

## Особенности разработки математической модели для распознавания болезней растений на основе байесовских сетей

*А.В. Чирков, П.С. Комаров, Е.В. Конюхов, Н.С. Миронов*

*Национальный исследовательский институт «Московский институт электронной техники»,  
г. Москва, Россия*

**Аннотация:** В данной статье предлагается точный и динамический метод диагностики болезней сельскохозяйственных культур. Предлагаемый метод использует байесовские сети для представления взаимосвязей между симптомами и болезнями сельскохозяйственных культур. Этот метод имеет одно главное отличие от существующих методов диагностики: он не использует все симптомы для постановки диагноза, а целенаправленно выбирает подмножество симптомов, которые наиболее важны для диагностики; выбор активного симптома основан на концепции марковского одеяла в байесовской сети. Теоретический анализ показывает, что предлагаемый метод позволяет значительно повысить эффективность диагностики болезней сельскохозяйственных культур.

**Ключевые слова:** распознавание болезней растений, математическая модель, байесовская сеть, машинное обучение, марковское одеяло.

**Введение.** Диагностика болезней сельскохозяйственных культур является важнейшей задачей аграрного производства. Болезни сельскохозяйственных культур стали одним из важнейших факторов, ограничивающих большие урожаи и высокое качество сельскохозяйственных культур, и вызывают колоссальные производственные и экономические потери в сельском хозяйстве. Например, ржавчина сои (разновидность болезни соевых бобов) вызвала снижение урожайности на 20%, и фермеры могли потерять около 11 миллионов долларов прибыли в 2006 году в США [1]. Болезни риса вызвали снижение урожайности на 40-50% в особенно тяжелое время в Китае [2]. Таким образом, проблема быстрой и точной диагностики болезней сельскохозяйственных культур является своевременной и актуальной.

В последние годы было разработано множество интеллектуальных методов диагностики [3 - 5] болезней сельскохозяйственных культур. Однако некоторые проблемы ограничивают существующие методы [6]. Во-первых,

---

существующие методы всегда используют для диагностики все симптомы заболевания, но из-за возникающего шума при сборе информации не все данные о симптомах являются точными, и не все симптомы являются важными для диагностики. Более того, использование большего количества симптомов потребует больших вычислительных затрат, что повлияет на скорость диагностики. Таким образом, необходимо повысить эффективность выбора оптимального подмножества симптомов, которые являются наиболее информативными и релевантными для диагностики. В условиях необходимости быстрого принятия решения при наличии большого количества симптомов хорошо подходит аппарат нечеткой логики [7]. Во-вторых, существует множество подходов к построению моделей диагностики. Наиболее широко используемыми подходами являются: классификация болезней при помощи метода опорных векторов и гиперспектральном отражении [8], экспертные системы оценки [9, 10], классификация болезней на основе оптимизированной кластеризации по ближайшим симптомам и самоорганизующейся карте Кохонена [11] и т.д. Несмотря на различия методов, модели диагностики всегда статичны: после того, как они сформированы, они не меняются. Но окружающая среда всегда динамична и меняется со временем; наиболее важные для диагностики симптомы в летнее время года могут отличаться от «зимних» симптомов. Следовательно, модель диагностики должна отражать временные изменения динамической среды.

**Теоретическая часть.** Байесовская сеть — это графическая модель для представления взаимосвязей между переменными. Для набора переменных  $X = \{X_1, X_2, \dots, X_n\}$  байесовская сеть представляет собой кортеж  $B = (G, \theta)$ , который включает в себя два компонента:

- $G$  — ориентированный ациклический граф, где каждый узел графа представляет собой переменную, а каждая дуга представляет прямую
-

зависимость между переменными;

- $\theta = \{P(X_i|\pi_i), 1 \leq i \leq n\}$  — представляет собой набор параметров, который представляет условное распределение вероятностей каждого узла с учетом значений их родительских узлов, где  $\pi_i$  представляет собой набор родительских узлов  $X_i$ ;  $\theta$  также называют таблицей условных вероятностей для каждого узла.

Совместное распределение вероятностей  $X$  может быть представлено, как произведение всех переменных таблицы условных вероятностей [12]:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i|\pi_i)$$

Простая структура байесовской сети изображена на рис. 1. В байесовской сети марковское одеяло [13] узла  $X_i$  (которое может быть представлено как  $MB(X_i)$  для краткости) представляет собой набор узлов, состоящий из его родительских узлов, его дочерних узлов и родительских узлов его дочерних узлов. Марковское одеяло переменной  $X_5$  равно  $\{X_2, X_3, X_4, X_7\}$ .

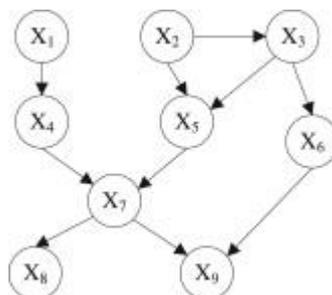


Рис. 1. – Простой пример структуры байесовской сети

Существует важное свойство марковского одеяла: марковское покрытие переменной  $X_i$  — это множество узлов, которое делает  $X_i$  независимым от всех других узлов сети. Такие узлы, принадлежащие марковскому одеялу, являются узлами, наиболее релевантными по отношению к  $X_i$ :

$$P(X_i | X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n) = P(X_i | MB(X_i))$$

**Предлагаемое решение.** В предлагаемом подходе симптомы

представлены узлами байесовских сетей. Вводится еще один узел под названием «Diseases» (заболевания); каждое значение этого узла представляет одну болезнь культуры. В случае, если есть болезни  $k$  для определенной культуры, то значение этого узла равно  $Diseases = \{d_1, \dots, d_k\}$ , где  $d_i$  представляет собой  $i$  – ую болезнь. Если симптомов всего  $n$ , то байесовская сеть имеет  $n + 1$  узлов:

$$\{Symptom_1, \dots, Symptom_n, Diseases\}$$

Таким образом, байесовская сеть может четко показать взаимосвязь между симптомами и заболеваниями. Результат диагностики можно получить путем вычисления апостериорного распределения вероятностей:

$$P(\{Diseases | Symptom_1, \dots, Symptom_n\})$$

Модель диагностики для 8 симптомов, основанная на байесовской сети в предлагаемом методе, изображена на рис. 2.

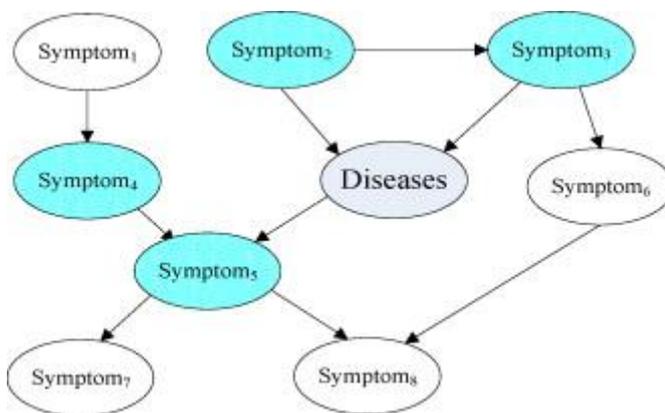


Рис. 2. – Простой пример структуры байесовской сети

Для краткости и удобства переменная  $Symptom_i$  представлена как  $S_i$ . Целью активного выбора симптомов является выбор подмножества  $S^* \subseteq \{S_1, \dots, S_n\}$ , а симптомы в нем являются наиболее информативными и релевантными для диагностики болезней сельскохозяйственных культур. Из свойства марковского одеяла, упомянутого ранее, известно, что в

байесовской сети узел *Disease* условно независим от других узлов при заданном марковском одеяле узла *Disease*:

$$P(Diseases|S_1, \dots, S_n) = P(Diseases|MB(Diseases))$$

Это означает, что при заданных значениях узлов, на вероятность внутри узла *Disease* влияет только его марковское одеяло. Поэтому марковское одеяло  $MB(Diseases)$  является наиболее информативным и релевантным для диагностики. Таким образом, мы можем выбрать только узлы (симптомы), принадлежащие  $MB(Diseases)$  для расчета результата диагностики. На рис. 2 марковское одеяло *Diseases* равно  $\{S_2, S_3, S_4, S_5\}$ , следовательно, необходимо использовать для диагностики эти симптомы.

Теперь необходимо определиться, как получить результат диагностики. Задачу вычисления результата диагностики можно сформулировать математически следующим образом: предположим, что у нас есть наблюдаемое значение симптомов, принадлежащих  $MB(Diseases)$ , и результат диагностики может быть получен путем вычисления условной апостериорной вероятности  $P(Diseases|MB(Diseases))$ .

Предполагая, что переменная *Diseases* имеет  $k$  возможных значений, другими словами, диапазон  $Diseases = \{d_1, d_2, \dots, d_k\}$ , необходимо выбрать  $d$  с наибольшим значением  $P(Diseases = d|MB(Diseases))$  в качестве окончательного результата диагностики:

$$Diagnosis\ Result = \arg \max_{d \in Diseases} P(Diseases = d|MB(Diseases))$$

Исходя из всего вышесказанного, если выбранное подмножество симптомов представляет собой марковское одеяло узла *Diseases* в байесовской сети, то результат диагностики вычисляется так:

### Diagnosis Result

$$= \arg \max_{d \in \text{Diseases}} P(\text{Diseases} = d | \pi(\text{Diseases})) \quad (1)$$
$$\cdot \prod_{S_i \in \text{Children}(\text{Diseases})} P(S_i | \pi(S_i))$$

, где  $\pi(\text{Diseases})$  — родительские узлы  $\text{Diseases}$ , а  $\text{Children}(\text{Diseases})$  — дочерние узлы  $\text{Diseases}$  в байесовской сети.

Благодаря свойствам марковского одеяла можно сделать вывод, что:

$$P(\text{Diseases} | \text{MB}(\text{Diseases})) = P(\text{Diseases} | S_1, \dots, S_n)$$
$$= \frac{P(\text{Diseases}, S_1, \dots, S_n)}{P(S_1, \dots, S_n)}$$

На основании [14], поскольку знаменатель  $P(S_1, \dots, S_n)$  не включает  $\text{Diseases}$ , это означает, что независимо от того, какое значение принимает переменная  $\text{Diseases}$ , результат для знаменателя  $P(S_1, \dots, S_n)$  будет одинаковым, поэтому  $P(S_1, \dots, S_n)$  можно рассматривать как константу. Более того, числитель  $P(\text{Diseases}, S_1, \dots, S_n)$  представляет собой совместное распределение вероятностей и может быть представлен таблицами условных вероятностей каждого из узлов.

Таким образом, приведенное выше уравнение имеет следующий вид:

$$\frac{P(\text{Diseases}, S_1, \dots, S_n)}{P(S_1, \dots, S_n)} = c \cdot P(\text{Diseases}, S_1, \dots, S_n)$$
$$= c \cdot \prod_{X_i \in \{\text{Diseases}, S_1, \dots, S_n\}} P(X_i | \pi(X_i))$$
$$= c \cdot P(\text{Diseases} | \pi(\text{Diseases}))$$
$$\cdot \prod_{S_i \in \text{Children}(\text{Diseases})} P(S_i | \pi(S_i))$$
$$\times \prod_{S_j \neq \text{Diseases} \cap S_j \notin \text{Children}(\text{Diseases})} P(S_j | \pi(S_j))$$

, где  $c$  - это замена константы  $P(S_1, \dots, S_n)$ , а приведенное выше уравнение делит произведение таблицы условных вероятностей на три части: таблица условных вероятностей узла *Diseases*, таблица условных вероятностей дочерних узлов *Diseases* и таблица условных вероятностей других узлов. Так как третья часть  $\prod_{S_j \neq Diseases \cap S_j \notin Children(Diseases)} P(S_j | \pi(S_j))$  не содержит узла *Diseases*, то она так же является константой.

Следовательно, уравнение можно упростить:

$$\begin{aligned} c \cdot P(Diseases, S_1, \dots, S_n) \cdot \prod_{S_i \in Children(Diseases)} P(S_i | \pi(S_i)) \\ \cdot \prod_{S_j \neq Diseases \cap S_j \notin Children(Diseases)} P(S_j | \pi(S_j)) \\ = c' \cdot P(Diseases | \pi(Diseases)) \cdot \prod_{S_i \in Children(Diseases)} P(S_i | \pi(S_i)) \end{aligned}$$

Данное уравнение демонстрирует то, что значение  $P(Diseases | MB(Diseases))$  пропорционально значению  $P(Diseases | \pi(Diseases)) \cdot \prod_{S_i \in Children(Diseases)} P(S_i | \pi(S_i))$ .

Из этого следует, что:

*Diagnosis Result*

$$\begin{aligned} &= \arg \max_{d \in Diseases} P(Diseases = d | MB(Diseases)) \\ &= \arg \max_{d \in Diseases} P(Diseases = d | \pi(Diseases)) \\ &\cdot \prod_{S_i \in Children(Diseases)} P(S_i | \pi(S_i)) \end{aligned}$$

**Выводы.** Уравнение (1) содержит таблицу условных вероятностей только узла и его дочерних узлов, которые получены непосредственно из байесовских сетей, поэтому для вычисления уравнения не требуется никакого вывода. Благодаря этому предлагаемый метод диагностики

является высокоэффективным. Модель диагностики на основе байесовской сети создана на основе знаний экспертов в области сельского хозяйства, или на основе машинного обучения, или при помощи комбинирования обоих способов. Таким образом, теоретически обосновано и доказано использование предложенного метода, основанного на байесовских сетях. Планируется дальнейшее развитие данного подхода при распознавании болезней растений и программная реализация на высокоуровневом языке программирования Python.

### Литература

1. Tocatlidou A., Passam H.C., Sideridis A.B., Yialouris C.P. Reasoning under uncertainty for plant disease diagnosis. *Expert Systems*, 19 (1). 2002. pp. 46-52.
  2. Kolhe S., Kamal R., Harvinder S.S., Gupta G.K. A web-based intelligent disease-diagnosis system using a new fuzzy-logic based approach for drawing the inferences in crops. *Computers and Electronics in Agriculture*, 76 (1). 2011. pp. 16-27.
  3. Пархоменко, П.П., Согомян Е.С. Основы технической диагностики. М.: Энергоиздат, 1981. 321 с.
  4. Основы технической диагностики / под ред. П.П. Пархоменко. М.: Энергия, 1976. 464 с.
  5. Чирков, А.В., Гагарина Л.Г. Разработка алгоритма на основе матриц смежности и нечеткой логики для распознавания болезней растений // Материалы научно-технической конференции "Микроэлектроника и информатика – 2022". М.: Национальный исследовательский университет "Московский институт электронной техники" (Москва), 2022. С. 129-135.
  6. Гагарина Л.Г. Современные проблемы информатики и вычислительной техники: учеб. пособие. М.: МИЭТ, 2017. 368 с.
  7. Чирков, А.В. Особенности применения аппарата нечеткой логики
-

для распознавания болезней растений // Системы компьютерной математики и их приложения, 2022, №23 URL: [fizmat.smolgu.ru/images/stories/SKMP/skmp-2022.pdf](http://fizmat.smolgu.ru/images/stories/SKMP/skmp-2022.pdf).

8. Rumpf T., Mahlein A.K., Steiner U., Oerke E.C., Dehne H.W., Plumer L. Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. *Computers and Electronics in Agriculture*, 74. 2010. pp. 91-99.

9. Devraj R. Jain. PulsExpert: an expert system for the diagnosis and control of diseases in pulse crops. *Expert Systems with Applications*, 38. 2011. pp. 11463-11471.

10. Gonzalez-Andujar J.L. Expert system for pests, diseases and weeds identification in olive crops. *Expert Systems with Applications*, 36. 2009. pp. 3278-3283.

11. Zhang K., Chai Y., Kuang J.J. Method for multi-disease diagnosis based on optimized symptom adjacent-searching clustering and SOM NN. *Transactions of the CSAE*, 27 (1). 2011. pp. 215-222.

12. de Campos L.M. A scoring function for learning Bayesian networks based on mutual information and conditional independence tests. *Journal of Machine Learning Research*, 7. 2006. pp. 2149-2187.

13. Pearl J. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, 1988. p. 584.

14. Buntine W. L. Operations for learning with graphical models. *Journal of Artificial Intelligence Research*, 2. 1994. pp. 159-225.

### References

1. Tocatlidou A., Passam H.C., Sideridis A.B., Yialouris C.P. Reasoning under uncertainty for plant disease diagnosis. *Expert Systems*, 19 (1). 2002. pp. 46-52.

2. Kolhe S. , Kamal R., Harvinder S.S., Gupta G.K. A web-based intelligent disease-diagnosis system using a new fuzzy-logic based approach for drawing the

inferences in crops. Computers and Electronics in Agriculture, 76 (1). 2011. pp. 16-27.

3. Parhomenko, P.P. Osnovy tehnicheckoj diagnostiki [Fundamentals of technical diagnostics]. M.: Jenergoizdat, 1981. 321 p.

4. Osnovy tehnicheckoj diagnostiki [Fundamentals of technical diagnostics]. M.: Jenergija, 1976. 464 p.

5. Chirkov, A.V., Gagarina L.G. Materialy nauchno-tehnicheckoj konferencii "Mikroelektronika i informatika 2022" (Materials of the scientific and technical conference "Microelectronics and Informatics 2022"). Moscow, 2022. pp. 129-135.

6. Gagarina L.G. Sovremennye problemy informatiki i vychislitel'noj tehniki: ucheb. posobie [Modern problems of informatics and computer technology: study guide]. M.: MIJeT, 2017. 368 p.

7. Chirkov, A.V. Sistemy komp'juternoj matematiki i ih prilozhenija, 2022, №23 URL: [fizmat.smolgu.ru/images/stories/SKMP/skmp-2022.pdf](http://fizmat.smolgu.ru/images/stories/SKMP/skmp-2022.pdf). Rumpf T., Mahlein A.K., Steiner U., Oerke E.C., Dehne H.W. , Plumer L. Computers and Electronics in Agriculture, 74. 2010. pp. 91-99.

8. Devraj R. Jain. Expert Systems with Applications, 38. 2011. pp. 11463-11471.

9. Gonzalez-Andujar J.L. Expert Systems with Applications, 36. 2009. pp. 3278-3283.

10. Zhang K., Chai Y., Kuang J.J. Transactions of the CSAE, 27 (1). 2011. pp. 215-222.

11. de Campos L.M. Journal of Machine Learning Research, 7. 2006. pp. 2149-2187.

12. Pearl J. Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann, 1988. p. 584.



13. Buntine W. L. Journal of Artificial Intelligence Research, 2. 1994. pp. 159-225.