Доклад

Теорема о прогнозе разнообразия

Беличева Дарья Михайловна

Содержание

# 1 Введение

**Цель**

Исследовать теорему о прогнозе разнообразия.

**Задачи**

* Описать понятие “мудрость толпы”;
* Дать теоретическое описание теоремы о прогнозе разнообразия;
* Привести пример, показывающий работу теоремы;
* Показать практическую реализацию расчетов по теореме о прогнозе разнообразия на языке программирования Julia.

**Актуальность**

При составлении численного прогноза, особенно в таких контекстах, как прогнозирование увеличения продаж, полезно объединить различные оценки. Коллективное понимание группы часто сравнивает ошибки и превосходит точку зрения только одного человека.

Коллективное понимание, или концепция “дебошир превосходит среднее значение”, предполагает, что объединенные суждения группы, как правило, более точны, чем индивидуальные оценки. Это происходит в первую очередь из-за баланса завышенных и заниженных оценок.

Теорема о прогнозе разнообразия применима к моделям, которые делают численные прогнозы или оценки. Она количественно оценивает влияние точности моделей и их разнообразия на точность их среднего [1].

# 2 Эксперимент Гальтона

В 1906 г. в городе Плимут (Великобритания) на сельской ярмарке был проведен эксперимент. Френсис Гальтон в качестве развлечения посетителей ярмарки предложил на глаз оценить вес выставленного на всеобщее обозрение быка и написать эту цифру на специальном билете. За правильные ответы организаторы шоу обещали призы. В результате в голосовании приняли участие около 800 человек — как заядлых фермеров, так и людей, далеких от скотоводческих дел. Собрав после этой ярмарки все билеты для анализа, Гальтон высчитал среднее арифметическое значение для всей выборки — 1197 фунтов. Реальный же вес быка оказался 1198 фунтов (543,4 кг). Каким-то непостижимым образом разношерстная публика дала ответ, максимально приближенный к реальному показателю. Т.е. ответ публики был точнее чем ответ отдельно взятого эксперта, например мясника или скотовода. Гальтон, который до этого свято верил в селекцию и превосходство одних людей над другими, был вынужден сменить вектор своих исследований.

С тех пор Френсис Гальтон изменил своё мнение по поводу мудрости толпы, он сказал:

“Результат кажется более достойным доверия к демократическому суждению, чем можно было ожидать.”

# 3 Мудрость толпы

Мудрость толпы: почему многие умнее немногих и как коллективная мудрость формирует бизнес, экономику, общества и нации, опубликованная в 2004 году, представляет собой книгу, написанную Джеймсом Суровецким об объединении информации в группах, что приводит к решениям, которые, по его утверждению, часто оказываются лучше, чем могли бы быть приняты любым отдельным членом группы. В книге представлены многочисленные тематические исследования и анекдоты, иллюстрирующие ее аргументацию, и затрагиваются несколько областей, в первую очередь экономика и психология[2].

Вступительный анекдот рассказывает об удивлении Фрэнсиса Гальтона по поводу того, что толпа на окружной ярмарке точно угадала вес быка, когда их индивидуальные предположения были усреднены (среднее значение было ближе к истинному весу разделанного быка, чем оценки большинства участников толпы)[3].

Мудрость толпы - это коллективное мнение группы людей, а не одного эксперта.

Исследования “Мудрости толпы” обычно объясняют превосходство средних значений толпы над индивидуальными суждениями устранением индивидуального шума - объяснение, предполагающее независимость индивидуальных суждений друг от друга. Таким образом, толпа склонна принимать наилучшие решения, если она состоит из различных мнений и идеологий.

Не все толпы (группы) обладают “мудростью”. Рассмотрим, например, толпы или обезумевших инвесторов в пузыре на фондовом рынке. Согласно Суровецкому, эти ключевые критерии отделяют мудрые толпы от иррациональных (см. в табл. 1):

Таблица 1: Пять элементов, необходимых для формирования мудрой толпы

| Критерии | Описание |
| --- | --- |
| Разнообразие мнений | У каждого человека должна быть личная информация, даже если это просто эксцентричная интерпретация известных фактов. |
| Независимость | Мнения людей не определяются мнениями окружающих. |
| Децентрализация | Люди могут специализироваться и использовать местные знания. |
| Агрегация | Существует некоторый механизм для превращения частных суждений в коллективное решение. |
| Доверие | Каждый человек верит, что коллективная группа справедлива. |

# 4 Теорема о прогнозе разнообразия

Скотт Э. Пейдж - американский социолог и заслуженный профессор Университета Джона Сили Брауна в области сложности, социальных наук и менеджмента в Мичиганском университете в Энн-Арборе. Он представил теорему прогнозирования разнообразия:

Квадрат ошибки коллективного прогнозирования равен среднему квадрату ошибки минус прогнозируемое разнообразие.

Следовательно, когда разнообразие в группе велико, ошибка толпы невелика.

*Определения*

* Средняя индивидуальная ошибка: среднее значение индивидуальных квадратов ошибок
* Коллективная ошибка: квадратическая ошибка коллективного прогнозирования
* Разнообразие прогнозов: среднее квадратическое расстояние от индивидуальных прогнозов до коллективного прогноза

Теорема прогнозирования разнообразия: Учитывая множество прогностических моделей, имеем

Коллективная ошибка = Средняя индивидуальная ошибка - Разнообразие прогнозирования

* — предсказание толпы (общая оценка параметра).
* — истинное значение параметра.
* — индивидуальные оценки параметра.

# 5 Пример работы теоремы

Допустим у нас есть трое людей (Катя, Костя и Мариян), которые предсказывают, сколько людей придет в закусочную на обед в определенный день (см. в табл. 2). Реальное число людей, пришедших на обед:

Таблица 2: Результаты предсказаний людей

| Человек | Предсказание |
| --- | --- |
| Катя | 10 |
| Костя | 16 |
| Мариян | 25 |

Чтобы получить среднюю оценку предсказаний, надо их сложить и поделить на количество:

Теперь давайте вычислим квадратичную ошибку каждого человека (см. в табл. 3):

Таблица 3: Квадратичная ошибка каждого человека

| Человек | Ошибка |
| --- | --- |
| Катя |  |
| Костя |  |
| Мариян |  |

Тогда средняя ошибка каждого человека составляет:

Давайте посмотрим, насколько точной была оценка толпы, посчитав ее ошибку:

То есть толпа ошиблась всего на 1. Таким образом, мы получаем мудрость толпы.

Чтобы понять, почему это имеет смысл, рассмотрим разнообразие (вариации в предсказаниях). Мы посмотрим на предсказание каждого человека и его отдаленность от среднего прогноза (см. в табл. 4).

Таблица 4: Отдаленность предсказания каждого человека от среднего прогноза

| Человек | Разнообразие |
| --- | --- |
| Катя |  |
| Костя |  |
| Мариян |  |

Тогда разнообразие предсказаний составляет:

Получается, что

то есть

Таким образом, мы пришли к теореме о прогнозе разнообразия (Diversity Prediction Theorem):

# 6 Практическая реализация

Подобные расчеты удобно выполнять на языке программирования Julia. Julia — высокоуровневый свободный язык программирования с динамической типизацией, созданный для математических вычислений. Будем использовать пакет Statistics, который содержит базовые функции статистики. В нем есть функция mean(), которая нужна для вычисления среднего значения для всех элементов в коллекции.

Функция для вычисления элементов теоремы о прогнозе разнообразия (средняя индивидуальная ошибка, коллективная ошибка, разнообразие прогнозов) будет выглядеть следующим образом:

function diversitytheorem(truth::T, pred::Vector{T}) where T<:Number  
 μ = mean(pred)  
 avgerr = mean((pred .- truth) .^ 2)  
 crderr = (μ - truth) ^ 2  
 divers = mean((pred .- μ) .^ 2)  
 avgerr, crderr, divers  
end  
  
for (t, s) in [(18, [10, 16, 25])]  
 avgerr, crderr, divers = diversitytheorem(t, s)  
 println("""  
 average-error : $avgerr  
 crowd-error : $crderr  
 diversity : $divers  
 """)  
end

В результате получаем:

average-error : 39.0  
crowd-error : 1.0  
diversity : 38.0

# 7 Выводы

Была исследована теорема о прогнозе разнообразия.

# Список литературы

1. Пейдж С. Модельное мышление. Как анализировать сложные явления с помощью математических моделей. 2015. 528 с.

2. The Wisdom of Crowds [Электронный ресурс]. URL: <https://en.wikipedia.org/wiki/The_Wisdom_of_Crowds#cite_note-1>.

3. Surowiecki J. The Wisdom of Crowds. Doubleday, Anchor, 2004. 336 с.