16 колонок

7-ой (по очереди, начиная с первого) столбец – выходные данные (статус игры)

Остальные столбцы – входные данные

20058 значений

# Информация по столбцам

1. Номер игры
2. Оценённость (true(81%)/false(19%))
3. Время начала
4. Время конца
5. Количество ходов (1-349)
6. Статус игры (resign(56%)/mate(32%)/other(13%))
7. Победитель (white(50%)/black(45%)/other(5%))
8. Приращение времени (10+0(38%)/15+0(7%)/other(55%))
9. ID игрока белыми фигурами
10. Рейтинг игрока белыми фигурами (784-2700)
11. ID игрока чёрными фигурами
12. Рейтинг игрока чёрными фигурами (789-2723)
13. Все ходы в стандартной шахматной записи (уникальные значения)
14. Эко открытия (А00(5%)/С00(4%)/other(91%))
15. Открытое имя/имя позиции (Van't Kruijs Opening(2%)/Sicilian Defense(2%)/Other(96%))
16. Начальный слой (количество ходов в начальной фазе) (1-28)

# Общая информация по датасету

Главная информация

Это набор из чуть более 20 000 игр, собранных от избранных пользователей на сайте Lichess.org, и как собрать еще. Я также буду загружать больше игр в будущем по мере их сбора. Этот набор содержит:

* ID игры;
* Номинальный (T / F);
* Время начала;
* Время окончания;
* Количество ходов;
* Статус игры;
* Победитель;
* Приращение времени;
* Идентификатор белого игрока;
* Рейтинг белых игроков;
* ID черного игрока;
* Рейтинг черных игроков;
* Все ходы в стандартной шахматной записи;
* Эко открытия (стандартизированный код для любого открытия, список здесь);
* Открытие имени;
* Начальный слой (количество ходов в начальной фазе)

Для каждой из этих отдельных игр от Lichess. Я собрал эти данные с помощью Lichess API, который позволяет собирать историю игр любого пользователя. Сложной частью был сбор имен пользователей для использования, однако API также позволяет выгружать всех пользователей в команде Lichess. На Lichess есть несколько команд с более чем 1500 игроками, так что это оказалось эффективным способом побудить пользователей собирать игры.

Возможное использование

В одной шахматной партии содержится много информации, не говоря уже о полном наборе данных из нескольких партий. Это в первую очередь игра в шаблоны, а наука о данных - это все, чтобы выявлять закономерности в данных, поэтому в прошлом шахматы были одной из самых популярных областей ИИ. Этот набор данных собирает всю информацию, доступную по 20 000 игр, и представляет ее в формате, который легко обрабатывать для анализа, например, что позволяет игроку выиграть черным или белым цветом, сколько мета (вне игры) факторы влияют на игру, соотношение дебютов и победы черных и белых и многое другое.