МАТЕМАТИЧЕСКИЕ МОДЕЛИ РАСПРЕДЕЛЁННЫХ ЦЕНТРОВ ОБСЛУЖИВАНИЯ ВЫЗОВОВ

Цыганков Н.И., Росляков А.В.
ПГАТИ, г. Самара
E-mail: rosl@infosfera.ru, euss@com.syzran.ru

Под распределённым ЦОВ (центром обслуживания вызовов) понимается совокупность территориально разнесенных и информационно связанных между собой региональных ЦОВ, предоставляющих услуги клиентам с помощью операторов или автоматически и использующих единые правила обработки и маршрутизации вызовов.

Распределённые центры могут иметь следующую архитектуру:

- наличие в сети центрального ЦОВ, в котором хранится база данных обо всех ЦОВ сети (только общая информация, т.к. не целесообразно чтобы при обслуживании каждого вызова запрашивалась информация из центрального ЦОВ);
- все ЦОВ обладают равными приоритетами.

Рассматривая распределённые ЦОВ необходимо отметить, существование централизованных ЦОВ. Принципиальным отличием распределённого ЦОВ от централизованного является возможность, в некоторых случаях, более экономично использовать ресурсы центра обслуживания вызовов, решается вопрос с нагрузкой в ЧНН (вызовы распределяются равномерно среди всех ЦОВ входящих в сеть).

Если, по причине аварийной ситуации один из ЦОВ находящийся в сети не может принимать вызовы, то вызовы обслуживаются другими ЦОВ, входящими в сеть. Достоинством использования распределённых ЦОВ является эффективное использование интегрированного решения сетевых ЦОВ с Интеллектуальными сетями.

Всем вышеперечисленным и объясняется принципиальные отличия распределённых ЦОВ от централизованных, которые необходимо отразить в математических моделях схем функционирования распределённого ЦОВ.

При рассмотрении работы распределённого ЦОВ можно выделить 3 варианта поступлении информации о вызове:

- Полная информация о вызывающем абоненте. Это может быть наличие/отсутствие приоритета, возможность обслуживания в кредит, набор междугороднего/международного номера и т.д.
- Полное отсутствие какой либо информации. Т.е. ЦОВ «не знает» пред. историю звонков абонента, отсутствует информация в базе данных ЦОВ об абонентах.
- Поступление не полной информации о звонке. Причём полнота информации подчиняется рандомизированной (случайной) величине.

Последний пункт является наиболее важным и интересным из приведённых выше. Рассмотреть его можно с помощью Байесовской модели обработки нечисловой, неточной и неполной информации о вызове. Одной из возможностей Байесовской модели это решение задачи определения вероятности события (в данном случае: насколько полная будет информация о звонке). Это решение описывается формулой полной вероятности:

$$P(p(B);w)=\sum_{i=1}^{m} w_i p(B/A_i),$$

где $p(B/A_i)$- вероятность события «B» при условии, что произошло альтернативное событие: $A_i(A_i \cap A_j = 0, i \neq j, A_1 \cup \ldots \cup A_m = \Omega)$. При этом весовой коэффициент w_i, с помощью которого и описываются данные звонка, есть вероятность i- той альтернативы A_i.

Главной задачей при этом является нахождение числовой информации о весовых коэффициентах: $w_1 \ldots w_n$, каждый из которых есть – часть информации, которая поступает вместе с вызовом. [1]

Пусть x_i – *вызовы, которые поступают в распределённый ЦОВ.*. Предполагается, что они одинаково распределены, причем это распределение зависит от некоторого фиксированного, но неизвестного параметра θ, такого что средним значением x_i для всех i является некоторая функция от θ, обозначаемая через $m(\theta)$. Целью является нахождение зависимости $E(m(\theta)|x)$, которая описывает достоверность получения информации о вызове, где $x = x_1, x_2 \ldots x_n$ есть совокупность значений x_i. Получаем выражение:

$$E[m(\theta)|X] = Zx + (1-Z)E[m(\theta)], \qquad \text{где } Z = \frac{n}{n + \dfrac{E[w^2(\theta)]}{V[m(\theta)]}}$$

Необходимо получить выражения для $E[m(\theta)], V[m(\theta)], E[w^2(\theta)]$.Допустим, что имеем некоторое количество «*N*» *с полной информацией о вызове поступивших за некоторое время* «*n*». Допустим, что это количество «N» одинаково распределенного за время «n», зависит от неизвестного параметра θ_i для конкретного вызова, $i = 1,2,...N$. X_{ij} - это то количество вызовов, на которые распространяется правило неточной и неполной информации, тогда:

$$E[m(\theta)] = \frac{\sum_{i=1}^{N}\sum_{j=1}^{n} x_{ij}}{Nn} = \overline{X},$$

$$E[w^2(\theta)] = \frac{1}{N}\left(\sum_{i=1}^{N}\frac{\sum_{j=1}^{n}\left(x_{ij} - \overline{x}_i\right)^2}{n-1}\right),$$

$$V[m(\theta)] = \frac{\sum_{i=1}^{N}\left(\overline{x_i - x}\right)^2}{N-1} - \frac{1}{n}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\sum_{j=1}^{n}\left(x_{ij} - \overline{x}_i\right)^2}{n-1}\right), \qquad \text{где } \overline{x}_i = \left(\sum_{j=1}^{n} x_{ij}\right)/n.$$

Вычислим весовые коэффициенты: $w_1 ... w_n$, каждый из которых есть – часть информации, которая поступает вместе с вызовом.

Чтобы вычислить аппроксимацию функции $E(m(\theta|x))$, необходимо рассмотреть лишь функции, имеющие вид:

$$w_1 + w_2\overline{x} = w_1 + w_2\sum_{j=1}^{n}\frac{x_j}{n},$$

где w_1, w_2- весовые коэффициенты, которые следует определить. Для вычисления коэффициентов используется выражение:

$$E\left[\left(E[m(\theta)\overline{x}] - w_1 - w_2\overline{x}\right)^2\right] \quad [1]$$

Форма, в которой будет решаться задача, такова: найти коэффициенты, минимизирующие выражение:

$$E\left[\left(m(\theta) - w_1 - w_2\overline{x}\right)^2\right] \quad [2]$$

Можно показать, что выражения $[1]$ и $[2]$ эквивалентны. Сделать это можно следующим образом:

$$E\left[\left(E[m(\theta)\overline{x}] - w_1 - w_2\overline{x}\right)^2\right] = E\left[\left[m(\theta) - E[m(\theta)|x] + \left(E[m(\theta)|x] - w_1 - w_2x\right)\right]^2\right] = E[A^2] + 2E[AB] + E[B^2]$$

$$A = m(\theta) - E[m(\theta)|x],$$

$$B = E[m(\theta)|x] - w_1 - w_2\overline{x}.$$

Заметим, теперь, что $E[AB]$ равно нулю, поскольку:

$$E[AB] = E[E[AB|x]] = E[BE[A|x]]$$

Используя тот факт, что все члены в B - функция X:

$$E[A|x] = E[m(\theta)|x] - E[E[m(\theta)|x]x] = E[m(\theta)|x] - E[m(\theta)|x] = 0$$

Следовательно,

$$E\left[\left(m(\theta) - w_1 - w_2x\right)^2\right] = E\left[\left(E[m(\theta)|x] - w_1 - w_2\overline{x}\right)^2\right] \quad [3]$$

Член A не содержит w_1 или w_2, поэтому из формулы $[3]$ ясно, что если w_1 или w_2 являются решением задачи $[2]$, то они являются также решениями $[1]$ и наоборот.

Решение можно найти, дифференцируя математическое ожидание выражения $[2]$ по w_1 и w_2 и приравнивая обе частные производные к нулю:

$$E\left[m(\theta) - w_1 - w_2\overline{x}\right] = 0,$$

$$E\left[\bar{x}\left(m(\theta)-w_1-w_2\bar{x}\right)\right]=0.$$

Откуда

$$w_1 = E[m(\theta)] - w_2 E\left[\bar{x}\right]$$

$$w_2 = \left[E\left[\bar{x}m(\theta)\right]\right] - E\left[m(\theta)E\left[\bar{x}\right]\right] / \left[E\left[\bar{x}^2\right] - E^2\left[\bar{x}\right]\right]$$

Можно предложить использование комбинирования моделей, т.к. в комплексе эти модели будут решать большее количество вопросов связанных с упрощением работы распределённого ЦОВ, первая из которых это модель**:** *M*/*M*/*N*. [2] Если рассматривать модель с несколькими ЦОВ, соединенных в сеть, то при поступлении вызова в один из ЦОВ, он параллельно поступает на обслуживание в другой ЦОВ. Вызов обслуживается одним из ЦОВ, то соответственно второму в который параллельно пришёл этот же вызов приходит информация об отказе в обслуживании. С помощью этой модели можно стохастически минимизировать количество вызовов в системе. В этой модели, характерно то, что для всех ЦОВ правило маршрутизации заявок, основанное на наименьшей длине очереди, является оптимальным. Недостатком этой модели является то, что нет возможности сразу, при поступлении вызова маршрутизировать вызов в нужный ЦОВ – на основании информации о среднем времени ожидания и длине очереди это могло бы существенно повысить эффективность работы распределённых ЦОВ. Нет приоритетов заявок. Все параметры ЦОВ идентичны. Модель не описывает нетерпеливые звонки. Вторая модель, представляет собой диффузионную модель для решения проблемы организации маршрутизации вызовов. [3] При этом вводится аппроксимация к центру с высокой загрузкой и большим количеством операторов. Система содержит несколько классов заявок и большое количество операторов, работающих параллельно и объединенных в группы. Проблема возникает в том, вызов какого класса адресовать данному оператору в конкретный момент времени. Интенсивность обслуживания зависит как от класса вызова, так и от оператора, где она обслуживается. Учитывается человеческий фактор. Недостатком этой модели является то, что модель хорошо описывает лишь такие ЦОВ, в которых нагрузка сопоставима с количеством операторов, т.е. при пиковой нагрузке – поэтому все результаты являются аппроксимационными.

Литература

1. Хованов Н.В. Анализ и синтез показателей при информационном дефиците. СПб., 1996.
2. Hordijk A., Koole G. On the shortest queue policy for the tandem parallel queue // Probability in the engineering and informational sciences. – 1992. – №6 – P.63-79.
3. Atar R. – A diffusion model of scheduling control in queueing systems with many servers // To appear in the Annals of probability. – 2004.