

MODELOWANIE PROBLEMÓW DECYZYJNYCH W INTEGROWANYM SYSTEMIE PRODUKCJI ROLNICZEJ

Hanna Hołaj

Rolniczy Zakład Doświadczalny „Jastków” Sp. z o. o.

Andrzej Kusz, Piotr Maksym, Andrzej W. Marciniak

Katedra Podstaw Techniki, Uniwersytet Przyrodniczy w Lublinie

Streszczenie. W artykule przedstawiono metodę modelowania problemu oceny zagrożenia w rolniczym procesie produkcyjnym i doboru konkretnego środka ochrony oraz wielkości adekwatnej dawki w integrowanym rolniczym procesie produkcyjnym. Podstawą decyzji jest diagnoza generatywna oparta o rozkład prawdopodobieństwa określony nad zbiorem możliwych decyzji. Pokazano w jaki sposób poprzez integrację niejednorodnych i niepewnych źródeł informacji można zmniejszyć niepewność w procesie podejmowania decyzji. Przedstawioną koncepcję budowy modelu oparto na technologii sieci bayesowskich.

Słowa kluczowe: produkcja rolnicza, modelowanie zagrożeń, integrowanie źródeł informacji, wspomaganie procesów decyzyjnych, sieci bayesowskie

Wstęp

Jednym z elementów paradygmatu integrowanej produkcji rolniczej jest odejście od binarnych, progowych decyzji odnośnie działań ochronnych w rolniczym procesie produkcyjnym. Dotychczas działania takie są wykonywane w oparciu o diagnozy dyskryminatywne, które matematycznie odpowiadają minimalizacji błędu związanego z niewykryciem zagrożenia w sytuacji gdy ono rzeczywiście wystąpiło przy określonym a priori, akceptowalnym poziomie błędu drugiego rodzaju (błędne rozpoznanie w sytuacji gdy zagrożenie nie jest istotne). W wyniku otrzymuje się progową funkcję dyskryminacyjną która obowiązuje tak długo jak długo nie zmieniają się konsekwencje błędów i rozkłady prawdopodobieństwa. W przypadku przekroczeniu progu szkodliwości zalecane jest wykonanie działania interwencyjnego, o z góry określonym natężeniu, nie zawsze dostosowanym do rzeczywistego stopnia zagrożenia.

Alternatywnym podejściem do problemu oceny zagrożenia jest odejście od funkcji dyskryminacyjnej na rzecz diagnozy generatywnej, opartej na wyznaczeniu rozkładu prawdopodobieństwa nad zbiorem możliwych decyzji dotyczących wyboru konkretnego środka ochrony oraz jego dawki. Wymaga to bardziej precyzyjnego określenia rozpoznawanego zagrożenia i jest możliwe poprzez integrację informacji pochodzących z heterogenicznych źródeł.

Istotnym elementem proponowanego podejścia do problemu ochrony w integrowanym systemie produkcji jest stosowanie działań prewencyjnych. Działania takie stosuje się bez testu, o ile prawdopodobieństwo wystąpienia zagrożenia w konkretnej uprawie jest znaczące, ale jeszcze nie wystąpiło. Do działań prewencyjnych są przypisane środki oznaczone jako stosowne do tych działań. Środków tych nie stosuje się w sytuacji gdy zagrożenie już wystąpiło [Korbas, Mrówczyński 2009].

Cel pracy

W pracy przedstawiono sposób konceptualizacji w języku sieci bayesowskich problemu podejmowania decyzji zgodnych z paradygmatem integrowanej produkcji rolniczej. Istotnym elementem opracowanej metodyki jest integracja niejednorodnych i niepewnych źródeł informacji oraz pokazanie na krzywych ROC (ang. Receiver Operating Characteristic) [Fawcett 2004] efektu zmniejszania tej niepewności w przypadku posiadania dokładnej informacji lub poprzez dodanie nowego źródła informacji.

Konceptualizacja problemu oceny zagrożeń

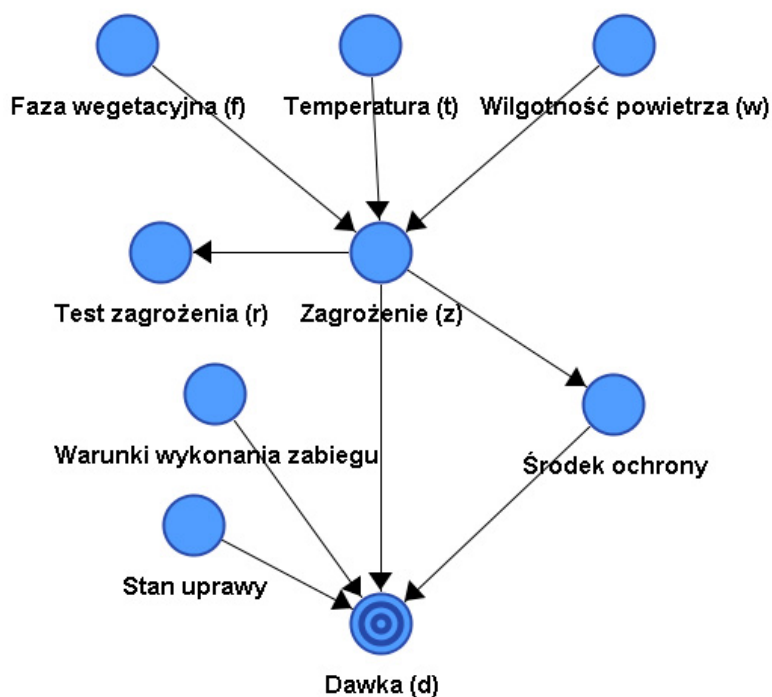
Konceptualizacja problemu oceny zagrożeń i wyboru odpowiedniego środka ochrony o odpowiednie dawce jest alternatywą do analogicznego problemu oceny zagrożeń i wyboru działań w kategoriach progów zagrożeń.

Graficzną postać tej konceptualizacji, a jednocześnie model problemu w sieci bayesowskiej, przedstawiono na rysunku 1 [Dokumentacja programu BayesiaLab]. W modelu tym wprowadzono następujące pojęcia: faza wegetacyjna uprawy (f), temperatura powietrza (t) i wilgotność powietrza (w). Węzły odpowiadające tym pojęciom są jednocześnie węzłami sieci bayesowskiej, reprezentującej boolowskie zmienne losowe, odpowiadające możliwym konkretyzacjiom odpowiednio fazy fenologicznej, temperatury i wilgotności powietrza (f , t , w).

Pojęciu zagrożenia odpowiada węzeł sieci bayesowskiej, który reprezentuje boolowską zmienną losową, nad którą rozkład prawdopodobieństwa określony jest funkcją logiczną AND zmiennych f (faza wegetacyjna), t i w (temperatura i wilgotność powietrza). Zamiast klasycznej funkcji AND stosujemy w tym przypadku funkcję logiczną NoisyAND co umożliwia uwzględnienie niepewności epistemicznej, dotyczącej wpływu czynników nie uwzględnionych w modelu. Węzeł ten po konkretyzacji reprezentuje zagrożenie (z), które jest funkcją logiczną zmiennych f , t , w .

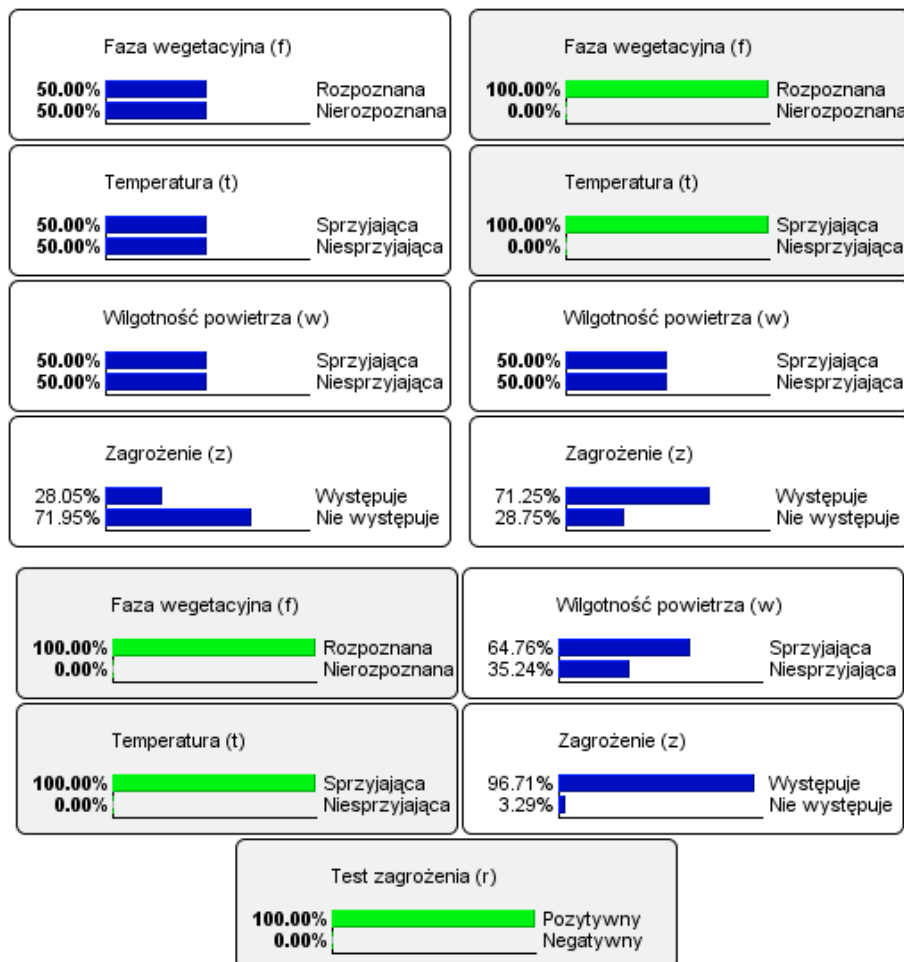
Węzeł *Zagrożenie* a priori reprezentuje potencjalne zagrożenie, które wystąpi w konkretnej fazie dopiero w momencie pojawienia się odpowiednich warunków temperaturowo-wilgotnościowych. Zagrożenie rzeczywiste zostaje określone po przeprowadzeniu pomiaru stopnia występowania zagrożenia. Aposterioryczny rozkład prawdopodobieństwa nad zbiorem wartości zmiennej reprezentującej zagrożenia jest wyznaczany po wykonaniu testu zagrożenia reprezentowanego w modelu przez zmienną *Test*.

W modelu środek ochronny reprezentowany jest przez zmienną boolowską, która jest konkretyzowana dla każdego z potencjalnych zagrożeń według tablicy decyzyjnej określonej przez producentów ochrony. Zagrożenie oraz skonkretyzowany dla niego środek ochronny, po uwzględnieniu stanu uprawy, określa odpowiednią dawkę. Binarystacja tej dawki to: dawka minimalna skuteczna oraz maksymalna dozwolona [Idkowiak 2009; Mrówczyński i in. 2011; Paradowski 2011].



Rys. 1. Model problemu oceny zagrożenia
Fig. 1. A model of the risk assessment problem

Typowy mechanizm funkcjonowania sieci przedstawiony jest na rysunku 2. Na rysunku tym pokazano jak zmienia się rozkład prawdopodobieństwa na zbiorze wartości zmiennej *Zagrożenie* w wyniku posiadania dodatkowych informacji. W tym przypadku uwzględnienie testu oceny zagrożenia znacznie zwiększa przekonanie o wystąpieniu zagrożenia (wzrost prawdopodobieństwa wystąpienia zagrożenia z 0,71 do 0,97).

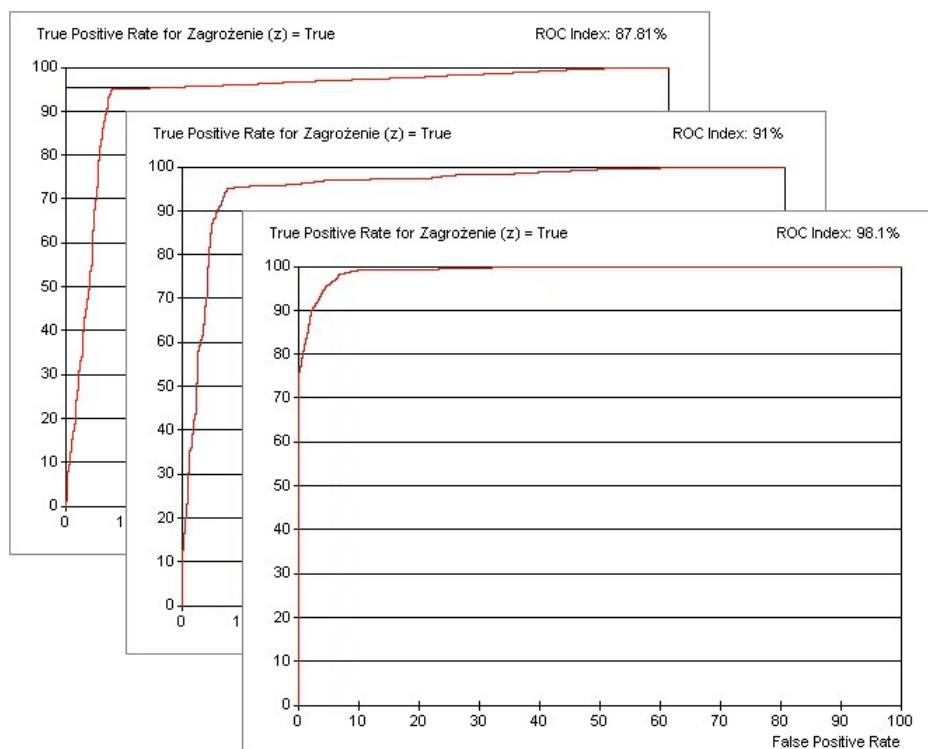


Rys. 2. Wnioskowanie o zagrożeniu
 Fig. 2. Inference about the threat

Na rysunku 3 pokazano w jaki sposób dodanie nowego źródła informacji zmniejsza stopień niepewności co do prawidłowego rozpoznania stopnia zagrożenia (zmienna *Zagrożenie*). Ocenę efektu zmniejszenia niepewności przeprowadzono w oparciu o analizę krzywej ROC. Dodanie nowego źródła informacji skutkuje wzrostem wskaźnika ROC. Wskaźnik ten wynosi 87,81% w przypadku gdy prognoza dawki jest oparta wyłącznie na rozpoznaniu fazy fenologicznej i 98,1% w sytuacji gdy ocena uwzględnia dodatkowe informacje o temperaturze i wilgotności powietrza.

Rozkład prawdopodobieństwa nad wartościami zmiennej *Dawka* środka ochrony jest punktem wyjścia do podjęcia decyzji, którą z tych wartości dawki należy zastosować. Model decyzyjny oparty na technologii sieci bayesowskiej przedstawiono na rysunku 4.

Oprócz węzłów reprezentujących zmienne losowe występują tu węzły reprezentujące decyzje (■) oraz użyteczność (◆). Zbiór decyzji obejmuje dwa działania dotyczące wyboru dawki: Dp-dawka minimalna skuteczna, Di-dawka maksymalna dozwolona. Konsekwencje decyzji zależne od stanu natury i wybranego działania reprezentowane są przez węzeł użyteczność zabiegu.



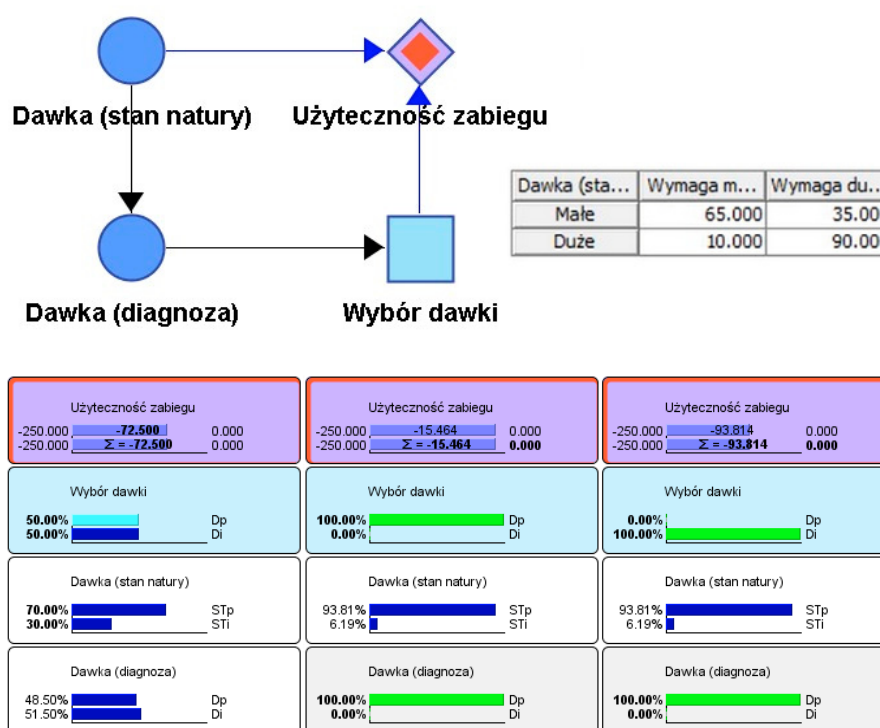
Rys. 3. Krzywe ROC ilustrujące efekt zmniejszenia niepewności poprzez dodanie dodatkowego źródła informacji

Fig. 3. The ROC curves illustrating the effect of reducing uncertainty by adding an additional source of information

W przypadku prawidłowych rozpoznań przypisano im wartość zero, natomiast wybierając dawkę minimalną skuteczną w sytuacji gdy powinna być zastosowana dawka maksymalna dozwolona przyjęto wartość – 250, a w przypadku przeciwnym – 100.

Zmienna *Dawka (stan natury)* reprezentuje wiedzę aprioryczną dotyczącą dawki, która faktycznie powinna być zastosowana w danej sytuacji. Zmienna ta przyjmuje wartości: STp (adekwatna dawka to dawka minimalna skuteczna) i STi (adekwatna dawka to dawka maksymalna dozwolona). Zmienna *Dawka (diagnoza)* przyjmuje takie same wartości jak zmienna *Dawka* występująca w modelu problemu oceny zagrożenia. Ponieważ zmienna

Dawka (diagnoza) jest zależna od zmiennej *Dawka (stan natury)* to warunkowy rozkład prawdopodobieństwa $P\{Dawka (diagnoza) | Dawka (stan natury)\}$ jest wyznaczany na podstawie sieci przedstawionej na rys. 1. Dla zmiennej *Dawka* wygenerowana zostaje krzywa ROC, na podstawie której, przy znajomości prawdopodobieństwa tej zmiennej (dawka minimalna skuteczna i dawka maksymalna dozwolona), można wyznaczyć prawdopodobieństwa prawidłowych i błędnych rozpoznań. Prawdopodobieństwa te są następnie eksportowane do modelu decyzyjnego (rys. 4), jako warunkowy rozkład prawdopodobieństwa zmiennej *Dawka (diagnoza)*.



Rys. 4. Model decyzji w sieci bayesowskiej i wyniki funkcjonowania modułu decyzyjnego
 Fig. 4. Decision model in the bayesian network and the results of functioning

Wyniki funkcjonowania modułu decyzyjnego przedstawiono w postaci rozkładów widocznych w dolnej części rysunku 4. Rozkład prawdopodobieństwa zmiennej *Dawka (diagnoza)* widoczny w pierwszej kolumnie został wyznaczony na podstawie apriorycznego rozkładu zmiennej *Dawka (stan natury)* oraz rozkładu warunkowego określonego na podstawie krzywej ROC wyznaczonej dla węzła *Dawka* (rys.1). Konkretyzacja wyników diagnozy (*Dawka (diagnoza) = Dp*) oraz wybór działania (*Wybór dawki = Dp*) skutkuje wyznaczeniem aposteriorycznego rozkładu prawdopodobieństwa nad zbiorem wartości zmiennej *Dawka (stan natury)* oraz oczekiwanej użyteczności, odpowiadającej wybranej

decyzji (Użyteczność zabiegu = -15,5). Jeżeli w analizowanej sytuacji zostanie wybrana dawka maksymalna dozwolona (Wybór dawki = D_i) to oczekiwana użyteczność wynosi -93,8 i jest to wynik gorszy niż w przypadku wyboru dawki minimalnej skutecznej.

Podsumowanie

Przedstawiona metoda modelowania problemów decyzyjnych dotyczących rozpoznawania zagrożeń, ich oceny i ewentualnej interwencji, może być użytecznym narzędziem wspomagającym proces podejmowania decyzji w przypadku integrowanej produkcji rolniczej.

Przedstawione moduły, które wykorzystują sieci bayesowskie, umożliwiają konceptualizację, specyfikację i analizę sytuacji problemowych w warunkach niedeterministycznych.

Wbudowane mechanizmy uczenia maszynowego i automatycznego wnioskowania zapewniają możliwość adaptacyjności modelu na podstawie danych empirycznych.

Bibliografia

- Fawcett T.** 2004. ROC Graphs: Notes and Practical Considerations for Researchers [online]. Machine Learning. [dostęp 10.06.2011]. Dostępny w internecie http://home.comcast.net/~tom.fawcett/public_html/papers/ROC101.pdf.
- Idkowiak M.** 2009. Bez ochrony fungicydowej nie ma wysokich plonów. *Agronom.pl*. IV/2009. Informator Grupy Osadkowskiej. s. 10-12.
- Korbas M., Mrówczyński M.** 2009. Integrowana produkcja pszenicy ozimej i jarej. Instytut Ochrony Roślin – PIB. Poznań. Maszynopis.
- Mrówczyński M., Korbas M., Szczepaniak W., Paradowski A.** 2011. Zeszyt Technologiczny. Pszenica ozima, jęczmień jary 2011. *Agrosan*. s. 17-52.
- Paradowski A.** 2011. Odchwaścić wiosną zboża ozime. *Agrotechnika*. Nr 3. Uprawa roli i roślin. s. 55-59.
- Dokumentacja programu BayesiaLab. [dostęp 10.06.2011]. Dostępny w internecie <http://www.bayesia.com>.

DECISION PROCESS MODELLING IN THE INTEGRATED AGRICULTURAL PRODUCTION SYSTEM

Abstract. The study presents the modelling method of a risk assessment problem in the agricultural production process and selection of an adequate protection measure and the size of the measure in the integrated technology of the plant production. The decision results from a generative diagnosis based on the distribution of decision probability under the set of possible decisions. The study presented how an integration of non-uniform and uncertain sources of information decreases uncertainty in the process of decision taking. The concept of a model structure was presented based on the technology of the Bayesian networks. Bayesian diagnosis model shows how integration of heterogeneous uncertain information sources decrease uncertainty in process of decision-making

Key words: agricultural production, modeling threats, integration of information sources, supporting decision-making processes, Bayesian networks

Adres do korespondencji:

Piotr Maksym; e-mail: piotr.maksym@up.lublin.pl
Katedra Podstaw Techniki
Uniwersytet Przyrodniczy w Lublinie
ul. Doświadczalna 50a
20-280 Lublin