

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ БІЛІМ ЖӘНЕ ҒЫЛЫМ МИНИСТРЛІГІ
Л.Н. ГУМИЛЕВ АТЫНДАҒЫ ЕУРАЗИЯ ҰЛТТЫҚ УНИВЕРСИТЕТІ



ЖАС ҒАЛЫМДАР КЕҢЕСІ



Студенттер мен жас ғалымдардың
«ҒЫЛЫМ ЖӘНЕ БІЛІМ - 2016» атты
XI Халықаралық ғылыми конференциясының
БАЯНДАМАЛАР ЖИНАҒЫ

СБОРНИК МАТЕРИАЛОВ
XI Международной научной конференции
студентов и молодых ученых
«НАУКА И ОБРАЗОВАНИЕ - 2016»

PROCEEDINGS
of the XI International Scientific Conference
for students and young scholars
«SCIENCE AND EDUCATION - 2016»

2016 жыл 14 сәуір
Астана

**ҚАЗАҚСТАН РЕСПУБЛИКАСЫ БІЛІМ ЖӘНЕ ҒЫЛЫМ МИНИСТРЛІГІ
Л.Н. ГУМИЛЕВ АТЫНДАҒЫ ЕУРАЗИЯ ҰЛТТЫҚ УНИВЕРСИТЕТІ**

**Студенттер мен жас ғалымдардың
«Ғылым және білім - 2016»
атты XI Халықаралық ғылыми конференциясының
БАЯНДАМАЛАР ЖИНАҒЫ**

**СБОРНИК МАТЕРИАЛОВ
XI Международной научной конференции
студентов и молодых ученых
«Наука и образование - 2016»**

**PROCEEDINGS
of the XI International Scientific Conference
for students and young scholars
«Science and education - 2016»**

2016 жыл 14 сәуір

Астана

ӘӨЖ 001:37(063)

КБЖ 72:74

F 96

F96 «Ғылым және білім – 2016» атты студенттер мен жас ғалымдардың XI Халық. ғыл. конф. = XI Межд. науч. конф. студентов и молодых ученых «Наука и образование - 2016» = The XI International Scientific Conference for students and young scholars «Science and education - 2016» . – Астана: <http://www.enu.kz/ru/nauka/nauka-i-obrazovanie/>, 2016. – б. (қазақша, орысша, ағылшынша).

ISBN 978-9965-31-764-4

Жинаққа студенттердің, магистранттардың, докторанттардың және жас ғалымдардың жаратылыстану-техникалық және гуманитарлық ғылымдардың өзекті мәселелері бойынша баяндамалары енгізілген.

The proceedings are the papers of students, undergraduates, doctoral students and young researchers on topical issues of natural and technical sciences and humanities.

В сборник вошли доклады студентов, магистрантов, докторантов и молодых ученых по актуальным вопросам естественно-технических и гуманитарных наук.

ӘӨЖ 001:37(063)

КБЖ 72:74

ISBN 978-9965-31-764-4

©Л.Н. Гумилев атындағы Еуразия
ұлттық университеті, 2016

3. Каминский А. В., Страхов Ю. И., Трейгер Е. М. Анализ практики оценки недвижимости. Учебно – практическое пособие. М.: Международная академия оценки и консалтинга. 2005. – 238 с.
4. Ахметов Е. С. Оценка земли и недвижимости: Учебное пособие. – Алматы, ТОО «Эвро», 2012. – 150 с.
5. Постолит А. В. Visual Studio. NET: разработка приложений баз данных. – СПб.: Питер, 2003. – 544 с.
6. Пауэрс Л., Снелл М. Microsoft Visual Studio 2008. – Санкт - Петербург, 2009. – 1200 с.
7. Кригель А., Трухнов Б. SQL. 2 – е издание. Библия пользователя. – Москва – Санкт – Петербург – Киев, 2010.

УДК 614.2

ИСПОЛЬЗОВАНИЕ ЭНТРОПИИ ПРИ ПОСТРОЕНИИ ДЕРЕВА РЕШЕНИЙ

Маркабаева Сая Алпысбаевна

Магистрант кафедры Информационные системы

ЕНУ им. Л.Н.Гумилева, Астана, Казахстан

Научный руководитель – Т.Т.Оспанова

Компьютерная программа обучается по мере накопления опыта относительно некоторого класса T и целевой функции P , если качество решения этих задач (относительно P) улучшается с получением нового опыта. Так мы назовем «Машинное обучение». В «Машинное обучение» наряду с нейронными и Байесовскими сетями входит понятие, деревья принятия решений.

Дерево принятия решений – это дерево. На нем есть метки:

- В узлах, не являющиеся листьями: атрибуты, по которым различаются случаи;
- В листьях: значения целевой функции;
- На ребрах: значения атрибута, из которого исходит ребро.

Чтобы прогнозировать новый случай, нужно спуститься по дереву до листа и выдать соответствующее значение. Рассмотрим следующую задачу.

Постановка задачи. Выиграет ли «Зенит» свой следующий матч? В таблице 1 показаны следующие значения параметров:

1. Выше ли находится соперник по турнирной таблице;
2. Дома ли играется матч;
3. Пропускает ли матч кто-либо и лидеров команды;
4. Идет ли дождь.

Мы знаем об исходах нескольких матчей и хотим предсказать исход следующего матча, параметры которого нам еще не встречались.

Таблица 1. Как играет Зенит

Соперник	Играем	Лидеры	Дождь	Победа
Выше	Дома	На месте	Да	Нет
Выше	Дома	На месте	Нет	Да
Выше	Дома	Пропускают	Нет	Да
Ниже	Дома	Пропускают	Нет	Да
Ниже	В гостях	Пропускают	Нет	Нет
Ниже	Дома	Пропускают	Да	Да
Выше	В гостях	На месте	Да	Нет
Ниже	В гостях	На месте	Нет	???

Методы решения. Теперь построим дерево решений.

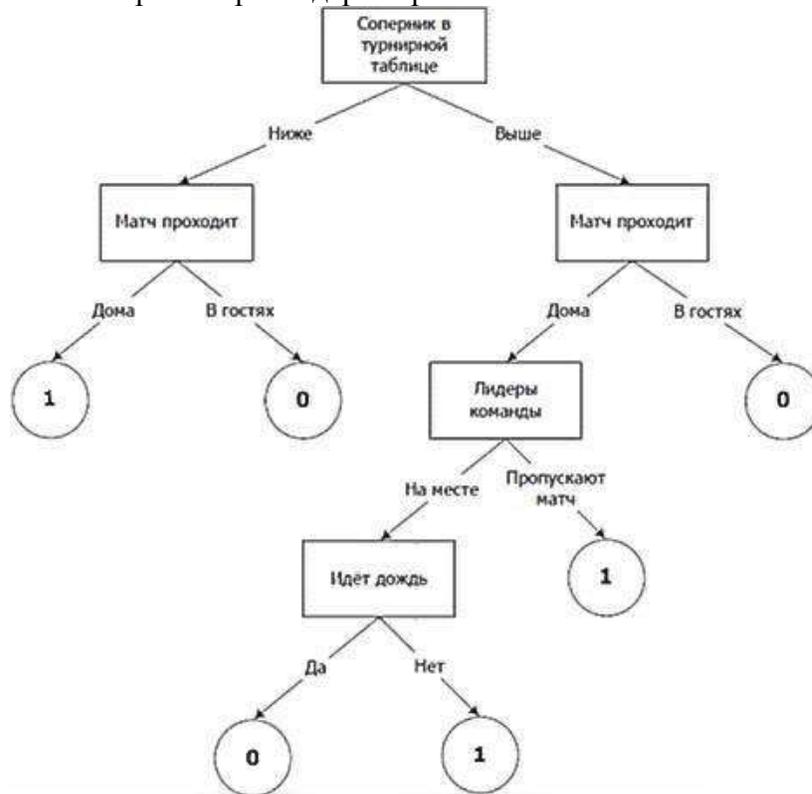


Рис. 1. Решающее дерево для прогнозирования игры в футбол

Спускаемся по дереву, выбирая нужные атрибуты, и получаем ответ: судя по нашему дереву Зенит этот матч должен проиграть.

Это у нас получилось большое дерево. А вот дерево для тех же самых данных, но куда меньше (Рис. 2).

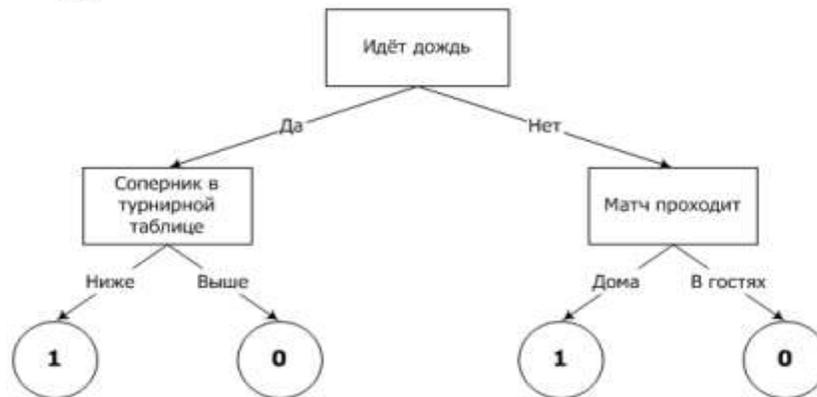


Рис. 2. Решающее дерево для прогнозирования игры в футбол

Поэтому, чтобы построить наиболее лучшее дерево необходимо ознакомиться с понятием энтропия. В общем виде постановка задачи для решающих деревьев выглядит следующим образом. Пусть дано множество объектов A (всего в A лежит N объектов, составляющих так называемую обучающую выборку), обладающих определенными независимыми характеристиками (атрибутами с конечным множеством значений; всего имеется $(M + 1)$ атрибутов). Множество первых M атрибутов обозначим как Q . Для заданного множества A

все $(M + 1)$ атрибутов известны. Для других (новых) элементов по известным первым M атрибутам требуется найти целевой $(M + 1)$ -ый атрибут. При этом на вход подается число N (элементов в обучающей выборке), число M .

Как правило, данный метод применяется для задач классификации и кластеризации. В данной работе предложен подход, который показывает способ применения данных деревьев к прогнозированию временных рядов.

Введём некоторые важные определения.

Определение 1. Энтропия $H(A, S) = -\sum_{i=1}^{S_n} \frac{|A_i|}{|A|} \log_2 \frac{|A_i|}{|A|}$, S - целевой атрибут, A_i - элементы из A , у которых атрибут S равен $i(a|A) = N$.

Определение 2. Прирост информации – определяется для каждого атрибута из Q по отношению целевому атрибуту S и показывает, какой из атрибутов Q дает максимальный прирост информации относительно значения атрибута S (т.е. относительно класса элемента). Прирост информации определяется по следующей формуле[1]:

$$Gain(A, Q) = H(A, S) - \sum_{i=1}^{Q_n} \frac{|A_i|}{N} H(A_i, S)$$

Теперь рассмотрим, как можно посчитать энтропию и прирост информации для нашей базы данных используя язык программирования C# [2].

Для работы используется программа Microsoft Visual Studio 2015. Создадим в ней класс ID3.

1. Создадим отдельный массив для атрибутов, отдельный двумерный массив для уникальных значений и двумерный массив для таблицы.

2. Создаем функцию для атрибутов в виде:

```
public void SetSName (int j, string name)
{
    SNames[j] = name;
}
```

j – номер столбца, $name$ – значение этих атрибутов.

3. Создадим функцию для нашей таблицы:

```
public void SetTable (int Line, int Attr, string value)
{
    STable[Line, Attr] = value;
}
```

$line$ - номер строки, $attr$ – столбец, $value$ – значение столбца

4. Создадим функцию для считывания одинаковых значений столбцов.

```
public void CalcSCount (int attr)
{
    int count = 1;
    SValue[0, attr] = STable[0, attr];
    for (int line = 1; line < N; line ++)
    {
        bool unique = true;
        for (int j = 0; j < line; j ++)
            if (STable [j, attr] == STable [line, attr])
            {
                unique = false;
            }
    }
}
```

```

        break;
    }
    if (unique)
    {
        SValue[count, attr] = STable[line, attr];
        count++;
    }
}
SCount [attr] = count;
}

```

5. Создадим функцию для считывания энтропии для всего атрибута:

```

public double Entropy (int attr)
{
    double sum = 0.0d;
    for (int q = 0; q < SCount[attr]; q++)
    {
        double ag = HowManyValues(attr, SValue[q, attr]) / (double)N;
        sum += ag * log2 (ag);
    }
    return -sum;
}

```

6. Создадим функцию для считывания энтропии для определенного атрибута:

```

public double Entropy (int attr, int i, int goal)
{
    double sum = 0.0d;
    int Ni = HowManyValues(attr, SValue [i, attr]);
    for (int q = 0; q < SCount[goal]; q++)
    {
        double ag = HowManyValues(attr, SValue[i, attr],
            goal, SValue[q, goal]) / (double)Ni;
        sum += ag * log2(ag);
    }
    return -sum;
}

```

7. Посчитаем прирост информации для атрибута Лидеры.

```

public double Gain (int attr, int goal)
{
    double e = Entropy(goal);
    int Qn = SCount [attr];
    double h = 0;

    for (int i = 0; i < Qn; i++)
    {
        int Ai = HowManyValues (attr, SValue [i, attr]);
        h += (Ai / (double) N) * Entropy(attr, i, goal);
    }
    return e - h;
}

```

Результаты работы. На рисунке 3 показано значение энтропии относительно целевого признака «Победа» и значение прироста информации для атрибута «Лидеры».

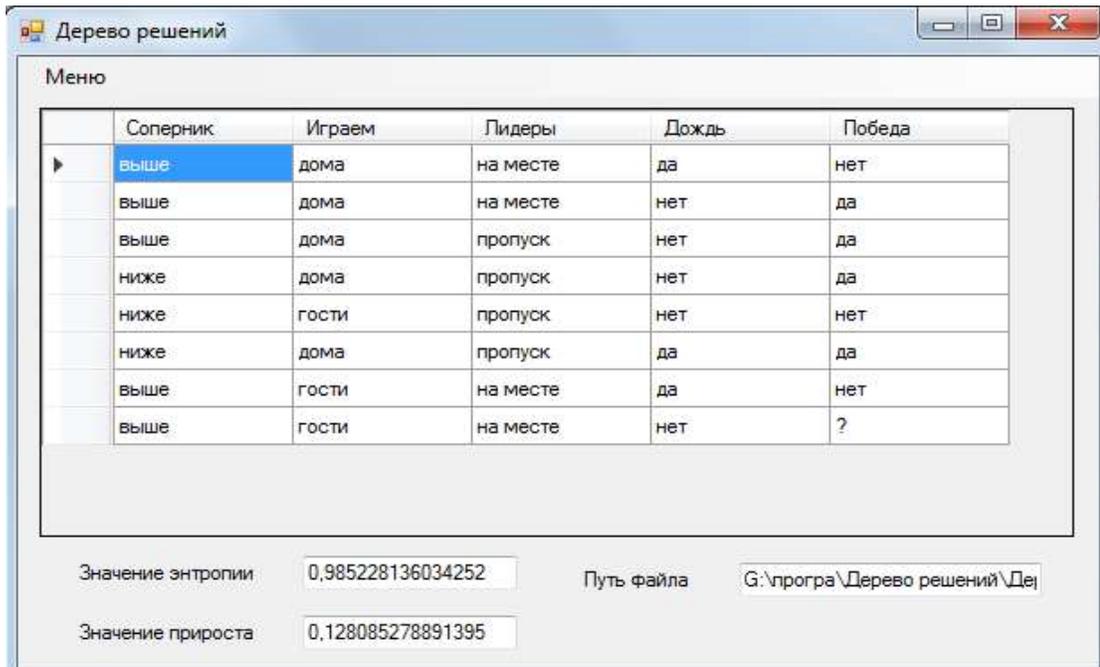


Рис. 3. Значение энтропии и прироста информации для атрибута «Лидеры»

И так далее, можем вычислить значение прироста информации для остальных атрибутов. Теперь можем строить дерево, выбирая на каждом этапе признак с максимальным приростом информации. В итоге получим следующее дерево (Рис. 4).

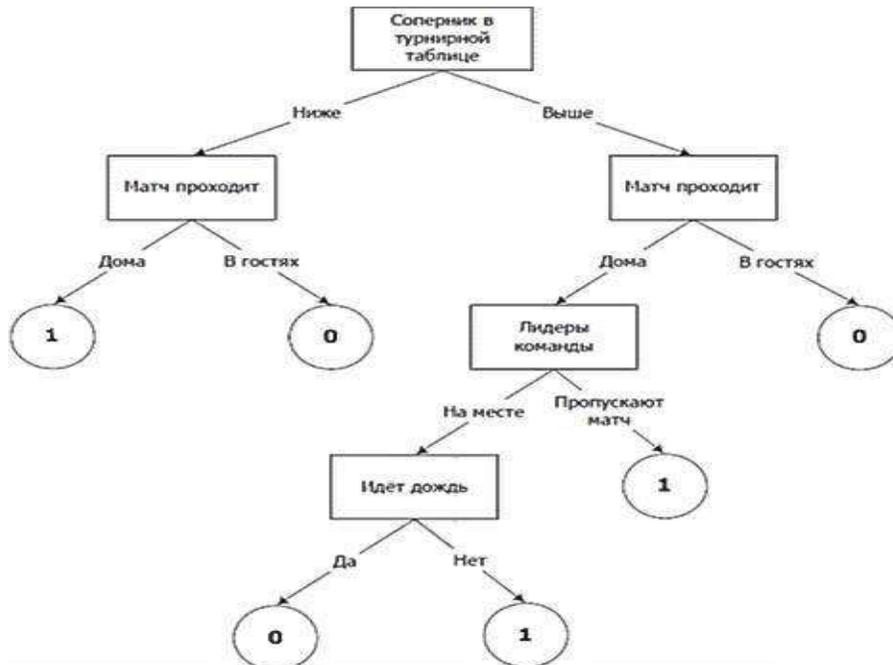


Рис. 4. Решающее дерево для прогнозирования игры в футбол

Итак чтобы получить оптимальное дерево решений сначала необходимо вычислить

энтропию для целевого атрибута и значение прироста информации для каждого атрибута отдельно.

Список использованных источников

1. Деревья принятия решений. Сергей Николенко. Машинное обучение – ИТМО, 2006. – Режим доступа: <http://www.amse.ru/archive/courses/2006/nikolenko/01-dectrees.pdf>
2. Уроки программирования с нуля С#. – Режим доступа: <http://mycsharp.ru/>

УДК 007.5

“ОҚЫТУШЫЛАРҒА АРНАЛҒАН ОҚУ КЕСТЕСІН ҚҰРУ” АВТОМАТТАНДЫРЫЛҒАН АҚПАРАТТЫҚ ЖҮЙЕСІН ҚҰРУ

Өтеген Сұңқар Дидарұлы

Л.Н.Гумилев атындағы ЕҰУ МАЖ-11 магистранты, Астана, Қазақстан
Научный руководитель – Садвакасова К.Ж.

Қазіргі уақытта қазақстандық оқу орындарындағы оқыту саласын кеңейту тенденциясы бақыланып отыр, ол оқытудың жаңа бағыттарының сәйкесінше факультеттердің, мамандықтардың және т.б. пайда болуына себепкер болып отыр. Алайда бұл үрдіс жаңа мүмкіндіктерді ашып қана қоймай, сонымен қатар, оқыту үрдісін жақсартудың жаңа мәселелерін ұсынады. Жоғарғы оқу орындарындағы оқу сапасын арттыру мәселесі оқыту үрдісін ұйымдастыруды тұрақты түрде жетілдіріп отыру қажеттілігімен шартталады және сонымен қатар, ең бірінші кезекте, негізгі ұйымдастыру мәселелерінің, соның ішінде, оқу кестесінің құрылымы мәселелерінің сапасын арттыру қажет.

Мамандықтарды, пәндерді, оқу топтарын саны жағынан арттыру жоғарғы оқу орындарындағы оқу кестесін құру жұмысын көп еңбекті қажет ететін үрдіске айналдырып отыр. Сонымен, берілген мәселенің ұйымдастыру шешімдерін іздеу туралы сұрақтар және осы үрдісті автоматтандыру бағдарламасы құрылды.

Сабақ кестесін жасау келесі тапсырмаларға бөлуге болады:

- 1 Негізгі анықтамалық енгізу;
- 2 Сабақтың жоспарын енгізу;
- 3 Шектеуді енгізу;
- 4 Сабақ кестесін құрастыру;



1-сурет.