

PROBABILISTYCZNE MODELE ZJAWISK PRZESTRZENNYCH W ROLNICTWIE

Andrzej Marciniak

Katedra Podstaw Techniki, Uniwersytet Przyrodniczy w Lublinie

Streszczenie. Niepewność, zarówno stochastyczna jak i epistemiczna, obecna w modelach zjawisk czaso-przestrzennych w rolnictwie uzasadnia zastosowanie metod probabilistycznych predykcji, wyjaśnianiu i aproksymacji obiektów przestrzennych. Z metodologicznego, obliczeniowego i inferencyjnego punktu widzenia odpowiednią technologią modelowania są tu sieci bayesowskie traktowane jako systemy reprezentacji wiedzy. W takim ujęciu modelowanie sprowadza się do translacji wiedzy z języka naturalnego na formalny i wykonywalny język sieci bayesowskich. Logiczną spójność i efektywność takiego rozumienia procesu modelowania pokazano na przykładzie budowy modelu aproksymacji i predykcji plonu pszenicy.

Słowa kluczowe: systemy informacji przestrzennej (GIS), probabilistyczna predykcja i interpolacja obiektów przestrzennych, sieci bayesowskie

Wstęp

Już w latach 80-tych ubiegłego wieku rozważany był problem: czy obiekty biologiczne uczestniczące w rolniczych procesach produkcyjnych należy traktować indywidualnie czy też statystycznie. W latach 90-tych popularna stała się zasada „speaking soil/plant/animal approach” interpretowana jako potrzeba dostarczenia glebie, roślinom, zwierzętom „tego czego potrzebują, wtedy kiedy potrzebują i w niezbędnej ilości” (site specific treatment).

Mniej więcej w tym samym czasie udostępniona została technologia globalnego pozycjonowania obiektów oraz pasywnego i aktywnego skanowania powierzchni Ziemi w różnych wymiarach przestrzeni informacyjnej. Technologie te zostały systemowo zintegrowane pod postacią GIS.

Rolnictwo było jedną z pierwszych dziedzin zastosowań GIS. Jest tak dlatego, że obiekty uczestniczące w rolniczych procesach produkcyjnych charakteryzują się znaczną rozległością czasoprzestrzenną. Ilość informacji, ich precyzja w tym dokładna lokalizacja w czasie i przestrzeni uzyskiwana w systemach GIS sprawiły że zasada „speaking plant approach” stała się praktycznie wykonywalna pod postacią technologii rolnictwa precyzyjnego.

Pozyskiwanie danych empirycznych dla systemów GIS i ich przetwarzanie jest kosztowne i dlatego stosuje się tu podejście statystyczne obejmujące próbkowanie, interpolację i aproksymację [Moller, Waagepetersen 2003]. Klasyczne metody próbkowania i interpolacji opracowane jeszcze w epoce przed „GIS-owskiej” mają charakter deterministyczny. Próbkowanie badanych obiektów i ich charakterystyk czasowo-przestrzennych odbywa się

zazwyczaj we wcześniej zaplanowanych, a zatem zdeterminowanych punktach pomiarowych. Finalnym produktem ma tu być mapa mierzonej charakterystyki. Wymaga to interpolacji pomiędzy punktami pomiarowymi, w węzłach interpolacyjnych. W efekcie powstaje ciągła (być może z punktami nieciągłości) mapa mierzonej wielkości, na której nie ma rozróżnienia pomiędzy miejscami w których dokonano pomiaru a miejscami w których dokonano interpolacji. W miejscach takich mamy często do czynienia z praktycznie istotną niepewnością epistemiczną.

Probabilistyczne modelowanie obiektów i zjawisk przestrzennych

Celem referatu jest propozycja zastosowania probabilistycznego podejścia do problemu interpolacji i predykcji (aproksymacji) obiektów przestrzennych.

Przyrodnicze, zwłaszcza biologiczne obiekty przestrzenne charakteryzują się kognitywną nieokreślonością zarówno co do ich granic zewnętrznych jak i wewnętrznego zróżnicowania charakterystyk przestrzennych i jako takie mogą być modelem matematycznej abstrakcji jaką są pola losowe [Dalang, Dozzi, Russo 2002]. Formalne podejście do modelowania pól losowych oparte jest na teorii stochastycznych równań różniczkowych cząstkowych. Jednak ze względu na wymóg obliczalności i jednoczesnej dokładności modeli („lepsze przybliżone obliczenia na dokładnych modelach niż dokładne obliczenia na modelach przybliżonych”) proponuje się tu zastosowanie technologii sieci bayesowskich, [Gelman (i in.) 2003].

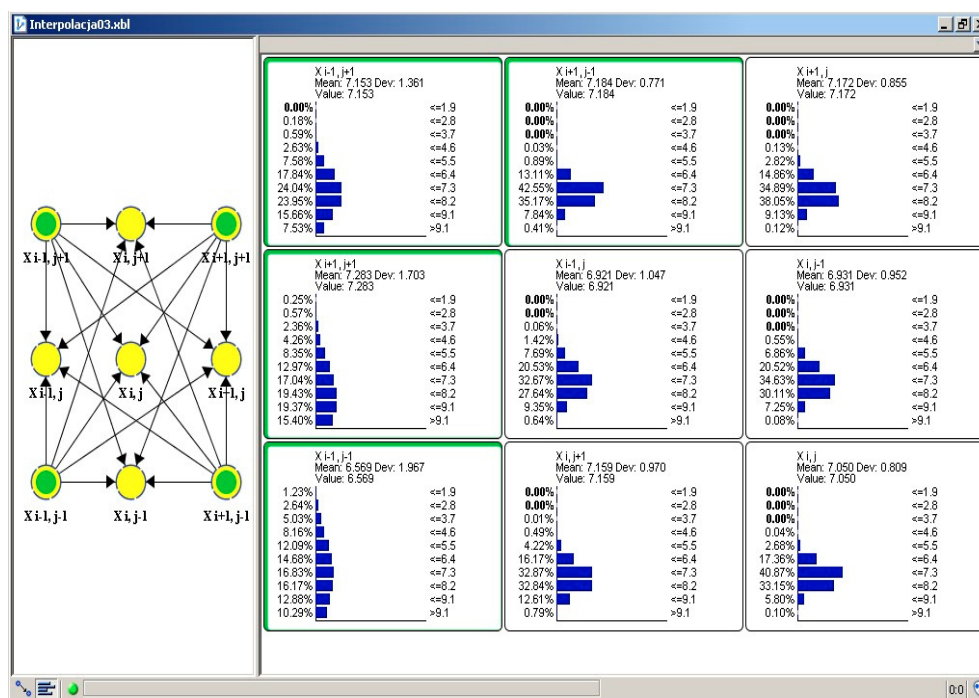
Język sieci bayesowskich charakteryzuje się dużą ekspresywnością i opartymi na logice mechanizmami wnioskowania, umożliwiającymi bogate i logicznie spójne wyjaśnianie zaobserwowanych faktów w kategoriach przyczynowo-skutkowych oraz przewidywanie logicznych konsekwencji różnych założeń modelowych. Ekspresywność języka i możliwość maszynowego uczenia na przykładach ma ważne znaczenie metodologiczne – budowanie modeli staje się procesem translacji wiedzy wyrażonej w języku naturalnym na język formalny i wykonywalny, procesem poszerzonym o automatyczne wykrywanie prawidłowości (zależności) w danych empirycznych i ich modelowanie.

Przykład

Wartości interpolowane nie mogą być traktowane jako obciążone tą samą dozą niepewności co wartości zmierzone. Niepewność ta nawet wzrasta jeżeli wartości w punktach pomiarowych też są niepewne. Dlatego należy założyć, że wartość mierzonej wielkości w punkcie interpolacji może być co najwyżej znana z dokładnością do rozkładu prawdopodobieństwa (być może o bardzo małej wariancji).

Na rysunku 1 przedstawiono sieć bayesowską reprezentującą problem interpolacji badanej wielkości w punkcie (i, j) w oparciu o znajomość rozkładu wartości pomiaru w sąsiednich punktach przy założeniu, że dokładna wartość w węzłach pomiarowych jest znana z dokładnością do rozkładu prawdopodobieństwa. Przykład wykonano w środowisku BayesiaLab - zaawansowanym systemie budowania modeli probabilistycznych w oparciu o procesy maszynowego uczenia na przykładach empirycznych i automatycznej inferencji [Bayesia (on-line) 2009].

W prawej części okna (rys. 1) przedstawiono rozkłady prawdopodobieństwa nad zbiorem wartości zmiennych w węzłach pomiarowych i w konsekwencji w węzle interpolowanym.



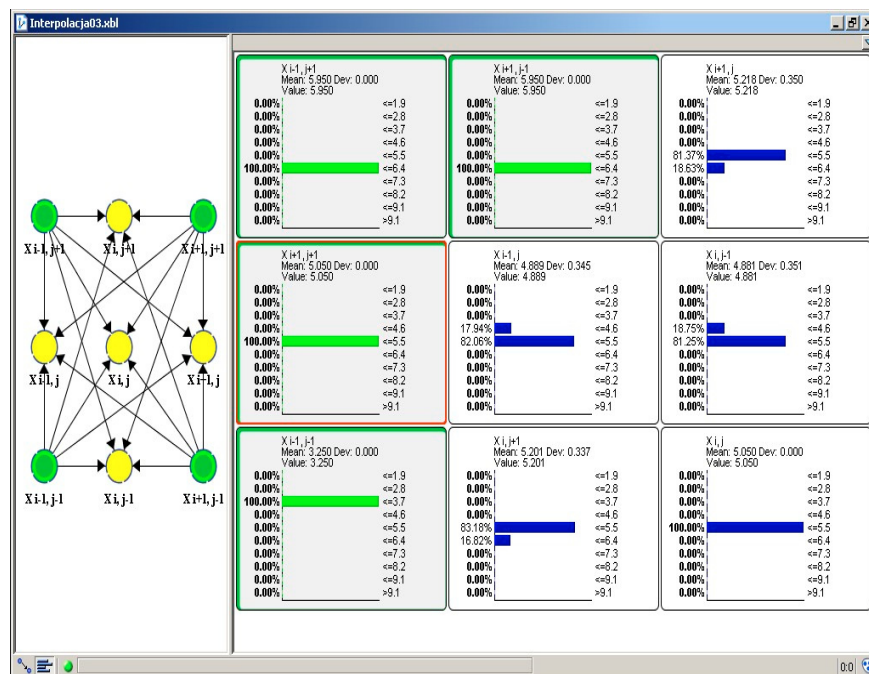
Źródło: opracowanie własne autora

Rys. 1. Wynik interpolacji plonu w węzle X_{ij} jeżeli jego wysokość w branych pod uwagę węzłach sąsiednich jest znana z dokładnością do rozkładu prawdopodobieństwa

Fig. 1. Result X_{ij} of yield interpolation in node (i, j) when yield in neighbouring nodes is known with accuracy to probability distribution

Na rysunku 2 przedstawiono sytuację, w której wartości zmierzone, znane są dokładnie a pomimo tego, zgodnie z przewidywaniami, wartość w węzle interpolacyjnym jest zmienną losową ale o znacznie mniejszej wariancji. Sieci przedstawione na rysunkach 1 i 2 można określić jako interpolatory bayesowskie.

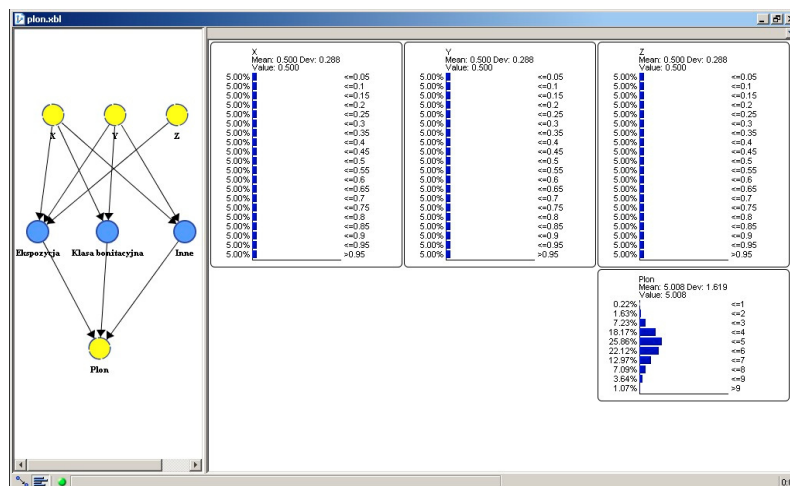
Odpowiednikiem klasycznego modelu aproksymacyjnego jest sieć bayesowska przedstawiona na rysunku 3. Sieć tę można określić jako predyktor bayesowski. Służy ona do predykcji (podstawowy rodzaj wnioskowania bayesowskiego) wartości aproksymowanej wielkości zgodnie ze znanymi relacjami przyczynowo-skutkowymi. Inny rodzaj wnioskowania bayesowskiego – od skutków do przyczyn, pokazano na rysunku 5. W tym przypadku zadajemy wysokość plonu i pytamy o lokalizację (z dokładnością do rozkładu prawdopodobieństwa) i inne atrybuty fragmentów pola o takim plonie.



Źródło: opracowanie własne autora

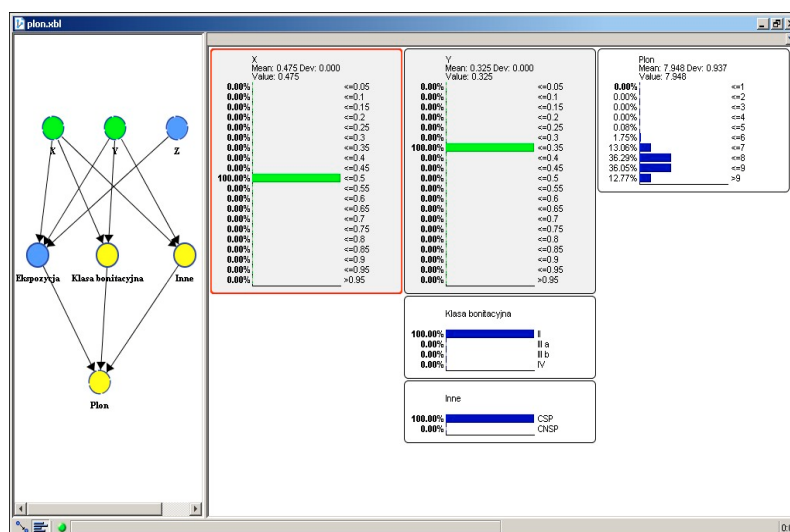
- Rys. 2. Wynik interpolacji plonu w węzle X_{ij} jeżeli jego wysokość w branych pod uwagę węzłach sąsiednich jest dokładnie znana
- Fig. 2. Result X_{ij} of yield interpolation in node (i, j) when yield in neighbouring nodes is exactly known

Wejściem tej sieci są współrzędne punktu a wyjściem w tym przypadku predykcja plonu. Sieć ta ma warstwę pośrednią którą są zmienne reprezentujące czynniki mające udokumentowany związek przyczynowo – skutkowy ze zmienną wynikową. Istotnym problemem jest wyznaczenie warunkowego rozkładu prawdopodobieństwa nad tymi zmiennymi. Problem rozwiązano w ten sposób, że wartości tych zmiennych dla danego pola i dla danego punktu na tym polu są wynikami klasyfikacji deterministycznego lub probabilistycznego klasyfikatora uczonego na przykładach. „Ekspozycja”, np. jest węzłem sieci ściśle deterministycznym natomiast „klasa bonitacji” jest węzłem reprezentującym klasyfikator bayesowski. Węzeł „inne” oznacza zmienną, której wartość „true” reprezentuje czynniki plonotwórcze natomiast wartość „false” czynniki niesprzyjające. Dodanie tego węzła ma związek z interpretacją prawdopodobieństwa bayesowskiego jako miary niepewności, której wartość aktualizowana jest poprzez fakty potwierdzające lub zaprzeczające. Warunkowy rozkład prawdopodobieństwa nad zmienną „Plon” otrzymywany jest bezpośrednio w procesie uczenia na danych empirycznych lub zadawany formułą reprezentującą w adekwatny sposób integrację informacji z węzłów poprzedzających.



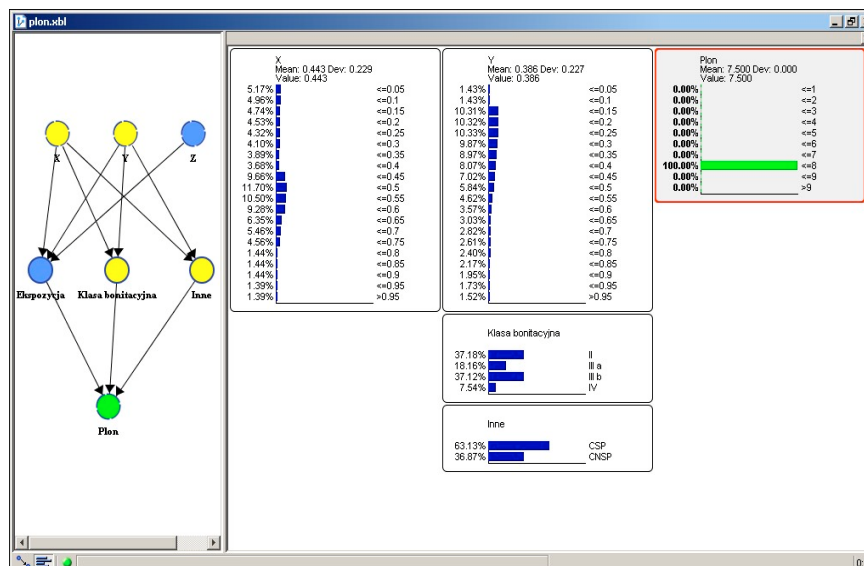
Źródło: opracowanie własne autora

Rys. 3. Wynik predykcji plonu z dokładnością do rozkładu prawdopodobieństwa w nieokreślonym miejscu pola
Fig. 3. Result of yield prediction with accuracy to probability distribution, without specification of site coordinates



Źródło: opracowanie własne autora

Rys. 4. Wynik predykcji plonu z dokładnością do rozkładu prawdopodobieństwa w określonym miejscu o współrzędnych x,y,z
Fig. 4. Result of yield prediction in site with exactly specified coordinates



Źródło: opracowanie własne autora

- Rys. 5. Odpowiedź na pytanie o miejsca na polu, w których przewidywany plon przyjmie wartość z przedziału 7-8 t·ha⁻¹
- Fig. 5. The answer to the question about the places on the field where predicted yield is in the range of 7- 8 t·ha⁻¹

Podsumowanie

Poszerzona analiza źródeł i natury niepewności w systemach informacji przestrzennej jest dobrym uzasadnieniem zastosowania probabilistycznego podejścia do problemu interpolacji i predykcji (aprosymacji) obiektów przestrzennych. Z metodologicznego, obliczeniowego i inferencyjnego punktu widzenia proponuje się tu wykorzystanie technologii bayesowskich sieci probabilistycznych. Skuteczność tej technologii pokazano na przykładzie bayesowskiego interpolatora i predyktora plonu.

Bibliografia

- Dalang R., Dozzi M., Russo F.** (ed.). 2002. Seminar on Stochastic Analysis, Random Fields and Applications IV. Centro Stefano Franscini. Ascona. Birkhäuser (Progress in Probability)
- Gelman A. i in.** 2003. Bayesian Data Analysis. Chapman & Hall/CRC.
- Moller J., Waagepetersen R. P.** 2003. A Statistical Inference and Simulation for Spatial Point Processes. Monographs on Statistics and Applied Probability. Chapman & Hall/CRC
- Bayesia (on-line) 2009. BayesiaLab - Bayesian network software. <http://www.bayesia.com/>

PROBABILISTIC MODELS OF SPATIAL PHENOMENA IN AGRICULTURE

Abstract. Uncertainty, both stochastic and epistemic, occurring in models of space-time phenomena in agriculture justifies application of probabilistic methods in predication, clarifying and approximation of spatial objects. From methodological, computational and inferential point of view, in this case proper modelling technologies include Bayesian networks treated as knowledge representation systems. From this perspective modelling comes down to translation of knowledge from natural language to formal and executable language of Bayesian networks. Logical coherence and effectiveness of this definition of modelling process is shown on the example of building a model of wheat crop approximation and prediction.

Key words: GIS, probabilistic prediction and approximation of spatial objects, Bayesian networks

Adres do korespondencji:

Marciniak Andrzej; e-mail: Andrzej.Marciniak@up.lublin.pl

Katedra Podstaw Techniki

Uniwersytet Przyrodniczy w Lublinie

ul. Doświadczalna 50A

20-280 Lublin