

СИНТЕЗ ВТОРИЧНОЙ СТРУКТУРЫ АЛГЕБРАИЧЕСКИХ БАЙЕСОВСКИХ СЕТЕЙ: ВИЗУАЛИЗАЦИЯ РАСПРЕДЕЛЕНИЙ ОТНОСИТЕЛЬНЫХ СЛОЖНОСТЕЙ ПРЯМОГО И ЖАДНОГО АЛГОРИТМОВ¹

Зотов М.А., студент кафедры системного программирования СПбГУ,
стажер лаб. ТиМПИ СПИИРАН, zotov1994@mail.ru

Аннотация

В настоящей работе визуализированы распределения отношений времен работы конкурирующих алгоритмов синтеза минимального графа смежности, а также проведено сравнение практического распределения оценок с теоретическим. Анализ графиков показал, что практические оценки и теоретические имеют схожий вид и поведение. Результаты работы подтверждают корректность предположения о том, что логарифмы указанных отношений имеют нормальное распределение.

Введение

В теории принятия решений часто необходимо пользоваться некоторыми знаниями о предметной области и показателями их достоверности. В качестве структур, позволяющих хранить в себе такие данные и их оценки, могут выступать байесовские сети доверия и алгебраические байесовские сети (АБС). АБС могут принимать вид так называемых графов смежности, а они, в свою очередь, могут быть минимальными графами смежности (по числу ребер). Для синтеза минимального графа смежности по заданному набору вершин - фрагментов знаний, - был разработан ряд алгоритмов, корректность которых была также доказана. В частности, в отношении нескольких алгоритмов[1, 4] были проведены вычислительные эксперименты по получению статистических относительных оценок сложностей. При вычислении таких оценок, на одном из шагов осуществлялось логарифмирование промежуточных результатов с целью получения нормального распределения. Получение

¹ Статья содержит материалы исследований, частично поддержанных грантом РФФИ 15-01-09001 – «Комбинированный логико-вероятностный графический подход к представлению и обработке систем знаний с неопределенностью: алгебраические байесовские сети и родственные модели».

такого распределения хорошо тем, что над ним удобно вычислять такие характеристики, как дисперсия и доверительный интервал. Целью настоящей работы является визуализация указанного распределения, с помощью которой, возможно, удастся выявить некоторые закономерности и особенности пространства входных данных, что может быть полезным при дальнейших исследованиях предметной области.

Определения и обозначения

Воспользуемся системой терминов и обозначений, сложившихся в [5, 6, 7].

Пусть задан конечный алфавит символов A , а непустые множества символов (без повторов) — слова — рассматриваются как возможные значения нагрузок вершин графов (в основном, графов смежности) и их ребер. Пусть имеется набор вершин $V = \{v_1, v_2, \dots, v_n\}$ и такие нагрузки $W = \{W_1, W_2, \dots, W_n\}$, причем $\forall u \in V$ W_u является нагрузкой для вершины u .

Определение. Назовем неориентированный граф $G = \langle V, E \rangle$ графом смежности, если он удовлетворяет следующим условиям:

1. $\forall u, v \in V : W_u \cap W_v \neq \emptyset \Rightarrow \exists$ путь P в графе $G : \forall s \in P \Rightarrow W_u \cap W_v \subseteq W_s$.
2. $\forall \{u, v\} \in E \Rightarrow W_u \cap W_v \neq \emptyset$.
3. $\nexists u, v \in V : W_u \subseteq W_v$, — нагрузка одной вершины графа не входит полностью в нагрузку любой другой вершины.

Граф смежности с минимальным и максимальным числом ребер мы будем называть *минимальным графом смежности (МГС)* и *максимальным графом смежности* соответственно.

На Рис. 1 приведён пример графа смежности, который, однако, не является МГС. На Рис. 2 приведён один из возможных МГС, полученных из графа, изображённом на Рис. 1, путём удаления двух рёбер.

Алгоритмы синтеза минимального графа смежности

Для синтеза МГС был предложен ряд алгоритмов, отличающихся друг от друга как скоростными характеристиками, так и целесообразностью применения в каждом конкретном случае. К алгоритмам синтеза МГС относятся:

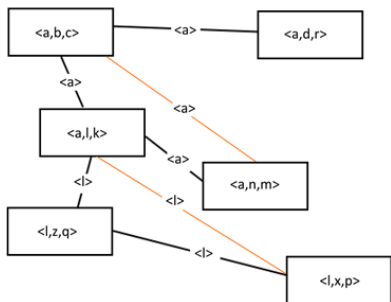


Рис. 1: Пример графа смежности

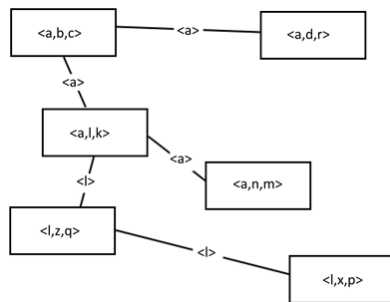


Рис. 2: Один из возможных МГС

- прямой;
- жадный;
- инкрементальный;
- декрементальный.

В отношении прямого и жадного алгоритмов были получены формальные оценки сложности, которые, однако, базируются на латентных свойствах пространства синтеза МГС, а также на свойствах МГС, которые будут известны только после синтеза МГС. Другими словами, полученные теоретические оценки сложности указанных алгоритмов оказались неприменимыми на практике.

Был предложен другой подход[4], который позволяет оценивать относительные сложности алгоритмов, базирываясь на статистических показателях; результаты таких эмпирических вычислений позволяют, с одной стороны, выбрать, в зависимости от характеристик пространства синтеза МГС, более подходящий алгоритм синтеза, с другой стороны, позволяют выявить возможные недостатки тех или иных алгоритмов, а также обнаружить и исследовать возможные пути улучшения алгоритмов.

Статистические относительные оценки сложности были получены для следующих алгоритмов:

- прямого и жадного;
- прямого и инкрементального;
- жадного и инкрементального;

- жадного и декрементального.

Подробное описание схемы проведения вычислительных экспериментов было дано в[3], поэтому отметим лишь, что выбранная схема проведения вычислительных экспериментов была разработана с целью решения следующих задач:

- уменьшение влияния состояния процессора и операционной системы на результат экспериментов (запуск вычислений M раз на *одних и тех же данных*);
- уменьшение влияния особенностей входных данных (запуск вычислений K раз на *различных данных*).

На выходе вычислительного эксперимента выдавалось K значений (статистик). Обработка этих значений (построение доверительного интервала, дисперсии, выделение квартилей, децилей и т.п.) и построение их визуализаций позволяло делать выводы о скоростных преимуществах конкурирующих алгоритмов.

Так, было установлено[2], что прямой алгоритм работает быстрее жадного на некотором поддиапазоне числа вершин графа (Рис. 3,4).

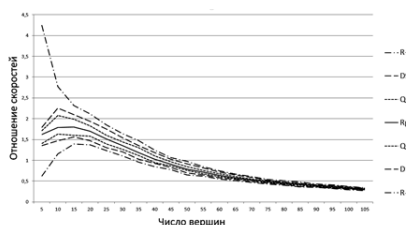


Рис. 3: 100% вершин 3-5 порядков

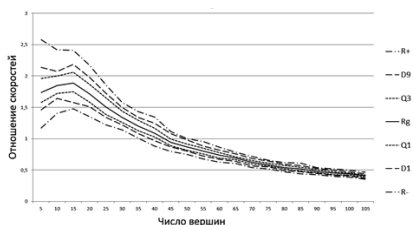


Рис. 4: 80% вершин 3-5 порядков, 20% вершин - 8-9 порядков

Было установлено[1], что инкрементальный алгоритм работает быстрее жадного и прямого алгоритмов (Рис. 5,6).

Обоснование корректности проведения вычислительных экспериментов

Как было сказано ранее, для получения относительных статистических оценок сложности, было сделано предположение о том, что логарифмы относительных времён работы конкурирующих алгоритмов

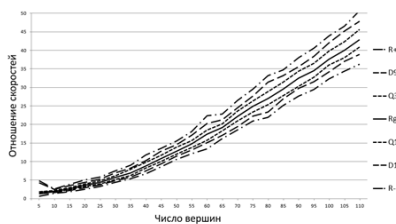


Рис. 5: Жадный и инкрементальный алгоритмы

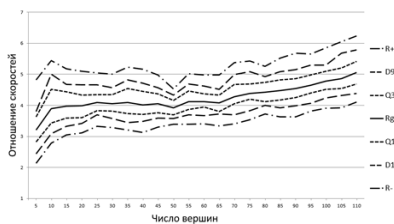


Рис. 6: Прямой и инкрементальный алгоритмы

распределены нормально. Для подтверждения этого необходимо также провести некоторые вычисления, которые смогут подтвердить изначальное предположение.

Необходимо будет построить гистограмму частоты повторения статистик для относительных времён работы алгоритмов, а также построить аналогичную гистограмму для *логарифмированных* относительных времён работы алгоритмов. Далее необходимо построить кривую нормального распределения. После получения необходимых графиков нужно будет наложить друг на друга кривую нормального распределения и гистограмму, полученную для логарифмированных относительных времён работы конкурирующих алгоритмов.

Построение частотных гистограмм

В качестве конкурирующих алгоритмов были выбраны прямой и жадный алгоритмы. Для них была построена гистограмма частоты повторения (точнее, попадания в отрезок) статистик. Гистограмма строилась следующим образом. На первом шаге полученные статистики относительных времён работ конкурирующих алгоритмов были упорядочены по возрастанию, далее высчитывался шаг статистики (разность максимального и минимального значений статистик делённая на число статистик), после этого считалась частота попадания статистик в каждый из шагов. Полученные данные были визуализированы для 35 вершин (Рис. 7 — 10).

Аналогичные вычисления были проведены и для графов мощностью в 45 и 55 вершин, приведём только кривые нормального распределения (Рис. 11,12).

Анализ рисунков позволяет сделать вывод о том, что пускай и имеются некоторые скачки и «неровности» в гистограммах, при наложении

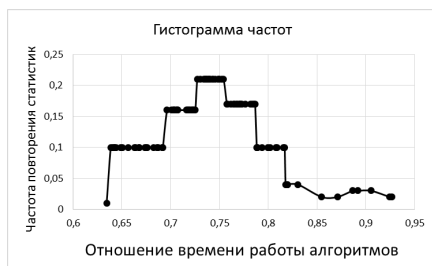


Рис. 7: Гистограмма для отношений времён работы алгоритмов



Рис. 8: Гистограмма для логарифмированных отношений времён работы алгоритмов

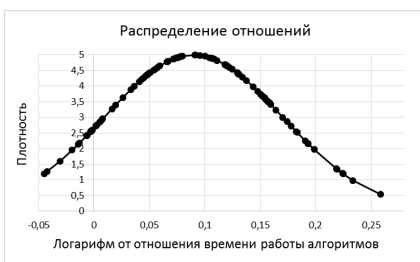


Рис. 9: Кривая нормального распределения для графа мощностью в 35 вершин

кривой нормального распределения чётко видно, что имеется совпадение сгущения точек, а также наблюдается общая тенденция роста и падения частот. Вычислительные эксперименты были проведены для вершин с распределениями:

- 90% — 2–4 порядков
- 7% — 5–7 порядков
- 3% — 8–10 порядков

Заключение

В статье резюмированы результаты предыдущих работ, связанных с построением относительных временных оценок сложностей конкурирующих алгоритмов синтеза минимальных графов смежности, а также

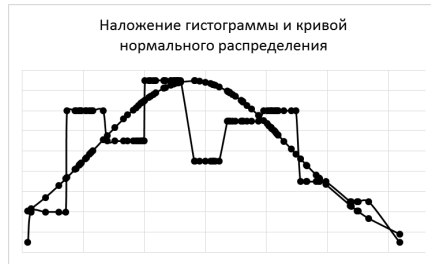


Рис. 10: Наложение кривой нормального распределения на гистограмму частот



Рис. 11: Кривая нормального распределения для графа мощностью в 45 вершин

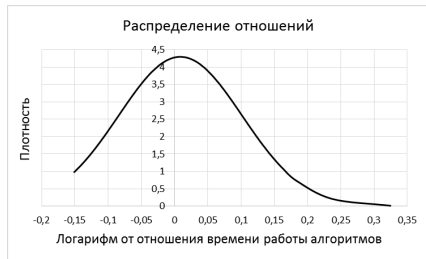


Рис. 12: Кривая нормального распределения для графа мощностью в 55 вершин

показана корректность предположения о том, что логарифмы относительных времён работ имеют нормальное распределение. Визуализированы кривая нормального распределения и гистограммы частот попадания статистик в каждый из шагов статистик.

Литература

- [1] Зотов М.А., Левенец Д.Г., Тулупьев А.Л., Золотин А.А. Синтез вторичной структуры алгебраических байесовских сетей: инкрементальный алгоритм и статистическая оценка его сложности // Научно-технический вестник информационных технологий, механики и оптики. 2016. Т. 16. № 1. С. 122–132.
- [2] Зотов М.А., Тулупьев А.Л. Вторичная структура алгебраических байесовских сетей: статистическая оценка сложности прямого ал-

горитма синтеза // SCM'2015: XVIII Международная конференция по мягким вычислениям и измерениям. Санкт-Петербург, 19-21 мая 2015 г. С. 158-162.

- [3] Зотов М.А., Тулупьев А.Л. Синтез вторичной структуры алгебраических байесовских сетей: методика статистической оценки сложности и компаративный анализ прямого и жадного алгоритмов // Компьютерные инструменты в образовании, 2015. № 1. С. 3–16
- [4] Зотов М.А., Тулупьев А.Л., Сироткин А.Л. Статистические оценки сложности прямого и жадного алгоритмов синтеза вторичной структуры алгебраических байесовских сетей // Нечеткие системы и мягкие вычисления. 2015. Т. 10. №1. С. 75–91.
- [5] Тулупьев А.Л., Николенко С.И., Сироткин А.В. Байесовские сети: логико-вероятностный подход. // СПб.: Наука, 2006. С. 607.
- [6] Тулупьев А.Л., Сироткин А.В., Николенко С.И. Байесовские сети доверия: логико-вероятностный вывод в ациклических направленных графах. // СПб.: Изд-во С.-Петерб. ун-та, 2009. С. 400.
- [7] Опарин В.В., Тулупьев А.Л. Синтез графа смежности с минимальным числом ребер: формализация алгоритма и анализ его корректности. // Тр. СПИИРАН. 2009. №11. С. 142–157.