**Сопоставление записей реляционных баз данных с использованием многоагентных систем.**

**Введение**

В работе рассматривается решение проблемы сопоставления записей реляционных баз данных, входящих в состав различных автоматизированных систем, имеющих одинаковое смысловое наполнение, но разные системы классификации и кодирования понятий. Решение задачи позволит свести к минимуму ручной труд оператора по сопоставлению понятий.

1. **Постановка задачи**

Каждую запись таблицы реляционной базы данных можно рассматривать как идентификатор (код, первичный ключ) и кортеж данных. Задача оператора при отождествлении понятий состоит в сравнении кортежей данных из двух таблиц и определении их совпадения.

Следующие рассуждения будут строиться на примере сведения объектов учета по ООСР.

Математически разность между двумя кортежами можно выразить как сумму разностей между каждой характеристикой в этом кортеже. Для объектов учета нескольких автоматизированных систем (АС) возможны два вида характеристик – числовые и текстовые.

Числовые характеристики сравниваются строго или нестрого. Строгое сравнение числовых характеристик подразумевает точное равенство (например, протяженность дороги в каждой из АС равняется 2456 км). Нестрогое сравнение числовых характеристик допускает попадание значения в определенный промежуток (координата Х объекта может лежать в пределах от 53,43 до 53,45). Для дальнейшего использования результатов сравнения примем, что равенство величин дает вес ребра равным 0, а неравенство – 1.

Сравнение текстовых характеристик может быть только нестрогим и основанным на вычисление расстояния Дамерау-Левенштейна. Расстояние Дамерау-Левенштейна – это мера разницы двух строк символов, определяемая как минимальное количество операций вставки, удаления, замены и транспозиции (перестановки двух соседних символов), необходимых для перевода одной строки в другую. Расстояние Дамерау-Левенштейна – целое число, для использования в нашей задаче разделим полученное расстояние на длину исходного слова. Получившееся дробное число и будет весом ребра между характеристиками.

Таким образом мы получим набор натуральных чисел – разницу между каждой характеристикой из сравниваемых кортежей. Просуммировав эти числа мы получим общее редакционное расстояние между кортежами.

В итоге, с математической точки зрения, мы получим граф с двумя наборами вершин и ребрами между вершинами. В получившемся графе необходимо наилучшим образом сопоставить точки из правого набора точкам левого набора, что приводит нас к решению задачи о назначениях.

Задача о назначениях – задача о наилучшем распределении некоторого числа работ между таким же числом исполнителей при условии взаимно однозначного соответствия между множествами работ и исполнителей. При отсутствии однозначного соответствия между множествами (одно множество больше другого) мы дополним меньшее множество, а веса ребер идущих к этим вершинам определим равным бесконечности. Для решения задачи необходимо найти оптимальное назначение из условия минимума общих затрат, которые равны сумме затрат исполнителей.

В математической модели задача представляется в виде двудольного графа, разбитого на два подмножества вершин X и Y одинаковой мощности n и множеством ребер U, соединяющих вершины из разных подмножеств. Информация о графе хранится в матрице чисел Dij, где i, j ∈ 1, 2,…, n, представляющих собой затраты на выполнение j-й работы i-м исполнителем. Требуется найти перестановку φ из элементов множества X, такую, что значение ЦФ равно:



Существует несколько способов решения задачи о назначениях. Рассмотрим решение задачи на основе муравьиных алгоритмов.

1. **Общие положения муравьиного алгоритма**

Основной идеей данного алгоритма является моделирование поведения муравьев, коллективной адаптации. Колония представляет собой систему с простыми правилами автономного поведения особей. Несмотря на примитивность поведения каждого отдельного муравья, поведение всей колонии оказывается достаточно разумным [МакКоннелл, 2004]. Таким образом, основой поведения муравьиной колонии служит низкоуровневое взаимодействие, благодаря которому, в целом, колония представляет собой разумную многоагентную систему. Взаимодействие осуществляется с помощью специального химического вещества – феромона, откладываемого муравьями на пройденном пути. При выборе направления движения муравей исходит не только из желания пройти кратчайший путь, но и из опыта других муравьев, информацию о котором получает непосредственно через уровень феромонов на каждом пути. Итак, концентрация феромона определяет желание особи выбрать тот или иной путь. Однако при таком подходе неизбежно попадание в локальный оптимум. Эта проблема решается благодаря испарению феромонов, которое является отрицательной обратной связью.

1. **Муравьиный алгоритм для решения задачи о назначениях**

Определим свойства муравья.

Каждый агент обладает собственной «памятью», в котором будет храниться список работ Jk, которые необходимо распределить муравью k.

Агенты обладают «зрением», обратно пропорциональным длине стоимости работы (длина ребра):

ηij = 1/Dij.

Каждый агент способен улавливать след феромона, который будет определять желание агента пройти по данному ребру, т.е. выбрать соответствующее ребру назначение. Уровень феромона в момент времени t на ребре Dij будет соответствовать τij(t).

Вероятность перехода муравья из вершины i в вершину j будет определяться следующим соотношением [Bonavear et al., 1999]:

, (1)

где α, β – параметры, задающие веса следа феромона. Они определяют «жадность» муравья. При α=0 муравей стремиться выбирать кратчайшее ребро, при β=0 – ребро с наибольшим количеством феромона. Рекомендуемые значения, полученные на основе экспериментальных исследований, варьируют от 1 до 5.

С течением времени желание, а значит и вероятность, выбора более оптимального назначения увеличивается, поскольку количество откладываемого феромона обратно пропорционально целевой функции и задается в следующем виде:

, (2)

где Q – параметр, имеющий значение порядка целевой функции оптимального назначения (задается лицом, принимающим решения), Lk(t) – целевая функция Tk(t).

Испарение феромона определяется следующим выражением:

, (3)

где m – количество муравьев, p – коэффициент испарения (0 ≤ p ≤ 1), определяющий долю оставшихся феромонов после каждой итерации.

В простом муравьином алгоритме начальное расположение колонии муравьев определяется следующим образом. Количество муравьев равно числу вершин в графе, и каждому муравью соответствует некоторая вершина. Однако в данном алгоритме все агенты изначально находятся в вершине x1(X – множество исполнителей). Размер колонии не ограничен в данной модели.

Приведем псевдокод муравьиного алгоритма для решения задачи о назначениях.

1. Ввод матрицы расстояний D.

2. Инициализация параметров алгоритма – α, β, Q.

3. Инициализация видимости ηij и начальной концентрации феромона.

4. Размещение муравьев в вершину x1.

5. Выбор начального назначения и определение L\*.

6. t = 0.

7. t=t+1.

8. k = 0.

9. k = k+1.

10. Реализовать назначения Tk(t) на основе выражения (1) и ЦФ Lk(t).

11. Если Lk(t) < L\*, то L\*= Lk(t), Т\* = Tk(t).

12. Если k < m, то перейти к п. 9.

13. Обновить следы феромона на всех ребрах на основе выражения (3).

14. Если t < T, то перейти к п. 7.

15. Вывод назначений T\* и его ЦФ L\*.

Временная сложность данного алгоритма зависит от времени жизни колонии, количества вершин графа и числа муравьев – O(t\*n2\*m).

**Заключение**

Экспериментальные исследования по решению задачи о назначениях с помощью муравьиного алгоритма проводилась различными группами исследователей. Их результаты показывают, что для графов с размерностью до 500 вершин «хорошее» решение находится за время порядка 3 секунд. Таким образом, применение этого алгоритма для решения поставленной в начале статьи задачи может значительно ускорить и упростить работу оператора по сведению данных.