

Piotr Maksym, Andrzej W. Marciniak, Ryszard Kostecki
Katedra Podstaw Techniki
Akademia Rolnicza w Lublinie

ZASTOSOWANIE SIECI BAYESOWSKICH DO MODELOWANIA ROLNICZEGO PROCESU PRODUKCYJNEGO

Streszczenie

Zastosowanie sieci bayesowskich do modelowania rolniczego procesu produkcyjnego oraz wspomagania decyzji odnośnie wyboru działań produkcyjnych przedstawiono z perspektywy budowania komputerowych systemów reprezentacji wiedzy. Metodykę budowania modeli sprowadzono do projektowania odpowiednich modułów rozpoznawania i predykcji i łączenia ich według określonych reguł. Zaproponowaną metodę omówiono na przykładzie modelowania produkcji pszenicy ozimej.

Słowa kluczowe: systemy reprezentacji wiedzy rolniczej, modelowanie procesów produkcyjnych, sieci bayesowskie, reguły modularyzacji i kompozycji

Wprowadzenie

Reprezentacja wiedzy w postaci modeli komputerowych, wbudowanych w informatyczną infrastrukturę zarządzania procesami jest jednym z paradygmatów systemów produkcji opartych na wiedzy i informacji. Technologia sieci bayesowskich pozwala na szybkie, niemal zautomatyzowane przekształcanie wiedzy ekspertów, wyrażonej werbalnie lub zapisanej w języku naturalnym, w postaci modeli komputerowych. Programy wspomagające tę technologię oparte są na językach programowania wizualnego i przejrzystych schematach konceptualizacyjnych, co w połączeniu z automatycznymi generatorami kodów źródłowych sprawia, że budowanie modeli komputerowych przestaje być domeną programistów. Pisanie kodu przestaje być najtrudniejszym etapem modelowania. Z kolei, jakość modeli nadal zależy od jakości konceptualizacji problematyki modelowanej dziedziny, a ta wymaga głębokiej wiedzy nie tylko teoretycznej ale również praktycznej i dlatego modele komputerowe konkretnej dziedziny przedmiotowej powinien tworzyć specjalista w tej dziedzinie. Stwierdzenie to dotyczy złożonych dziedzin empirycznych, takich jak inżynieria rolnicza.

Celem przedstawionej pracy jest próba pokazania sposobu (metody) przejścia od nieformalnej konceptualizacji rolniczego procesu produkcji do modelu w postaci sieci bayesowskiej. Ten rodzaj modeli, oprócz zalet wyżej wspomnianych, dostarcza przejrzystej konceptualizacji niepewności i ryzyka, nieodłącznie tkwiących w produkcji rolniczej, oraz ich miar ilościowych w postaci rozkładów prawdopodobieństwa. Bayesowski mechanizm rozumowania probabilistycznego, bardzo prosty w swojej istocie i dzięki temu uniwersalny, ma zadziwiające bogactwo praktycznych interpretacji. Jego wadą jest złożoność obliczeniowa ale opracowane w ostatnich latach algorytmy wnioskowania przybliżonego sprawiają, że nawet duże sieci są obliczeniowo efektywne.

Konceptualizacja przedmiotu modelowania

Rolnicze procesy produkcyjne należą do tej kategorii bytów, które widzimy i opisujemy i o których rozumiemy tak jak je poznajemy. Stąd, jako punkt wyjścia do zbudowania uporządkowanej konceptualizacji działań składających się na proces produkcyjny przyjęto schemat elementarnego cyklu kognitywnego: *<rozpoznanie problemu>* → *<wybór działania>* → *<wykonanie>*↵. Poszczególne fazy tego cyklu należy zinterpretować odpowiednio do specyfiki modelowanego procesu. W typowym procesie produkcji roślinnej można wyróżnić trzy kategorie działań:

- działania inicjujące proces – wybór stanowiska pod uprawę, uprawa poźniwna i przedsiewna oraz siew,
- działania kontroli i ochrony procesu – kontrola stanu uprawy pod kątem typowych możliwych zagrożeń i wynikające z niej działania uprzedzające, interwencyjne i korekcyjne,
- działania kończące proces – zbiór plonu, w tym ewentualne działania przygotowujące zbiór i kończące cykl uprawowy.

Wszystkie te działania są temporalnie uporządkowane i skutki, główne oraz uboczne, działań wcześniejszych mają wpływ na wybór działań późniejszych w zakresie jednego, a nawet wielu cykli produkcyjnych. Jeżeli do tego dodamy aktywny udział czynników pogodowych i biologicznych, negatywnych lub sprzyjających, to mamy do czynienia z problemem, w którym wybór działań zależy nie tylko od wyniku rozpoznania sytuacji poprzedzającej, ale również od predykcji ich skutków - chwilowych i opóźnionych, krótko i długo terminowych. Trafność predykcji skutków działań technologicznych wymaga dokładnych i sprzężonych ze sobą prognoz biologicznych i agrometeorologicznych. Prognozy biologiczne dotyczą zdarzeń fenologicznych w cyklu rozwojowym uprawianej rośliny i w sprzężonych z nim cyklach rozwojowych agrofagów.

Reprezentacja rolniczego procesu produkcyjnego w języku sieci bayesowskich

Przedstawiony wyżej schemat konceptualizacyjny implikuje określoną modularyzację sieci bayesowskich i topologię poszczególnych modułów. Modularyzacja sieci jest tu istotna co najmniej z trzech powodów:

- moduły mają sztywną topologię odwzorowującą istotną wiedzę przyczynowo skutkową,
- algorytmy uczenia maszynowego na poziomie estymacji warunkowych rozkładów prawdopodobieństwa są znacznie efektywniejsze niż algorytmy jednoczesnego uczenia topologii i rozkładów prawdopodobieństwa, a pobranie danych do uczenia maszynowego nie wymaga przeszukiwania długich ścieżek w bazach danych,
- moduły są składnikami do wielokrotnego użycia, a mając odpowiednio bogate repozytorium modułów można proces budowy modelu sprowadzić do automatyzowalnej operacji wyboru i składania modułów.

Niezbędne są dwa rodzaje modułów – diagnostyczne i predykcyjne. Moduły diagnostyczne służą do rozpoznawania ogólnie rozumianego stanu plantacji, np. rodzaju, zakresu i poziomu zagrożeń na podstawie samoistnie pojawiających się i obserwowalnych ich objawów, a nawet zanim one wystąpią. Sieć odpowiadająca takim modułom zawiera dwa rodzaje węzłów. Węzły wstępujące, reprezentujące przyczyny, np. zagrożenia (bezpośrednio nie obserwowalne) oraz węzły zstępujące, reprezentujące objawy. Rozpoznanie (diagnoza) jest tu wynikiem wnioskowania polegającego na wyjaśnianiu obserwowalnych skutków w kategoriach nieobserwowalnych przyczyn. Specyfikujemy objawy i otrzymujemy diagnozy odnośnie przyczyn z dokładnością do rozkładu prawdopodobieństwa nad możliwymi przyczynami. Moduły diagnostyczne mogą być w zależności od potrzeb dalej rozbudowywane. Można w nich umieszczać dłuższe łańcuchy przyczynowo skutkowe, jeżeli objawy nie są bezpośrednimi skutkami diagnozowanych przyczyn. Poprzez dodawanie węzłów reprezentujących symptomy, których obserwacja wymaga dodatkowych pomiarów oraz szeregowanie symptomów według kosztów pomiarów, diagnostyczna sieć bayesowska odwzorowuje również pełny cykl rozumowania hipotetyczno-dedukcyjnego. Dla dedukcyjnego domknięcia cyklu wystarczy zafiksować najbardziej prawdopodobną hipotezę jako „twardy fakt” i znaleźć wśród niesprawdzonych jeszcze symptomów taki, który najlepiej lub najtaniej weryfikuje daną hipotezę.

Drugi z wyżej wspomnianych rodzajów modułów potrzebny jest do predykcji procesów w zależności od podjętych decyzji przed wykonaniem działań z nich wyni-

kających. Do tego potrzebne są sieci nazywane dynamicznymi. Sieć dynamiczna opisuje ewolucję łącznego rozkładu prawdopodobieństwa w czasie. Aktualnie, sieci dynamiczne tworzone są z dwóch sieci o takiej samej topologii (oprócz tych węzłów sieci, które nie ewoluują w czasie), ale o rozkładach prawdopodobieństwa odpowiadających dwóm przekrojom czasowym modelowanego procesu, odległym o stały krok czasu [Murphy 2002]. Przechodzenie przez n kolejnych kroków czasu odbywa się za pomocą pętli iteracyjnej. Zmianę rozkładów prawdopodobieństwa pomiędzy dowolnymi sąsiednimi przekrojami w czasie można w sposób jawny uzależnić od numeru kroku czasu.

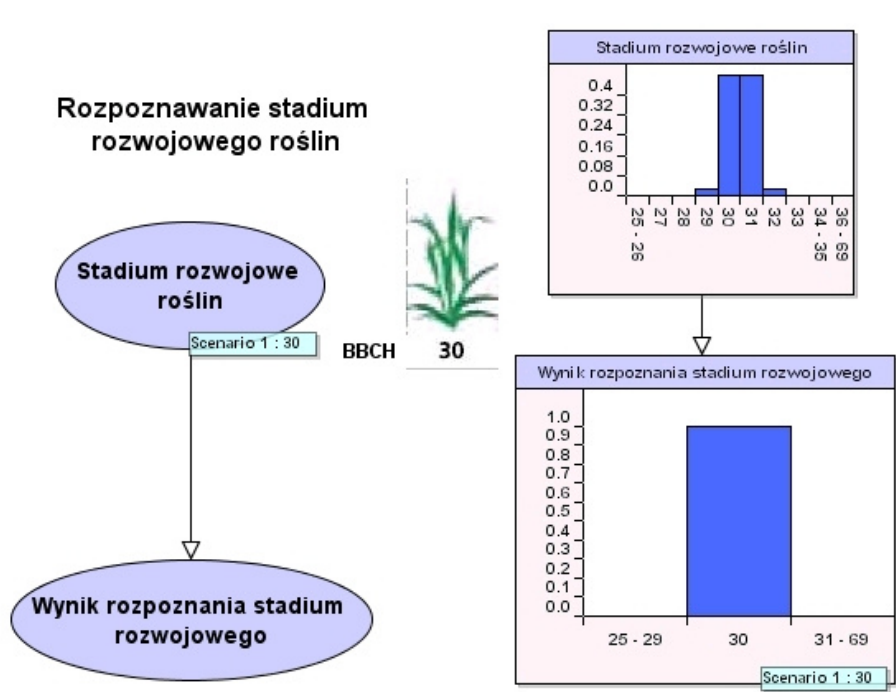
Chociaż sieć bayesowska nie jest modelem optymalizacyjnym w tradycyjnym znaczeniu, to może być użyta do optymalizacji decyzji. Do tego celu może być wykorzystany fakt, że formalnie, dowolnym węzłom sieci można przypisać pożądaną wartość („twardy fakt”) lub pożądaną rozkład prawdopodobieństwa („miękki fakt”) i obliczyć aposterioryczny rozkład prawdopodobieństwa nad wartościami pozostałych węzłów. Ten mechanizm nazywa się rozumowaniem „*what if*”. Jeżeli w sieci wyodrębnimy węzły reprezentujące cel optymalizacji (może ich być kilka), ograniczenia oraz zmienne decyzyjne, to poszukiwanie optymalnego rozwiązania problemu, można sprowadzić do przeszukiwania zakresów wartości zmiennych decyzyjnych aż spełnione zostanie kryterium optymalizacji. Ponieważ węzły w sieci bayesowskiej reprezentują zmienne losowe, to jako kryterium optymalizacji przyjmujemy maksymalizację (minimalizację) pewnych statystyk związanych z rozkładem prawdopodobieństwa zmiennej reprezentującej cel optymalizacji.

Przykład

Jako dziedzinę modelowania wybrano proces produkcji pszenicy ozimej. Stosowany tu model pojęciowy i schematy rozumowania mogą być użyte do modelowania również innych upraw.

Wiedzę przedmiotową z zakresu produkcji pszenicy ozimej zaczerpnięto przede wszystkim z zaleceń Instytutu Ochrony Roślin [www.ior.poznan.pl].

Na rysunku 1 pokazano elementarny moduł diagnostyczny. Celem rozpoznawania jest tu faza - a dokładniej stadium rozwojowe pszenicy wyrażone w jednostkach rozszerzonej skali fenologicznej BBCH [Meier U. (ed.) 2001]. Skala fenologiczna jest rodzajem zegara biologicznego uprawianej rośliny, z którym zsynchronizowane są zegary biologiczne agrofagów. Stąd, od poprawnego rozpoznania stadium rozwojowego roślin zależy trafność lokalizacji w czasie wszelkich działań plonochronnych.



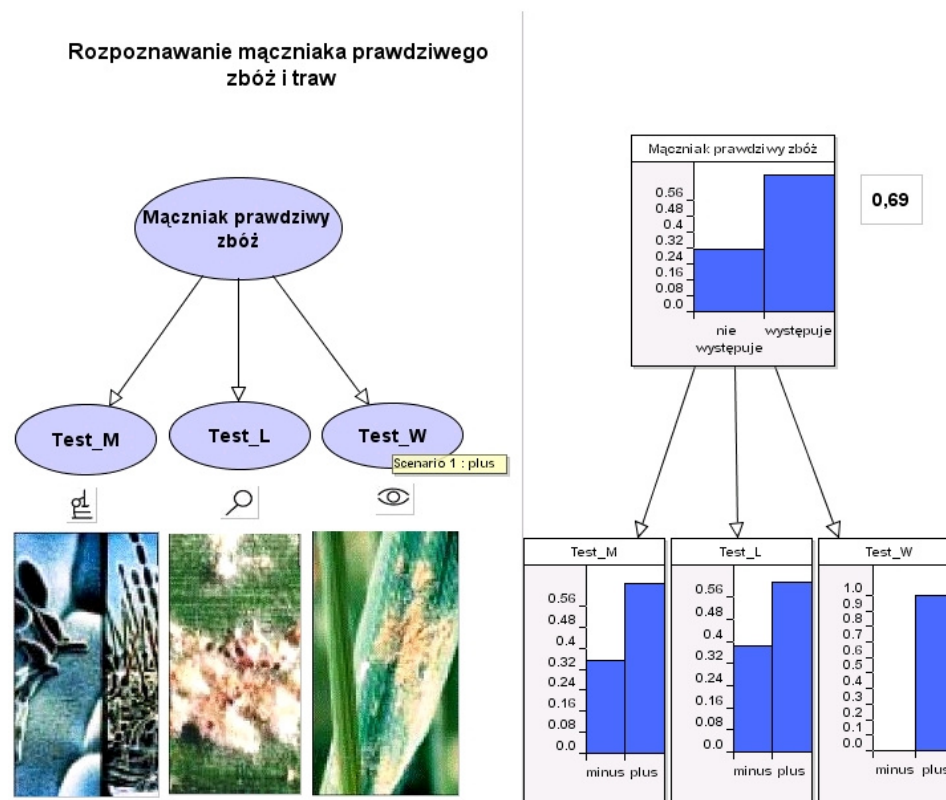
Rys. 1. Rozpoznawanie stadium rozwojowe roślin

Fig. 1. Plants developmental stage recognition

Sieć pokazana na rys. 1 dotyczy rozpoznawania stadium rozwojowe poprzez analizę pokroju roślin i składa się z dwóch węzłów. Węzeł wstępujący reprezentuje rzeczywiste stadium rozwojowe a węzeł zstępujący – wynik rozpoznawania. Pokazane rozkłady prawdopodobieństwa dotyczą sytuacji, w której jeżeli wynik rozpoznawania jest jednoznacznie BBCH 30 to rzeczywiste stadium rozwojowe jest z jednakowo dużym prawdopodobieństwem BBCH 30 lub BBCH 31 i pomijalnie małym prawdopodobieństwem BBCH 28 lub BBCH 32. Ten rozkład jest miarą niepewności co do rzeczywistego stadium rozwojowe. Łatwo zauważyć, że taki model może być również użyty do testowania układu rozpoznającego, niezależnie czy jest to „oko → mózg” ludzkiego eksperta czy komputerowy system rozpoznawania obrazów.

Śledząc wskazania zegara fenologicznego, wiemy kiedy należy przeprowadzić lustrację uprawy pod kątem występowania specyficznych i groźnych dla uprawianej rośliny agrofagów. Do podjęcia decyzji odnośnie zabiegu ochronnego wymagane jest rozpoznania agrofaga oraz zakresu i stopnia porażenia uprawy.

Elementarny model problemu rozpoznawania agrofaga jest analogiczny jak na rys. 1. Jego konkretyzację pokazano na rysunku 2 na przykładzie rozpoznawania mączniaka prawdziwego zbóż i traw. Węzeł wstępujący sieci reprezentuje obiekt rozpoznawania, tj. wystąpienie lub nie konkretnego patogena, a węzeł zstępujący wynik rozpoznawania (wynik określonego testu na występowanie).



Rys. 2. Rozpoznawanie mączniaka prawdziwego

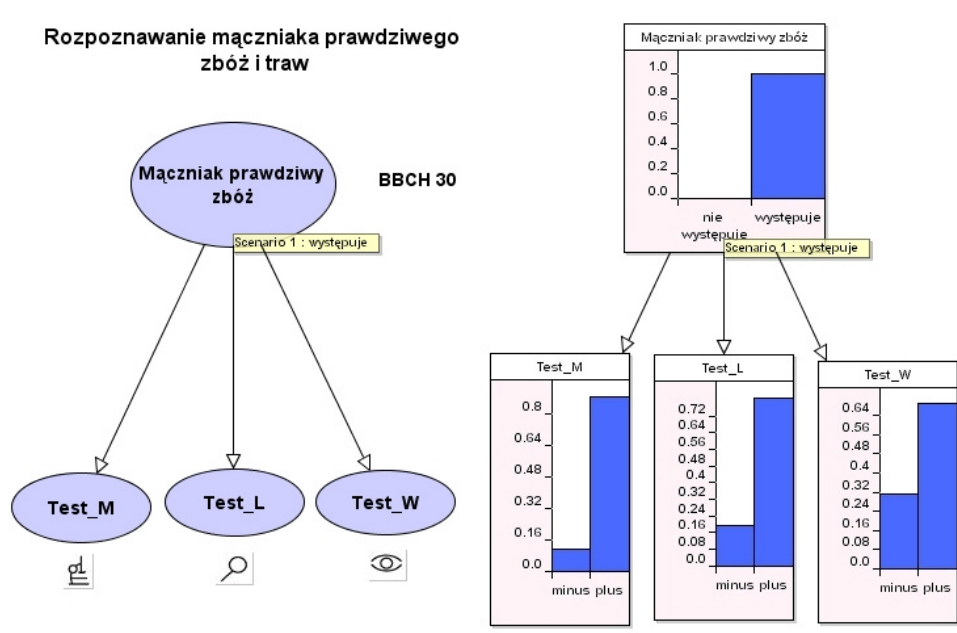
Fig. 2. Wheat powdery mildew (*Blumeria graminis f.sp. tritici*) recognition

Aktualnie stosowaną metodą rozpoznawania jest analiza wzrokowa obrazu objawów na liściach i porównywanie ich z obrazami wzorcowymi. Ażeby zmniejszyć ryzyko błędów rozpoznawania, oprócz obserwacji zmian patologicznych gołym okiem, zaleca się obserwacje pod lupą lub mikroskopem. Na ogół wystarcza to do uniknięcia pomyłek w przypadku występowania zmian wywoływanych przez czynniki abiotyczne [Syngenta 2005]. Dlatego właśnie sieć pokazana na rysunku 2¹⁾

¹⁾ Użyte tu wzorcowe obrazy objawów zostały zaczerpnięte z atlasu chorób zbóż [Syngenta 2005].

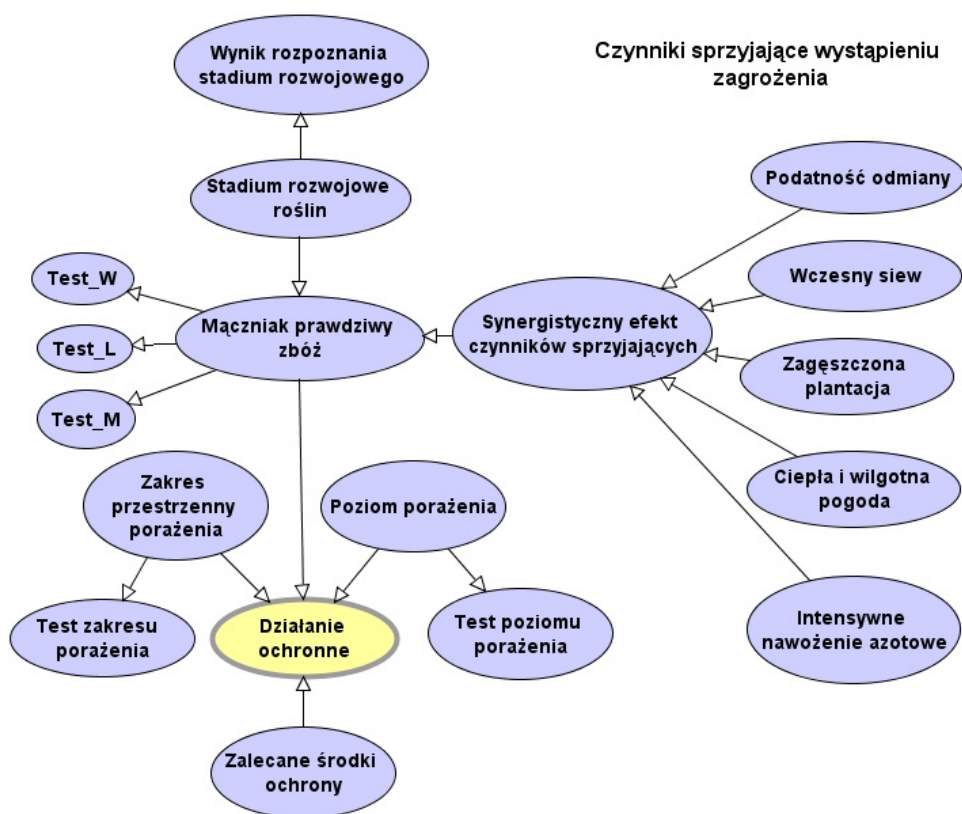
ma trzy węzły zstępujące, oznaczone etykietami Test W, Test L i Test M, reprezentującymi odpowiednio: obserwację nieuzbrojonym okiem, obserwację pod lupą i pod mikroskopem. Rozkłady prawdopodobieństwa pokazane na rys. 2 odnoszą się do sytuacji, gdy rozpoznanie wykonano tylko gołym okiem i jego wynik uznano jako jednoznacznie pozytywny. Wówczas prawdopodobieństwo tego, że patogen rzeczywiście występuje jest równe 0,69. Pozostałe testy dają odpowiednio wynik 0,80 i 0,90.

Jak już wcześniej wspomniano, model taki może być użyty do oceny jakości układu rozpoznającego. Każdy z pokazanych tu trzech testów ma inną wrażliwość i specyficzność. Miarą wrażliwości testu jest warunkowe prawdopodobieństwo pozytywnego wyniku pod warunkiem, że rozpoznawany patogen rzeczywiście występuje. Stąd, ażeby obliczyć wrażliwość testów należy zadać występowanie patogena jako „twardy fakt”. Pokazano to na rysunku 3. Jak należało oczekiwać, największą wrażliwość ma test „obserwacja pod mikroskopem”. Specyficzność testu definiujemy jako warunkowe prawdopodobieństwo tego, że wynik testu będzie negatywny, pod warunkiem, że patogen nie występuje. Specyficzność obliczamy analogicznie jak wrażliwość, zadając niewystępowanie patogena jako „twardy fakt”.



Rys. 3. Wrażliwość testów
Fig. 3. Tests sensibility

Sposób połączenia modułów diagnostycznych z pozostałymi modułami sieci wynika z odpowiedzi na pytanie o to, jakie węzły sieci określają rozkład prawdopodobieństwa wystąpienia diagnozowanej przyczyny lub stanu oraz jakich węzłów warunkowe prawdopodobieństwo zależy od występowania zagrożenia. Odpowiedź na drugie pytanie jest oczywista – są to węzły decyzyjne. Odpowiedź na pierwsze pytanie zależy od tego, co jest przedmiotem diagnozowania. W przypadku diagnozowania zagrożeń powodowanych przez czynniki biologiczne będzie to węzeł reprezentujący stadium rozwojowe uprawianej rośliny oraz węzły reprezentujące czynniki sprzyjające i być może czynniki powstrzymujące wystąpienie zagrożenia. Taką przykładową konstrukcją pokazano na rysunku 4.



Rys. 4. Przykład łączenia modułów
 Fig. 4. Example of module assembling

Podsumowanie

Problem modelowania rolniczego procesu produkcyjnego za pomocą sieci bayesowskich przedstawiono z perspektywy projektowania komputerowych systemów reprezentacji wiedzy. Konceptualizację i wynikające z niej mechanizmy rozumowania odnośnie złożonego procesu produkcji oparto na uniwersalnym schemacie pojęciowym: - <rozpoznanie problemu> → <wybór działania> → <wykonanie> ↓. Ten schemat konceptualizacyjny implikuje określoną modularyzację sieci bayesowskich i topologię modułów. Wyodrębniono dwa rodzaje modułów: diagnostyczne i prognostyczne. Odpowiadają one dwóm znanym mechanizmom rozumowania probabilistycznego w sieciach bayesowskich – od skutków do przyczyn i od przyczyn do skutków.

Bibliografia

Meier U. (ed.) 2001. Growth stages of mono-and dicotyledonous plants. BBCH Monograph. 2nd Edition, 2001.

Murphy K. P. 2002. Dynamic Bayesian Networks. <http://www.ai.mit.edu/~murphyk>

Instytut Ochrony Roślin. <http://www.ior.poznan.pl/>

Syngenta Crop Protection 2005. Atlas chorób zbóż powodowanych przez grzyby. Rozpoznawanie - Ocena – Zwalczenie. <http://www.syngenta.pl>

APPLICATION OF BAYESIAN NETWORKS IN MODELING OF AGRICULTURAL PRODUCTION PROCESS

Summary

Problem of agricultural production modeling with Bayesian Networks is discussed from the perspective of building knowledge representation systems in agricultural domain. Universal conceptualization schema of acting in changing and uncertain environments as cognitive cycles, implies the need of two elementary modules of BN, namely the recognition (diagnostic) and prediction module. Building BN

model of agricultural production process can be decomposed into the design of elementary modules repository and then appropriate composing entire models with modules. Conceptualization and design of diagnostic and predictive modules, and then their linkage rules was shown on the examples concerned winter wheat production.

Key words: agricultural knowledge representation systems, production process modeling, Bayesian networks, modularization and composition rules