

Адаптивный рандомизированный алгоритм выделения сообществ в графах

Тимофей Проданов

Добрый день, я Тимофей Проданов и я буду рассказывать об адаптивном выделении сообществ в графах.

Слайд 1

Последние двадцать лет развивается изучение сложных сетей, то есть графов с неправильной, сложной структурой. Сложные сети были с успехом применены в различных областях, например в биоинформатике и социологии. Сложные сети, построенные по реальным системам, часто разбиваются на группы тесно связанных узлов графа, сообщества. Способность находить и анализировать подобные группы узлов даёт большие возможности в изучении реальных систем, например, сообщества в интернете представляют распространённые темы. Поиск таких групп узлов в называют выделением сообществ или кластеризацией.

Слайд 2:
Выделение
сообществ

В 2004 году Марк Ньюман представил целевую функцию модулярность, показывающую качество разбиения графа на сообщества, определение модулярности изображено на слайде. Модулярность удобна в использовании, так как подсчёт выигрыша от объединения двух сообществ занимает одну операцию, а объединение двух сообществ — от количества соседей одного из сообществ. Таким образом, задачу выделения сообществ рассматривают как задачу максимизации модулярности.

Эффективными оказались рандомизированные алгоритмы выделения сообществ в графах. В 2010 году Овельгёне и Гейер-Шульц предложили рандомизированный жадный алгоритм выделения сообществ с параметром k . На каждой итерации рассматривается k случайных сообществ, у каждого сообщества исследуются соседи, и затем соединяется лучшая пара.

Слайд 3:
Алгоритмы
выделения
сообществ

В 2012 году те же учёные представили схему кластеризации основных группа графа. Сначала s начальных алгоритмов выделяют сообщества, а те узлы, которые начальные алгоритмы назначили в разные сообщества, распределяет по сообществам финальный алгоритм. В итеративной схеме узлы, относительно которых начальные алгоритмы не сошлись во мнении, снова распределяют по сообществам начальные алгоритмы.

В качестве начальных и финального алгоритмов можно использовать рандомизированный жадный алгоритм с разными параметрами.

Качество работы этих алгоритмов критично зависит от их параметров, и остаётся открытым вопрос об адаптивных версиях алгоритмов, которые были бы работоспособны на большем количестве задач.

Слайд 4:
SPSA.
Постановка
задачи

Для создания таких модификаций используется стохастическая аппроксимация. Алгоритм одновременно возмущаемой стохастической аппроксимации, или по другому *SPSA*, предполагает разбиение улучшаемого алгоритма на два n шагов. В течении нечётного шага улучшаемый алгоритм использует один параметр, а во время чётного — другой, после чего вычисляются следующие два значения параметра.

В работе рассматривается применение алгоритма *SPSA* к двум рандомизированным алгоритмам выделения сообществ для получения модификаций, способных хорошо работать на большем количестве входных графов.

Применимость *SPSA* обоснована теоретически для выпуклой в среднем функции качества. На слайде изображена медианная модулярность разбиений алгоритмом *RG* двух графов при разных значениях параметра k . На каждом графе модулярность либо достигает максимума при небольшом k , либо постоянно увеличивается при росте k . Однако время работы

Слайд 5:
Применимость
SPSA.
Функция
качества

линейно зависит от k , поэтому предлагается функция качества, зависящая и от модулярности, и от параметра k , она показана на слайде. Такая функция в среднем выпуклая и подходит для работы *SPSA*.

Предлагается адаптивный рандомизированный жадный алгоритм или *ARG*, построенный на основе *RG*. Алгоритм идентичен рандомизированному жадному алгоритму с параметром в k . Однако действие *ARG* разбито на шаги длиной в σ итераций, после окончания шага параметр k меняется.

В функции качества используется медиана прироста модулярности за последние σ шагов. Использовать прирост модулярности опасно, так как алгоритм рандомизированный и даже при плохом k иногда модулярность будет сильно возрастать.

Для получения следующих двух параметров k некоторая текущая оценка параметра возмущается в обе стороны. Следующая оценка будет ближе к тому возмущению, которое дало меньшее значение функции качества, формула вычисления изображена на слайде.

Слайд 6:
Адаптивный
рандомизи-
рованный
жадный алго-
ритм