

УДК 631.526.32:633.367

ИДЕНТИФИКАЦИЯ ПОДОБНЫХ СОРТОВ ЛЮПИНА ЖЕЛТОГО С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМА К-БЛИЖАЙШИХ СОСЕДЕЙ

В. Б. КУТОВЕНКО

*Национальный университет биоресурсов и природопользования Украины,
г. Киев, Украина, 03041*

Н. С. ОРЛЕНКО, Н. П. КОСТЕНКО, С. П. ЛЕКАРЬ

*Украинский институт экспертизы сортов растений,
г. Киев, Украина, 03041*

(Поступила в редакцию 15.05.2019)

Мониторинг рынка кормовых культур показывает расширение сети животноводческих хозяйств, у которых возникает необеспеченная потребность в недорогом растительном белке. Особое место в производстве растительного белка отводится зернобобовым культурам, которые содержат в 2–3 раза больше белковых веществ по сравнению с зерновыми. Одной из таких культур является люпин, который хорошо приспособлен к произрастанию в природно-климатических условиях Белоруссии, Украины, Польши и средней полосы России. Важность кормового и сидерального направления этой культуры подтверждает широкий спектр научных исследований [1–8].

*Люпин – одна из высокобелковых культур, которая широко используется в практике земледелия Беларуси, Польши, России и Украины. Государственный реестр сортов растений, пригодных для распространения в Украине, по состоянию на апрель 2019 года насчитывает 8 сортов люпина желтого (*Lupinus luteus* L.). Процесс точной идентификации и группировки новых сортов по морфологическим признакам требует применения средств интеллектуального анализа данных. В статье показана возможность использования алгоритма *k*-ближайших соседей, как одного из средств интеллектуального анализа данных, для поиска подобных по морфологическим признакам общеизвестных сортов люпина желтого. В процессе исследования были использованы результаты полевых испытаний люпина желтого, а именно: описания морфологических качественных (OL), количественных (QN) и псевдокачественных (PQ) признаков, кодовые формулы признаков и построение компьютерной модели общеизвестных сортов люпина желтого, которые зарегистрированы в Украине с использованием алгоритма ближайших соседей. В результате были установлены наиболее подобные общеизвестные сорта люпина желтого по морфологическим признакам.*

Ключевые слова: *люпин желтый, морфологические признаки, общеизвестные сорта, алгоритм ближайших соседей.*

Monitoring of the forage crop market shows the expansion of network of livestock farms in which there is an unsecured need for inexpensive vegetable protein. A special place in the production of vegetable protein is given to leguminous crops, which contain 2–3 times more protein substances compared to cereals. One of such crops is lupine, which is well adapted to growing in the natural-climatic conditions of Belarus, Ukraine, Poland and central Russia. The importance of the feed and green-manure directions of this crop is confirmed by a wide range of scientific studies.

*Lupine is one of the high-protein crops that is widely used in agriculture in Belarus, Poland, Russia and Ukraine. The state register of plant varieties suitable for distribution in Ukraine, as of April 2019, has 8 varieties of yellow lupine (*Lupinus luteus* L.). The process of accurate identification and grouping of new varieties according to morphological characteristics requires the use of data mining. The article shows the possibility of using the algorithm of *k*-nearest neighbors, as one of the means of data mining, to search for similar morphological signs of the well-known varieties of yellow lupine. During the study, the results of field tests of yellow lupine were used, namely: descriptions of morphological qualitative (OL), quantitative (QN) and pseudo-quality (PQ) signs, code formulas of signs and the construction of a computer model of well-known varieties of yellow lupine, which were registered in Ukraine using nearest neighbours algorithm. As a result, the most similar well-known varieties of yellow lupine were identified by morphological characteristics.*

Key words: *yellow lupine, morphological signs, well-known varieties, algorithm of nearest neighbors.*

Введение

Государственный реестр сортов растений, пригодных для распространения в Украине на 26 марта 2019 года насчитывает восемь сортов люпина желтого [9].

Целью статьи является определение наиболее подобных морфологических признаков среди общеизвестных сортов с целью дальнейшего использования выделенных групп подобных сортов для проведения экспертизы на отличимость сортов-кандидатов люпина желтого после первого года испытания.

Авторами статьи была поставлена задача идентифицировать общеизвестные сорта из базы данных на группы подобных сортов люпина желтого с использованием инструмента machine learning – алгоритма ближайшего соседа.

В отечественной и зарубежной литературе широко представлено использование средств machine learning и в частности алгоритма ближайшего соседа [10–12].

Согласно этому алгоритму, предполагается, что уже имеется какое-то количество объектов с точной классификацией (в нашем случае – подобных сортов растений), и нужно выработать правило, которое позволяет отнести новый сорт к одному из возможных классов (набору подобных по морфологическим признакам сортов растений).

KNN алгоритм осуществляет подбор коэффициента k – количество записей, которые будут считаться близкими с использованием таких правил:

- 1) $((x,y) \geq 0, d(x,y) = 0$ тогда и только тогда, когда $x = y$;
- 2) $d(x,y) = d(y,x)$;
- 3) $d(x,z) \leq d(x,y) + d(y,z)$, при условии, что точки x, y, z не лежат на одной прямой; где x, y, z – векторы признаков сравниваемых объектов.

Упорядочение значений атрибутов проводится с использованием расстояния Евклида, которая используется в качестве меры сходства наблюдений и рассчитывается по формуле:

$$D_E = \sqrt{\sum_1^n (x_i - y_i)^2},$$

где n – количество атрибутов.

В случае использования переменных, которые относятся к количественной шкале могут быть нормализованы с использованием формулы:

$$X^x = \frac{X - \min(X)}{\max(X) - \min(X)}$$

Расчеты были выполнены с использованием статистического пакета IBM SPSS Statistics «Statistical Package for the Social Sciences» [13].

Основная часть

Использованы данные результатов испытаний сортов люпина желтого на отличимость, однородность и стабильность (ООС) (tests of a selection achievement for distinctiveness, uniformity and stability) за период 2002–2018 гг., которые проводились Украинским институтом экспертизы сортов растений в соответствии с международными требованиями и в соответствии с Методикой проведения экспертизы сортов люпина белого, узколистого, желтого (*Lupinus albus* L., *L. angustifolius* L., *L. luteus* L.) на отличимость, однородность и стабильность [14].

Для компьютерной обработки данных результатов исследований морфологических признаков использованы номинальная и порядковая шкалы. Номинальная шкала применялась для группирования подобных сортов по следующим морфологическим признакам: антоциановая окраска стебля в фазе бутонизации, интенсивность зеленого цвета листка в фазе бутонизации, окраска крыльев цветка, окраска кончика лодочки цветка, наличие орнаментации семян, окраска орнаментации семян, распределение орнаментации семян, плотность орнаментации семян, а порядковая шкала для признаков: наличие горького вещества в семенах, высота растения в вегетативной стадии, время цветения, высота растения в фазе начала цветения, длина центрального листка (верхушечный листочек), ширина центрального листка (верхушечный листочек), тип роста растения, время наступления зеленой спелости боба, высота крепления первого соцветия растения в период зеленой спелости (от поверхности почвы до первого соцветия), высота растения в период зеленой спелости, длина боба, время созревания, масса 100 штук семян.

Для построения компьютерной модели сортов люпина желтого были использованы данные результатов описания морфологических признаков общеизвестных сортов и обработаны с использованием алгоритма k -ближайших соседей (k -nearest neighbor

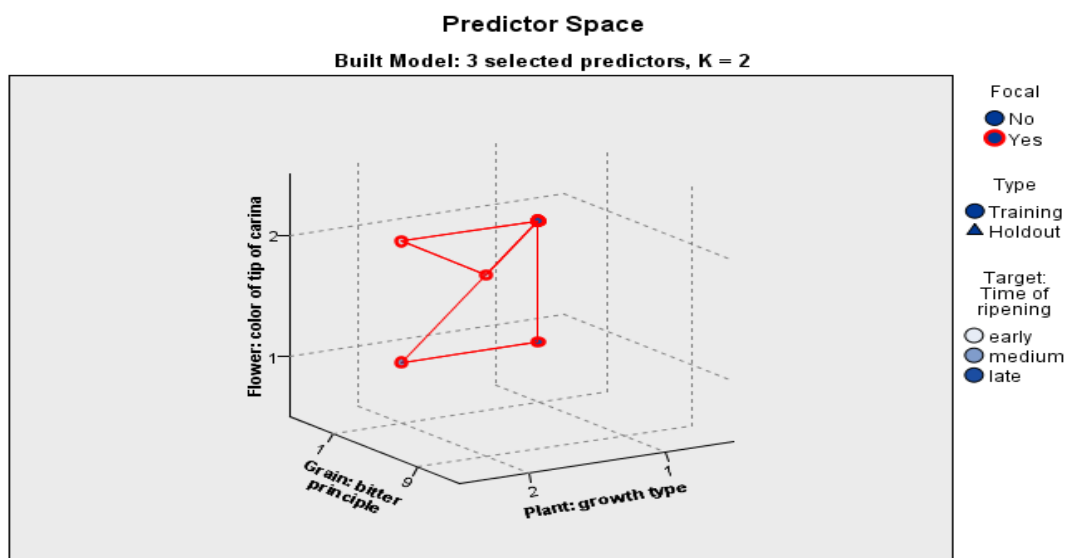
algorithm, KNN). Для модели были заданы следующие параметры: метка наблюдения – название сорта и показатели 14 морфологических признаков. В качестве фокусных признаков: наличие горького вещества в семенах, тип роста растения, высота растения в период зеленой спелости, окраска крыльев цветка, три из которых относятся к порядковой шкале и один к номинальной.

Фокусные признаки были использованы пакетом IBM SPSS для перекрестной проверки. Это дало возможность получить более точные результаты классификации высоты растения в вегетативной стадии. В качестве идентификатора фокусного наблюдения была задана переменная, которая содержит данные о времени цветения, а в качестве зависимой переменной – переменная, содержащая данные о времени наступления зеленой спелости растения. Для расчетов параметров модели была выбрана метрика Евклида.

Фокусные наблюдения модели сортов люпина желтого и их k ближайших соседей отображены на диаграмме рассеяния или диаграмме пространства с целевой переменной (рис.1).

показатели модели – признаки «наличие горького вещества в растении» (Grain: bitter principle), «тип роста растения» (Plant: growth type) и «окраска крыльев цветка» (Flower: color of wings). Точки на диаграмме показывают значения степени проявления признаков общеизвестных сортов, которые находятся в контрольной группе.

Выделено восемь сортов люпина желтого, которые использовались в качестве обучающей выборки модели. В качестве фокусных наблюдений были использованы следующие признаки: наличие горького вещества в растении, тип роста растения, окраска крыльев цветка и высота растения в фазе зеленой спелости (Plant: height at green ripening).



This chart is a lower-dimensional projection of the predictor space, which contains a total of 8 predictors.

Рис. 1. Диаграмма пространства модели сортов люпина желтого

Интерактивная диаграмма пространства показателей состоит из трех осей, которые представляют

В результате расчета модели сформирована таблица наиболее подобных общеизвестных сортов люпина желтого.

Таблица 1. Подобные сорта люпина желтого

Название сорта (фокусного)	Название первого подобного сорта	Название второго подобного сорта	Расстояние между фокусным и первым	Расстояние между фокусным и вторым
Агат Полісся	Світязь	Макарівський	1,414	2,449
Макарівський	Світязь	Агат Полісся	2,449	2,449
Олімп	Рябчик	Серпневий	2,449	2,449
Рапсодія	Смена	Серпневий	2,449	2,828

Рябчик	Серпневий	Олімп	2,449	2,449
Світязь	Агат Полісся	Макарівський	1,414	2,449
Серпневий	Рябчик	Олімп	2,449	2,449
Смена	Світязь	Агат Полісся	2,449	2,449

Как свидетельствуют данные табл. 1, наиболее подобными сортами являются сорта Агат Полісся и Світязь (расстояние составляет 1,414), наиболее отличным был сорт Рапсодія (расстояние до наиболее подобного сорта составляет 2,449, а до следующего подобного 2,828).

У сортов Агат Полісся, Світязь и Макарівський отсутствует горькое вещество в семенах, тип роста растений является детерминантным, окраска кончиков лодочки цветка сине-черная, время наступления зеленой спелости боба – среднее, высота растения в фазе зеленой спелости – средняя, отсутствует антициановая окраска стебля растения, время цветения – среднее. Однако, у сортов Агат Полісся и Світязь высота растения в фазе зеленой спелости и в фазе начала цветения – низкая, а у сорта Макарівський – высокая, масса 100 штук семян у сортов Агат Полісся и Світязь – средняя, а у сорта Макарівський – высокая и средняя соответственно. Эти три сорта отличаются по признаку окраски кончика крыльев цветка. Так, у сорта Агат Полісся была окраска светло-желтая, у сорта Світязь – розовая, а у сорта Макарівський – синева-белая.

У сортов Олімп, Рябчик и Серпневий отсутствует горькое вещество в семенах, окраска кончиков лодочки цветка сине-черная, срок наступления зеленой спелости боба и зеленой спелости растения – ранний. Однако сорта Рябчик и Серпневий по типу роста – детерминантные, а сорт Олімп – индетерминантный. Антациановая окраска стебля сортов Олімп и Серпневий – слабая, а у сорта Рябчик – средняя. Масса 100 штук семян была у сортов Олімп и Серпневий – средняя, а у сорта Рябчик – малая. Эти три сорта отличаются по высоте в фазе зеленой спелости растения. Так, у сорта Олімп высота растения – низкая, у сорта Рябчик – средняя, а у сорта Серпневий – высокая.

Сорт Рапсодія наиболее отличается от всех других сортов. Только у этого сорта позднее время цветения, наступление зеленой спелости боба и время зеленой спелости растения – поздний. Кроме того, этот сорт отличается длинным размером боба. Сорт Рапсодія наиболее подобен к сорту Смена. Общим у них является отсутствие горького вещества в семенах, отсутствие орнаментации семени, желтая окраска кончика лодочки цветка, высокая масса 100 штук семян и отсутствие антоциановой окраски стебля растения.

Дополнительное представление модели сортов люпина желтого демонстрирует диаграмма важности переменных, которая состоит из элементов модели (рис. 2, 3 и 4). Как свидетельствует рис. 2, горькое вещество в семени присутствует только у сорта Пеликан.

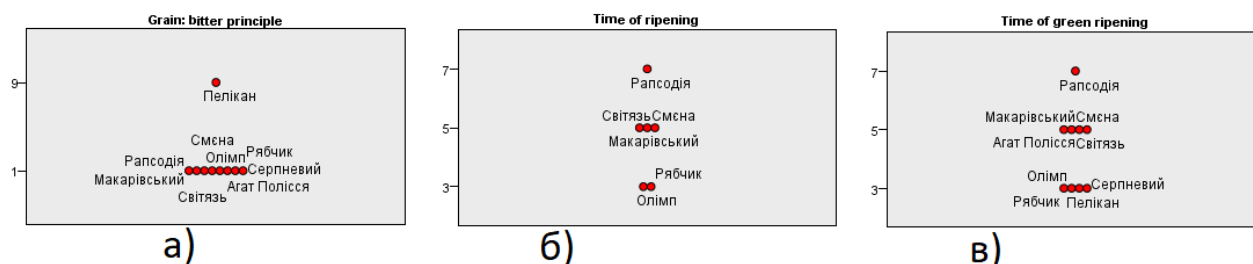


Рис. 2. Проявления признаков: а) наличие горького вещества в семени (Grain: bitter principle), б) время наступления спелости (Time of ripening), в) время наступления зеленой спелости боба (Time of green ripening)

Только сорт Рапсодія имеет поздний срок спелости, Агат Полісся, Макарівський, Світязь и Смена имеют средний срок спелости, а Олімп Серпневий, Рябчик и Пеликан являются ранними.

Рис. 3 показывает, что тип роста сортов Смена и Олімп является индетерминантным, а у сортов Рапсодія, Світязь, Рябчик и Макарівський – детерминантным.

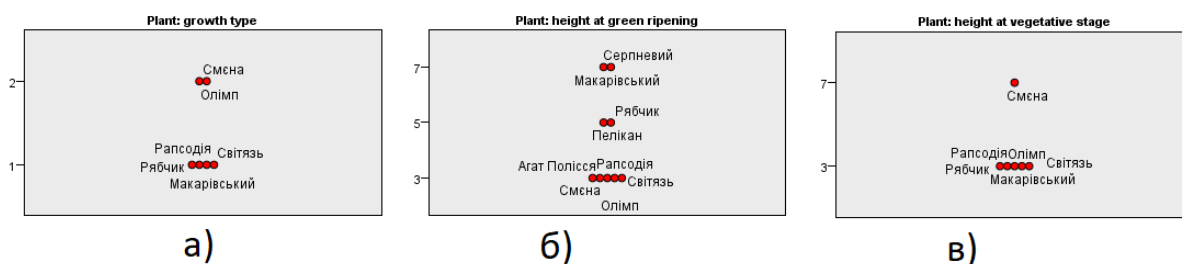


Рис. 3. Проявления признаков: а) тип роста растения (Plant: growth type), б) высота растения в фазе зеленой спелости (Plant: height at green ripening) в) высота в вегетативной стадии (Plant: height at vegetative stage)

Только сорт Смена является высоким в вегетативной стадии, но не смотря на это высокими в стадии зеленой спелости являются сорта Серпневий и Макарівський.

Длинный боб имеют сорта Рапсодія и Серпневий, а все другие – низкий. Цвет кончика лодочки цветка желтый у сортов Рапсодія и Смена, а у всех других сортов – сине-черный. Антоциановая окраска стебля в фазе бутонизации сорта Рябчик умеренное, у сорта Олімп – слабое и отсутствует у всех других сортов.

Заклучение

Была построена модель сортов люпина желтого с использованием алгоритма k-ближайших соседей, что относится к средствам интеллектуального анализа данных (Machine Learning). Выявлены наиболее подобные и наиболее отличимые сорта люпина желтого (*Lupinus luteus* L.). Таким образом, определены наиболее подобные между собой сорта: Агат Полесся и Свитязь, затем следуют сорта Макаровский и Серпневий. А в следующую группу подобных по морфологическим признакам вошли сорта: Олимп, Свитязь и Агат Полесся. Сорт Рапсодия имеет значительное отличие от других сортов люпина желтого, в связи с этим не был включен не в одну из групп. В результате моделирования сформирована выборка выделенных групп подобных сортов для проведения экспертизы на отличимость сортов-кандидатов люпина желтого после первого года испытания.

ЛИТЕРАТУРА

1. Анохина, В. С. Гомологическая изменчивость признаков разных видов люпина / В. С. Анохина, И. Б. Саук, И. Ю. Романчук; под ред. В. М. Юрина / Физиологические, биохимические и молекулярные основы функционирования биосистем: труды белорусского государственного университета. – Минск, 2015. – Том 1. – С. 111–115.
2. Молекулярно-генетическая и биохимическая характеристика образцов стержневой коллекции люпина желтого / И. Ю. Романчук [и др.] // Физиологические, биохимические и молекулярные основы функционирования биосистем: труды белорусского государственного университета; под ред. В. М. Юрина. – Минск, 2015. – Том 1. – С. 111–115.
3. Использование алкалоидных и белковых комплексов в качестве маркерных признаков генетической характеристики образцов разных видов люпина / И. Ю. Романчук [и др.] // Журнал Белорусского государственного университета. Биология. – 2017. – №1. – С. 49–55.
4. Анализ межсортного полиморфизма люпина желтого (*Lupinus luteus* L.) с использованием EST-SSR и SRAP-RGA маркеров / Е. Н. Сысолятин [и др.] // Весці НАН Беларусі. Серыя біялагічных навук. – 2018. – Т. 63, № 3. – С. 298–306.
5. Агеева, П. А. Люпин – перспективная высокобелковая кормовая культура для различных регионов Российской Федерации / А. П. Агеева, М. И. Лукашевич, Н. А. Почутина // Нива Татарстана. – 2013. – № 4–5. – С. 35–37.
6. Bahmat M. I., Mazur V. A., Didur I. M., Pansyryeva H. V., & Telekalo N. V. (2018). Bioenergy efficiency of the usage of biopreparations for the growth of white lupine in the conditions of the Right-Bank Forest-Steppe of Ukraine. *Ukrainian Journal of Ecology*, 8(3): 203208.
7. Mazur V. A., Didur I. M., Pansyryeva H. V., & Telekalo N. V. (2018). Energy-economic efficiency of growth of grain-crop cultures in the conditions of right-bank Forest-Steppe of Ukraine. *Ukrainian Journal of Ecology*, 8(4): 26-33. (in Ukraine).
8. Mazur, V.A., Mazur, K.V., Pansyryeva, H.V., Alekseev, O.O. (2018). Ecological and economic evaluation of varietal resources *Lupinus albus* L. in Ukraine. *Ukrainian Journal of Ecology*, 8(4), 148-153 (in Ukraine).
9. Державний реєстр сортів рослин, придатних для поширення в Україні станом на 26.03. 2019 р. URL: <http://sops.gov.ua/uploads/page/5bbdf6a297647.pdf>
10. Лесковец, Ю. Анализ больших наборов данных / Ю. Лесковец, А. Раджараман. — М.: ДМК, 2016. — 498 с.
11. Марманис, Х. Алгоритмы интеллектуального Интернета. Передовые методики сбора, анализа и обработки данных / Х. Марманис, Д. Бабенко. — М.: Символ, 2011. — 480 с.
12. Brett Lantz. *Machine Learning with R*. Pack Publishing. Birmongham-Mumbai, 2013.
13. Наследов, А. Д. *IMB SPSS Statistics 20 и AMOS: профессиональный статистический анализ данных* / А. Д. Наследов. – СПб.: Питер, 2013. – 416 с.
14. Методика проведения экспертизы сортів люпину білого, вузьколистого, жовтого (*Lupinus albus* L., *L. angustifolius* L., *L. luteus* L.) на відмінність, однорідність та стабільність Методики проведення експертизи сортів рослин групи кормових та коренеплідних на відмінність, однорідність і стабільність URL: <https://sops.gov.ua/uploads/page/5a5f1cc42cc8a.pdf>.