

венство $b_{c_i} < 0$ (для решаемой практической задачи такая ситуация может возникнуть в случае недостаточного количества данных для вычисления параметра b_c). В этом случае неравенство (2) будет иметь вид $0 \leq L_{c_i} - y_{c_i} \leq p_{c_i} \leq U_{c_i}$. Для решения такой задачи предлагается перейти к задаче поиска приближенного решения. Основная идея состоит в следующем. Пусть $y_i = x_i^2$. Без ограничения общности рассуждений полагаем $0 \leq x_i \leq 1$. Тогда $y_i \leq x_i$, $y_i \geq 0$, $y_i \geq 2x_i - 1$ (т.к. $(x_i - 1)^2 \geq 0$), т.е. вместо x_i^2 можно использовать параметр y_i и дополнительные ограничения (рисунок 1). Для получения более точного результата можно увеличить количество ограничивающих прямых.

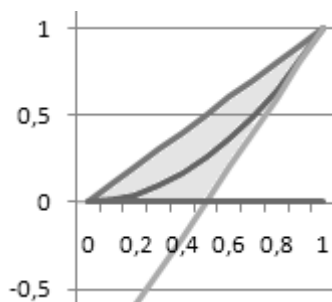


Рис. 1

Список использованных источников:

1. Химмельблау, Д. Прикладное нелинейное программирование. – М.: Мир, 1975.
2. IBM ILOG CP Optimizer V2.3. User's Manual. – 252 p.

ПОДХОД К ОПТИМИЗАЦИИ ХРАНЕНИЯ И ОБРАБОТКИ ДАННЫХ В СИСТЕМАХ СБОРА СТАТИСТИКИ

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Курдесов Н.Н.

Поддубная О.Н. доцент,
кандидат физико-математических наук

Вместе со стремительным накоплением информации развиваются и технологии анализа данных. Увеличивается количество областей, где работа с качественно большими объемами информации становится более актуальной. Постоянно происходит увеличение скорости потока данных в организационный процесс: экономика, банковская деятельность, маркетинг, веб-аналитика, и др. В современных системах хранения и обработки больших объемов информации одной из главных задач является обеспечение возможности быстрой выборки требуемых данных.

При выборе способа хранения большого объема информации необходимо учитывать специфику её выборки. Одной из черт в системах сбора статистики является то, что накапливаемые данные распределяются последовательно во времени, а выборка, как правило, делается по объектам сбора этих данных. Например: приложение собирает информацию о работе некоторого количества компьютеров в сети. Обработанные данные от каждого из компьютеров будут последовательно во времени записываться на диск. На данный момент для хранения больших объемов данных используются накопители на жестких магнитных дисках. Их особенностью является то, что скорость последовательного чтения/записи может быть в несколько раз выше, чем скорость случайного чтения/записи. При выборе информации конкретного объекта (в нашем случае статистики работы компьютера) время выборки будет сопоставимо с временем случайного чтения.

На рисунке 1 проиллюстрировано распределение данных на диске, накапливаемых последовательно во времени:

Данная проблема остаётся актуальной даже при использовании индексов в современных СУБД, вроде MySQL или Postgres. Они ускоряют поиск информации, расположенной на диске, но не могут ускорить её чтение.

Решением данной проблемы является подход перераспределения данных на диске таким образом, чтобы информация, которая в последствии будет использоваться для выборки, располагалась на нём последовательно.

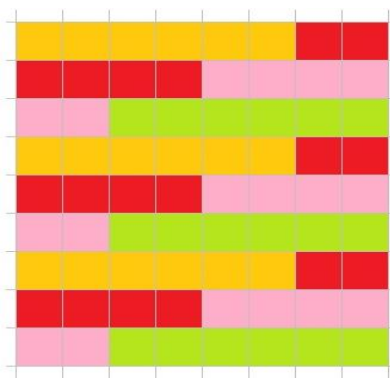


Рис. 1 – Распределение накапливаемых данных

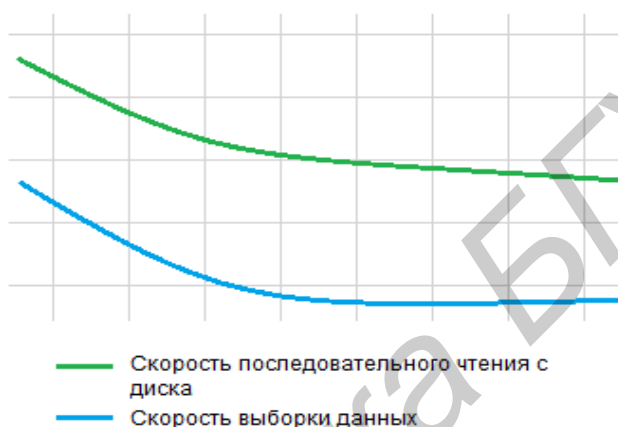


Рис. 2 – Сравнение скоростей линейного чтения с диска и выборки данных до оптимизации

Одним из самых простых способов является алгоритм пообъектного перебора с последующим чтением/записью его информации. В общем случае алгоритм в упрощенном виде выглядит следующим образом:

```
for obj in objects: # перебираем объекты
    msgs = obj.msgs() # получаем все сообщения объекта из БД
    for msg in msgs:
        obj.remove(msg) # удаляем сообщения от объекта из БД
    for msg in msgs:
        obj.add(msg) # записываем сообщения, на диске при этом они будут записаны последовательно
```

На рисунке 3 показаны данные на диске, после их перераспределения:

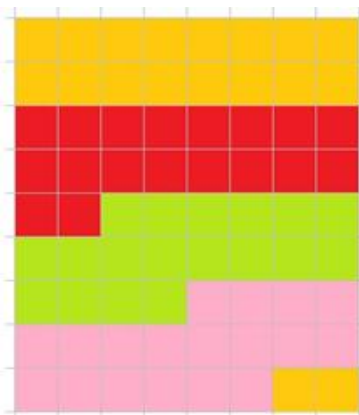


Рис. 3 – Перераспределение данных на диске для оптимизации выборки

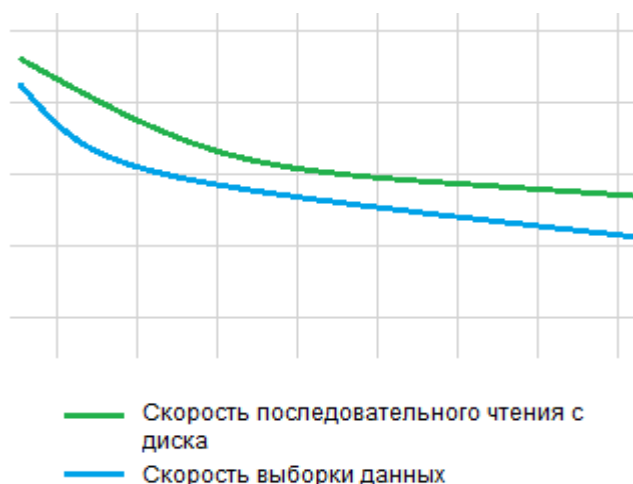


Рис. 4 – Сравнение скоростей линейного чтения с диска и выборки данных после оптимизации

Это решение позволяет достигать скорости выборки данных сопоставимой со скоростью линейного чтения, и не зависит от используемых СУБД.

Список использованных источников:

1. Моррисон, Алан. Большие Данные: как извлечь из них информацию/ АланМоррисон // Ежеквартальный журнал–Москва, 2010. – 31 с.
2. Свободная общедоступная многоязычная универсальная энциклопедия [Электронный ресурс], http://en.wikipedia.org/wiki/Hard_disk_drive
3. Свободная общедоступная многоязычная универсальная энциклопедия [Электронный ресурс]: <https://ru.wikipedia.org/wiki/IOPS>.

МАКСИМИЗАЦИЯ ВЛИЯНИЯ В СОЦИАЛЬНЫХ СЕТЯХ

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Лычковский А. В.

Волорова Н. А. – к-т. техн. наук, доцент

Социальная сеть отношений между людьми играет важную роль в распространении информации, идей и влияния среди ее членов. Новое мнение или инновация, которые появляются в мире, могут стать невостребованными или внести значительный вклад в развитие общества. Для того, чтобы понять степень принятия новых идей, нужно понять какова динамика их распространения в социальной сети. Эта динамика, в свою очередь, определяется степенью влияния одного объекта на другой, так что решение одного объекта приводит к принятию другим схожего решения.

Процессы, происходящие в социальных сетях, имеют долгую историю исследования. Некоторые из ранних исследований касались использования новых медицинских препаратов и инноваций в сфере сельского хозяйства [1, 2]. Также исследовались проблемы распространения информации о новом продукте при использовании «вирусной рекламы» [3,4,5]. Схожие процессы могут быть найдены и в инженерных сетях, например, проблема каскадного отключения энергетических систем [6].

Последние работы в этой области, касаются проблемы маркетинга [5]. Предположим, что мы имеем данные о некоторой сети, включая оценку степени влияния одного объекта на другой. В этой сети мы хотим рекламировать новый продукт таким образом, чтобы максимальная ее часть приняла продукт. Для этого можно воспользоваться «вирусной рекламой». Ее суть в том, чтобы вовлечь некоторое начальное множество «влиятельных» объектов из социальной сети, продемонстрировать им продукт, например, раздавая бесплатные образцы. Далее ожидается, что эти объекты будут делиться своим мнением с своими друзьями, потом друзья со своими друзьями и так далее. При такой формулировке возникает вопрос, как выбрать начальное множество «влиятельных» объектов таким образом, чтобы как можно большая часть социальной сети приобрела продукт.

Математическая модель описанной выше задачи может быть представлена в следующем виде. Имеется ориентированный взвешенный граф $G = (V, E)$ и модель M процесса распространения идеи в G . Для любого подмножества $A \subseteq V$ пускай $\delta(A)$ определяет ожидаемое число объектов, которые приняли идею, если A было начальным множеством. Таким образом входными значениями для задачи будут являться граф $G = (V, E)$ и натуральное число K , которое определяет максимальное число элементов в