona zdrowia, medycyna.

REQUIREMENT PROGNOSING FOR MEDICAL SERVICES

Summary. This article covers selected problems on requisition for medical services total amount prognosis, based on statistic methods. Article describes usefulness of analytical function extension in junction with least squares method and its numerical implementation as a way of prognosis based on time series of values.

Keywords: prognosis, healthcare, medicine.

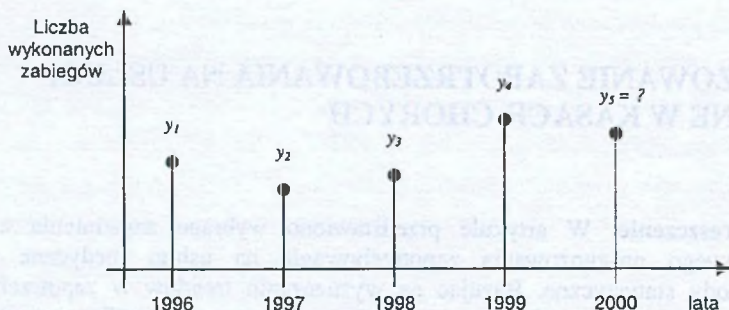
1. Wyznaczanie trendu na podstawie szeregu czasowego

Miarodajna prognoza zapotrzebowania na usługi medyczne ma kluczowe znaczenie dla kasy chorych na etapie planowania wielkości i struktury budżetu.

Powstaje pytanie, czy można wykorzystać do tego celu rzeczywiste dane o realizowanych usługach medycznych w latach poprzednich. Mamy wtedy do czynienia z danymi występującymi w postaci tzw. szeregu czasowego, w którym obserwacje (pomiar) są zależne. Metody analizy takich ciągów obserwacji zależnych opisano m.in. w [5].

Z formalnego punktu widzenia problem prognozy zapotrzebowania na usługi medyczne można scharakteryzować w następujący sposób:

- dany jest uporządkowany w czasie (czas jest zmienną niezależną) zbiór rzeczywistych wartości zapotrzebowania na usługi medyczne, realizowane w kolejnych latach: y_1, y_2, \dots, y_n ,
- na tej podstawie chcemy przewidzieć zapotrzebowanie na takie usługi w roku $n+1$, patrz rys. 1.



Rys. 1. Reprezentacja przykładowego szeregu czasowego
Fig. 1. Example of time series

Wybór metody prognozowania wartości szeregu czasowego zależy od wielu czynników [5]. Najważniejsze z nich to wielkość zbioru danych historycznych (czyli tzw. pamięć szeregu czasowego) i ścisła zależność tych danych, której miarą może być np. funkcja autokorelacji.

Gdy danych historycznych jest dużo i są ze sobą silnie skorelowane do rozwiązania problemu prognozy wartości szeregu czasowego można wykorzystać znane i sprawdzone modele, przedstawione m.in. w [5].

W interesującym nas przypadku prognozowania zapotrzebowania na usługi medyczne liczba danych historycznych jest niewielka (lata 1996 – 1999), a dodatkowo dane te często są niepełne. Trudno więc ustalić, jak silnie są ze sobą powiązane. W tej sytuacji prognoza zapotrzebowania na usługi medyczne obciążona będzie niedokładnością wynikającą z ograniczonej pewności danych źródłowych.

1.1. Wyznaczanie trendu metodą aproksymacji

W celu ustalenia trendu wykorzystano podejście polegające na realizacji dwóch podstawowych kroków prognozowania:

- dla danego zbioru wartości $\{y_1, y_2, \dots, y_n\}$ liczb rzeczywistych (prezentujących zapotrzebowanie na wybraną usługę medyczną w kolejnych latach), stosując założony mechanizm optymalizacji, można dobrać funkcję $F(x_i, p_0, p_1, \dots, p_n)$, która te wartości

najlepiej przybliża – ze względu na wydajność, najczęściej korzysta się z metody najmniejszych kwadratów,

- tak znaną postać funkcji można wykorzystać do prognozy, obliczając jej wartość dla argumentu $n+1$.

Z matematycznego punktu widzenia w kroku 1 trzeba rozwiązać zadanie aproksymacji nieznannej funkcji f , danej zbiorem dyskretnych wartości, za pomocą znanej funkcji F poprzez określenie parametrów w wyniku minimalizacji np. błędu średniokwadratowego. Ponieważ funkcję f można przybliżyć różnymi funkcjami F , należącymi do różnych klas funkcji, trzeba zastosować mechanizm doboru takiej funkcji F , która najlepiej przybliży funkcję f w sensie przyjętego kryterium jakości.

W kroku 2 do znalezienia prognozowanej wartości należy wykorzystać analityczne przedłużenie wybranej funkcji.

Należy zauważyć, że tak postawione zagadnienie cechuje się autokorekcją generowanych wyników. Prognozowanie bazuje na rzeczywistych danych, w związku z tym nie występuje zjawisko akumulacji błędów w kolejnych latach.

1.2. Dobór metody aproksymacji

Dla potrzeb rozwiązania zagadnienia przyjęto, że ogólna postać funkcji F jest zawężona do funkcji liniowej, logarytmicznej lub wykładniczej ze względu na charakter zjawiska, a optymalizacja prowadzi do obliczenia parametrów p_0, p_1, \dots, p_n funkcji F należących do rozpatrywanych grup, a następnie wybrania najlepszej w sensie przyjętej kategorii jakości aproksymacji. Parametry funkcji F są liczbami rzeczywistymi.

Typowe metody aproksymacji polegają na takim doborze parametrów p_0, p_1, \dots, p_n przyjętej funkcji, aby spełnić założone kryterium dokładności aproksymacji.

Przykładowe metody aproksymacji to:

- metoda wybranych punktów,
- metoda średnich,
- metoda sumowania bezwzględnych wartości,
- metoda najmniejszych kwadratów.

Metoda najmniejszych kwadratów daje jakościowo najlepsze wyniki i jest prosta w realizacji przy zastosowaniu aproksymacji w klasie wielomianów uogólnionych. Jest też dobrze udokumentowana w literaturze [1],[2].

Metoda ta opiera się na kryterium minimalizacji zdefiniowanym w następujący sposób:

$$\min_{p_0, p_1, \dots, p_k} \left(\sum_{i=1}^n [F(x_i, p_0, p_1, \dots, p_k) - y_i]^2 \right). \quad (1)$$

Doświadczenie osób pracujących w kasach chorych jak również osób badających zagadnienia trendów w Zakładzie Ubezpieczeń Społecznych wskazuje funkcje liniowe, logarytmiczne i wykładnicze jako najlepiej przybliżające charakter badanego zjawiska. Dodatkowym kryterium wyboru tych klas funkcji jest fakt „wygładzania” odchylek z wykonania poszczególnych usług medycznych, związanego ze zjawiskami o charakterze nieokresowym, takimi jak wyjątkowe epidemie, strajki w służbie zdrowia i inne.

1.3. Podsumowanie przeglądu metod, optymalizacja i jakość aproksymacji

Opracowany algorytm aproksymacji znajdujący się w [1] jest podobny dla wszystkich przedstawionych funkcji aproksymujących. Jest on szybki i niezbyt złożony, co pozwala na przeprowadzenie optymalizacji (oszacowania jakości) i doboru najlepiej pasującej funkcji do badanego zagadnienia według następującego schematu:

- przeprowadzenie aproksymacji średniokwadratowej dla założonych funkcji, obliczenie parametrów funkcji F ,
- obliczenie sumy kwadratów odchyleń wartości funkcji F_i , od wartości rzeczywistych, w punktach – latach – stanowiących podstawę aproksymacji,
- wybranie funkcji, której odpowiada najmniejsza suma obliczona w poprzednim kroku i wyliczenie jej wartość w punkcie prognozowanym, wykorzystując analityczne przedłużenie tej funkcji.

Powyższe kryterium oceny jakości jest proste i szybkie w implementacji numerycznej, a jednocześnie daje zadowalające efekty.

2. Przygotowanie danych źródłowych

Prawidłowe przygotowanie danych wykorzystywanych na etapie wyznaczania trendu ma kluczowe znaczenie dla otrzymanego wyniku. Pomijając przypadki, w których danych źródłowych brakuje, należy uwzględnić niespójność danych pochodzących z różnych systemów komputerowych służby zdrowia z lat poprzedzających działalność kas chorych.

2.1. Zależność danych od populacji

Zależność liczby wykonanych usług medycznych od populacji jest oczywista. Należy tutaj wspomnieć o zmianie granic województw oraz o dynamicznej zmianie liczby podopiecznych kasy chorych. W modelowym programie przyjęto jako wartości bezwzględne liczbę usług medycznych przypadających na 1000 mieszkańców badanego obszaru (najczęściej w skali województwa). Taki sposób przygotowania danych uniezależnia również

od błędu związanego z migracją ludności oraz umożliwia sparametryzowanie algorytmu dla przewidywanej populacji w okresie prognozowanym.

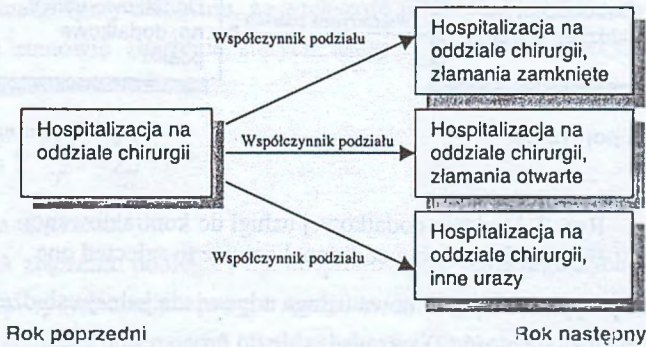
2.2. Translator usług medycznych

W poszczególnych latach opieka zdrowotna dysponowała różnym zapleczem technicznym do wykonywania usług medycznych. W praktyce oznacza to, że usługi medyczne kontraktowane przez Kasę np. w roku 1999 różnią się od usług medycznych, które będą kontraktowane w roku 2000 zarówno szczegółowością jak i jednostką miary, w jakiej wyrażona jest usługa.

Podczas przygotowywania danych wykorzystywanych do prognozowania należy zwrócić uwagę na wyjątkowe sytuacje złączenia lub podziału usług na przestrzeni lat.

Kasa może uszczegółowić kontraktowaną usługę w roku prognozowanym (w porównaniu z rokiem poprzednim) poprzez podział jednej usługi z roku poprzedniego na kilka usług w roku prognozowanym, patrz rys. 2.

Sytuacja taka wymaga wprowadzenia współczynnika podziału dla ilości usług – przykładowo łączna liczba osób hospitalizowanych na oddziale chirurgii ogólnej zostaje rozdzielona wg współczynników na trzy nowe usługi .



Rys. 2. Uszczegółowienie kontraktowanej usługi
 Fig. 2. Splitting service into better described one

Problem ten sprowadza się do zachowania następującej równości:

$$L_r = \sum_{i=1}^n \alpha_i M_{i,r+1}, \quad \text{gdzie} \quad \alpha_i = \frac{1}{A_i}, \quad (2)$$

r – rok,

L – współczynnik liczby zakontraktowanych jednostek danej usługi / liczbę ubezpieczonych,

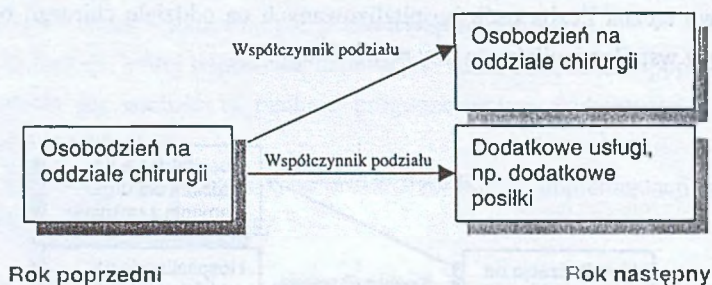
M – współczynnik liczby zakontraktowanych jednostek danej usługi / liczbę ubezpieczonych w roku następnym, stanowiący jedną z nowych usług wchodzących do podziału,

A – współczynnik trendu dla danej usługi.

Istnieje jednakże przypadek, w którym następuje zachwianie tej równości.

Kasa chce kontraktować dodatkowo do wybranej usługi inną, np. w celu podwyższenia standardu pobytu w szpitalu, patrz rys. 3. Powoduje to zwiększenie łącznej ilości świadczonych usług w sposób proporcjonalny do liczby usługi rozpatrywanej.

Oczywiście może zaistnieć sytuacja, w której Kasa rozbije kontraktowaną w roku poprzednim usługę na kilka nowych, przy czym suma współczynników podziału będzie mniejsza od jedności.



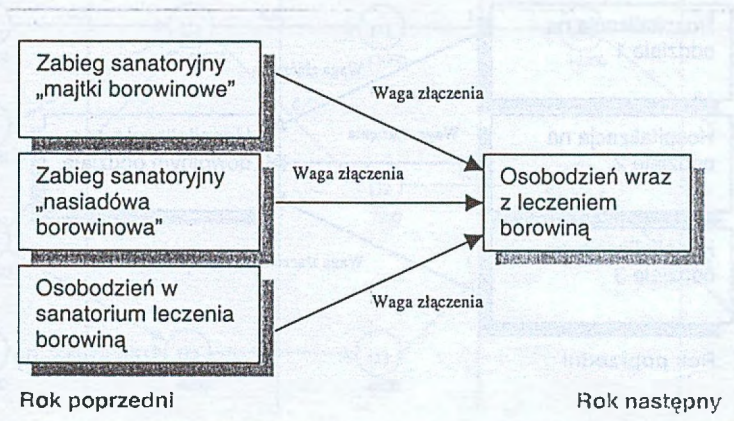
Rys. 3. Dodanie dodatkowej usługi do kontraktowanej
Fig. 3. Contracting additional service to selected one

W większości przypadków jedna nowa usługa odpowiada jednej usłudze z roku poprzedniego, a zatem zależność (2) sprowadza się do prostszej postaci:

$$L_r = \frac{M_{r+1}}{A}. \quad (3)$$

Na podstawie informacji otrzymanych od pracowników działu analiz i prognozowania kas chorych można wywnioskować, że taka sytuacja będzie miała miejsce w około 85% wszystkich przypadków.

Najbardziej rzadkim przypadkiem jest koncentracja usług.



Rys. 4. Grupowanie usług, przypadek 1

Fig. 4. Grouping services, case 1

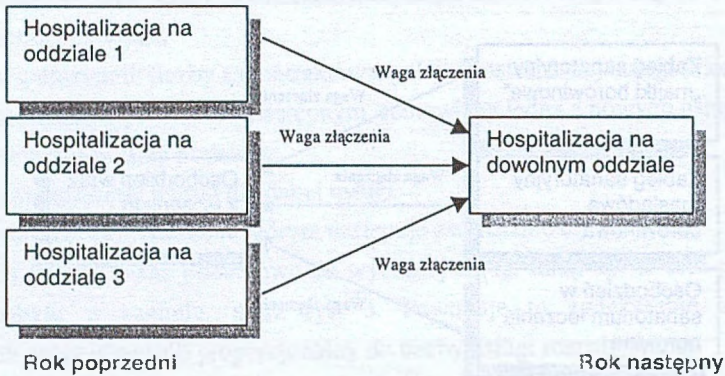
Kasa w następnym roku kontraktuje usługę stanowiącą złączenie usług z lat poprzednich. Wtedy poszczególne usługi z roku poprzedniego tworzą usługę prognozowaną z określonymi współczynnikami, zwanymi wagami złączenia, patrz rys. 4.

W takiej sytuacji (przy założeniu, że większość pacjentów korzysta z obu usług), nowa usługa powinna stanowić złączenie starych usług z wagami proporcjonalnymi, tak aby zachować zależność:

$$\sum_{i=1}^n L_{r,i} W_i = \frac{M_{r+1}}{A}, \sum_{i=1}^n W_i = 1 \quad (4)$$

W – waga złączenia.

Gdyby jednak złączeniu podlegały np. hospitalizacje z poszczególnych oddziałów w celu zakontraktowania ryczałtowego pobytu pacjenta w szpitalu (na dowolnym oddziale) należałoby złączyć usługi z wagami 1 tak, aby zakontraktowana liczba w roku kolejnym była sumą hospitalizacji w roku poprzednim, zmienioną o współczynnik trendu, patrz rys. 5.



Rys. 5. Grupowanie usług, przypadek 2

Fig. 5. Grouping services, case 2

Wtedy zachowana zostanie zależność:

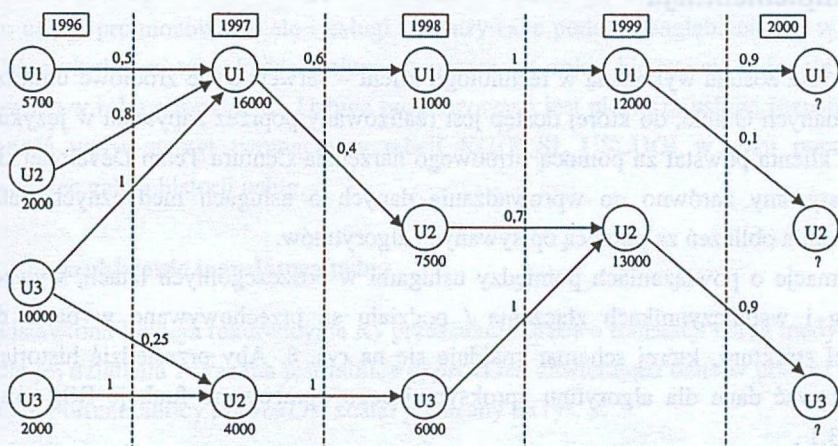
$$\sum_{i=1}^n L_{r,j} = \frac{M_{r+1}}{A}. \quad (5)$$

Poza wyżej wymienionymi zagadnieniami należało uwzględnić możliwość zmiany jednostki miary, w której wyrażono liczbę usług medycznych.

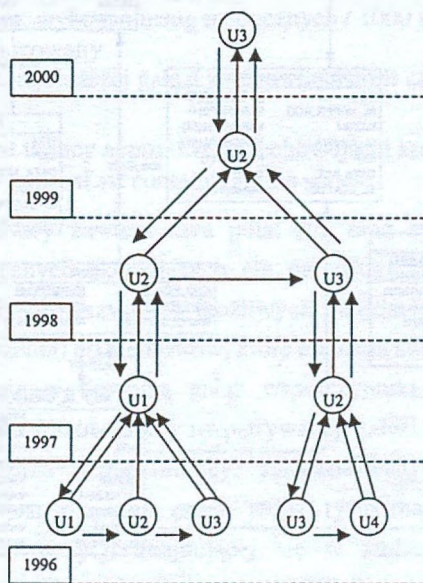
Przykładowo w 1999 roku pobyt pacjenta w szpitalu był opłacany ryczałtem (jednostką miary była hospitalizacja), niezależnie od ilości dni spędzonych w placówce. W związku z tym zdarzały się sytuacje przedwczesnych zwolnień ze szpitali. W roku 2000 kasa zdecydowała się kontraktować osobodni, zamiast hospitalizacji. Stąd potrzeba wprowadzenia współczynnika podziału / wagi złączenia, nawet w przypadku, gdy złączenie w poszczególnych latach jest złączeniem prostym. Wtedy współczynnik ten należy rozumieć jako współczynnik zmiany jednostki, w której wyrażono liczbę usług w poszczególnych latach.

Tak postawione zagadnienie translatora usług medycznych pozwala na zobrazowanie historii prognozowanych usług w postaci grafu. Przykładowy graf historii usług U1, U2 i U3 (prognoza na rok 2000) przedstawiono na rys. 6.

Szczególnej uwagi wymaga przypadek prognozowania usługi U3. Jak widać z grafu, z usługi U3 w roku 1996 do usługi U2 w roku 1999 można dotrzeć dwiema ścieżkami. Z punktu widzenia algorytmu przeszukiwania grafu translacji usługa U3 w roku 1996 jest dołączana do danych źródłowych dwukrotnie, za każdym razem z innym współczynnikiem złączenia (wynikającym z przejścia różnymi ścieżkami od U3 w roku 2000 do U3 w roku 1996). Tak przedstawiony graf jest rozwijany do postaci drzewa. Przykładowe drzewo usług oraz kolejność przeszukiwania przedstawiono na rys. 7.



Rys. 6. Przykładowy graf historii usług
Fig. 6. Sample services' history graph



Rys. 7. Drzewo usług medycznych dla prognozowanej usługi U3 w roku 2000
Fig. 7. History tree for U3 service, prognosis - year 2000

ww. tabelami. Należy zwrócić uwagę na fakt, że w tabeli KC_P_SL_USLUGI znajdują się nie tylko usługi prognozowane, ale i usługi rozpatrywane podczas zagłębiania się w historii usług. W związku z tym każda usługa tworząca w roku bieżącym staje się usługą rozpatrywaną w roku poprzednim. Usługa prognozowana jest pierwszą usługą rozpatrywaną. Nieobecność wpisu usługi tworzącej w tabeli KC_P_SL_USLUGI w roku poprzednim oznacza koniec gałęzi historii usług.

3.1. Przeszukiwanie translatora usług

Przedstawiona funkcja rekurencyjna *RS* przeszukuje drzewo translacji usług medycznych. W efekcie jej działania zwracana jest tablica *approxList*, zawierająca dane w postaci szeregu czasowego. Format tablicy *approxList* został pokazany na rys. 9.

Rok:	R_{n-1}	R_{n-2}	...	R_{i+1}	R_i
Współczynnik:	L_{n-1}	L_{n-2}	...	L_{i+1}	L_i

R – kolejne lata w historii usługi

L – szereg wsp. wykonań usług medycznych / 1000 mieszkańców

n – rok prognozowany

i – głębokość najdłuższej gałęzi w drzewie historii usług

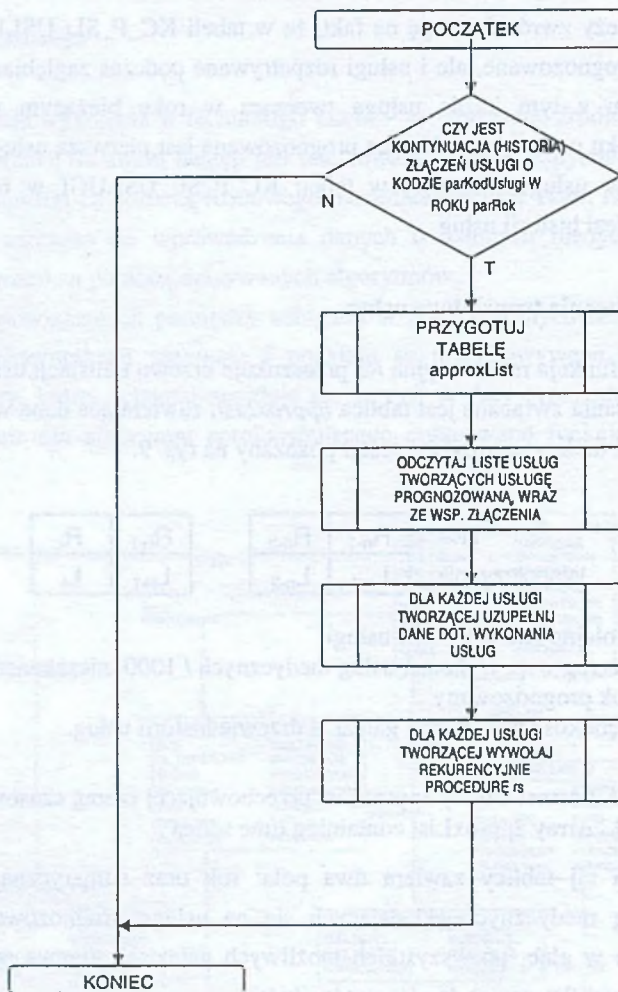
Rys. 9. Format tablicy *approxList* przechowującej szereg czasowy

Fig. 9. Array *approxList* containing time series

Każdy element tej tablicy zawiera dwa pola: rok oraz sumaryczną wartość liczby wykonanych usług medycznych składających się na usługę prognozowaną. Wartość ta wynika z przejścia w głąb, po wszystkich możliwych gałęziach drzewa prowadzących od usługi prognozowanej (korzenia) do elementów, które nie mają swojego poprzednika (liści).

Podczas zagłębiania się w dowolną gałąź współczynniki złączenia są wyznaczane narastająco. Na każdym poziomie (roku) rozpatrywanej gałęzi liczba wykonanych usług medycznych jest wyznaczana przez bieżący, zakumulowany współczynnik złączenia. Tak wyliczona wartość jest dzielona przez jedną tysięczną populacji województwa, a następnie dodawana do pozycji znajdującej się w tablicy *approxList*, w miejscu odpowiadającym bieżącemu rokowi. Podczas przejść sąsiednimi gałęziami wyliczone wartości są dodawane do tych, które znajdują się już w tablicy, w odpowiednich pozycjach odpowiadających rokowi. Wartości w poszczególnych polach L_i stanowią szereg czasowy i są przekazywane do odrębnej funkcji realizującej algorytm aproksymacji średniokwadratowej.

Schemat blokowy w postaci uogólnionej został przedstawiony na rys. 10.



Rys. 10. Funkcja RS

Fig. 10. RS function

Każde wywołanie procedury *RS* odpowiada zagłębieniu się w kolejny poziom historii usługi medycznej, począwszy od usługi prognozowanej, a skończywszy na usłudze, która nie ma żadnego poprzednika. Przeszukanie poziomu wiąże się z czterema krokami:

- Przygotowanie pozycji w tablicy *approxlist*:

W tym kroku wykonywane jest zerowanie pozycji wartości współczynnika wykonanych usług medycznych, jeśli jest to pierwsze przejście przez rozpatrywany rok. Dodatkowo niezbędne jest wpisanie rozpatrywanego roku. Lista lat określa wartości znajdujące się na osi odciętych podczas prognozowania, patrz rys. 1.

- odczytanie listy usług tworzących usługę rozpatrywaną:

Lista jest odczytywana z tabel bazy danych i przechowywana w tymczasowej tabeli *sCodeArray* (zmienna lokalna procedury *RS*, struktura na rys. 11) poprzez wykonanie zapytania:

```
select s.kod_uslugi, t.wspolczynnik, s.kod from  
KC_P_TRANS t, KC_P_SL_USLUGI_STARE s  
where t.kod_stary=s.kod and kod_nowy=:kod_uslugi_rozpatr',
```

gdzie tabela *t* przechowuje dane o translacji usług, tabela *s* przechowuje dane o usługach.

- Uzupełnienie tablicy współczynników wykonanych usług medycznych, przypadających na 1000 ubezpieczonych, dla wszystkich usług tworzących usługę rozpatrywaną:

Dla każdej usługi, która znajduje się w tablicy *sCodeArray*, na podstawie kodu wewnętrznego odczytywane są dane o liczbie wykonanych usług medycznych. Następnie liczby te (po wymnożeniu przez zakumulowany współczynnik złączenia usługi) są dodawane do pozycji odpowiadającej bieżącemu rokowi w tablicy *approxList*. Współczynnik zakumulowany oznacza współczynnik przekazany do procedury *RS* (czyli współczynnik wynikający z dotychczasowego przejścia po gałęzi, dla pierwszego wywołania równy 1), mnożony przez współczynnik złączenia rozpatrywanej usługi z usługą tworzącą.

- Wywołanie rekurencyjne procedury *RS* dla każdej usługi tworzącej usługę prognozowaną:

Parametrami procedury *RS* są: rok do rozpatrzenia, kod usługi rozpatrywanej, poziom zagłębienia, zakumulowany współczynnik złączenia. Jeżeli procedura nie natrafi na wpis w tabeli *KC_P_SL_USLUGI*, odpowiadający kodowi usługi rozpatrywanej (przekazanemu do procedury), zakończy akcję, kończąc w ten sposób przeszukiwanie gałęzi.

Po wyliczeniu wartości prognozowanej (korzystając z przedłużenia analitycznego funkcji) otrzymujemy współczynnik liczby usług przypadający na 1000 mieszkańców województwa. Pozwala to na przeprowadzenie symulacji zapotrzebowania na usługi w zależności od zakładanej populacji ubezpieczonych w roku prognozowanym. Wprowadzenie współczynnika liczby usług medycznych na liczbę mieszkańców było niezbędne z powodu różnic w liczebności populacji, której dotyczą dane z wykonań usług medycznych (m.in. z powodu zmian granic administracyjnych województw oraz dowolnością w przystępowaniu do kas chorych).

Kod usługi:	S_0	S_1	...	S_{j-1}	S_j
Współczynnik:	W_0	W_1		W_{j-1}	W_j
Kod wewnętrzny:	K_0	K_1		K_{j-1}	K_j

S – kody usług tworzących rozpatrywaną usługę

W – współczynniki złączenia usługi o kodzie S z rozpatrywaną usługą

K – wewnętrzne kody (klucze główne tabeli przechowującej usługi tworzące)

j – liczba usług tworzących rozpatrywaną usługę

Rys. 11. Tablica sCodeArray przechowująca kody usług tworzących

Fig. 11. sCodeArray containing codes of services preceding considered one

4. Zastosowanie

Przedstawiona metoda została praktycznie przetestowana w Śląskiej Regionalnej Kasie Chorych oraz dostarczona jako element systemu informatycznego do siedmiu innych regionalnych kas chorych.

LITERATURA

1. Majchrzak E., Mochnacki B.: Metody numeryczne – podstawy teoretyczne, aspekty praktyczne i algorytmy. Wydawnictwo Politechniki Śląskiej, 1996.
2. Fortuna Z., Macukow B., Wąsowski J.: Podręczniki Akademickie - Metody numeryczne. WNT, Warszawa, 1993.
3. Ralston A.: Wstęp do analizy numerycznej. PWN, Warszawa. 1983.
4. Jankowscy J. i M.: Przegląd metod numerycznych. Cz.1. WNT, Warszawa. 1981.
5. Box G.E.P., Jenkins G.M.: Analiza szeregów czasowych. Prognozowanie i sterowanie. PWN, 1983.
6. Serwis WWW Zakładu Informatyki i Badań Statystycznych Centrum Organizacji i Ekonomiki Ochrony Zdrowia: <http://www.coieoz.com.pl> .

Recenzent: Dr inż. Krystian Żymetka

Wpłynęło do Redakcji 25 marca 2003 r.

Abstract

The article covers various aspects of prognosis of total amount requisition for medical services. The method is based on trend prognosis using statistics and numerical algorithm software implementation.

Chapter one gives base problem description (see fig. 1), and describes base idea of trend prognosis using analytical function extension. Also considers optimal function selection that fits time series of discrete values using least squares method (see eq. (1)).

Chapter two covers selected problems and solutions on preparing data to comply with requirements of the method described in chapter one. As medical services are time and population dependent, chapter shows the way to allow medical service evolution (in the terms of time) providing so called "medical service translator" and explains typical situations such as splitting a service (fig.2 and equation (2)), additional service contracting (fig.3 and equation 3) or services grouping (fig. 4 and 5). Chapter two also shows graphic presentation of "medical services translator" as "services' history graph" (see fig.7), describes briefly graph into tree translation, and tree searching algorithm.

Chapter three covers various aspects of software implementation of considered problem. It shows service translator data structure as a part of SQL database (fig. 8) and describes recurrent algorithm (RS function, see fig. 10) searching through the translator for selected service's history (evolution). The second part of chapter describes all necessary steps to prepare data that fits least squares method requirements such as making data population independent. This part also shows necessary steps to achieve final result, which is a service prognosis for selected population. Also two most important internal data structures are briefly described (fig. 9 and 11).

Adres

Piotr CZEKALSKI: Politechnika Śląska, Instytut Informatyki, ul. Akademicka 16, 44-101 Gliwice, Polska, bella@boy.iinf.polsl.gliwice.pl