

Представление и обработка запросов на основе множеств объектов с оценками

А.С. Ярыгина

Научный руководитель: Б.А. Новиков

Цели работы

- **Единообразный подход** к построению **сложных запросов** в различных сценариях **поиска**

Мотивировка



- Дом не дальше 500 метров от метро
- Уютный дом
- Мне нравится этот
- Точно не такой

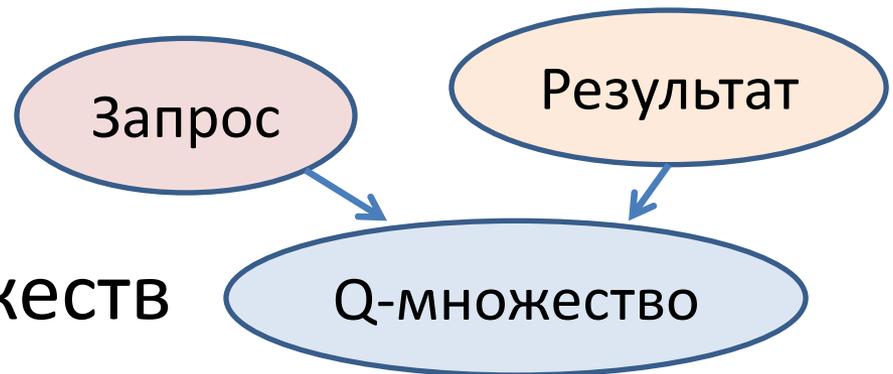


Постановка задачи

- Единообразный способ описания сложных видов поиска:
 - Описание поиска сложных объектов
 - Описание сложных сценариев поиска
- Необходимо:
 - Комбинирование разнородных типов условий в одном запросе
 - Приведение разнотипных условий к сопоставимой форме
- Набор высокоуровневых операций
 - Синтез
 - Калибровка
- Выражение различных сценариев поиска в терминах построенного набора операций

Запросы и результаты

- Q-множеством называется множество S вместе с определенной на нем функцией оценки $score: S \rightarrow [0,1]$



- Построение Q-множеств

$$score(a) = sim(q, a)$$

Задача синтеза

- Синтез данных и коллекций
 - Комбинирование цвета и текстуры при поиске изображений по содержанию в одной коллекции
- Интуитивные представления пользователя
 - Эффект снятия сливок
 - Эффект хора

Операции модели

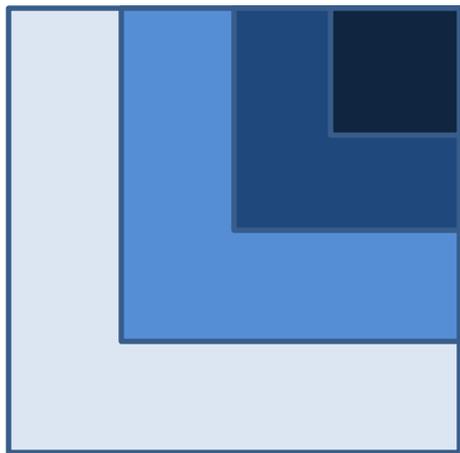
Синтез

- Симметричность
- Сохраняет упорядоченность
- Эффект хора
- Эффект снятия сливок

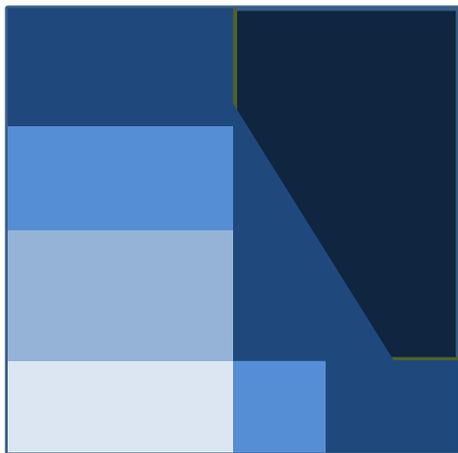
Калибровка (нормализация и усиление/ослабление)

- Чувствительность к выбросам
- Скос
- Качество

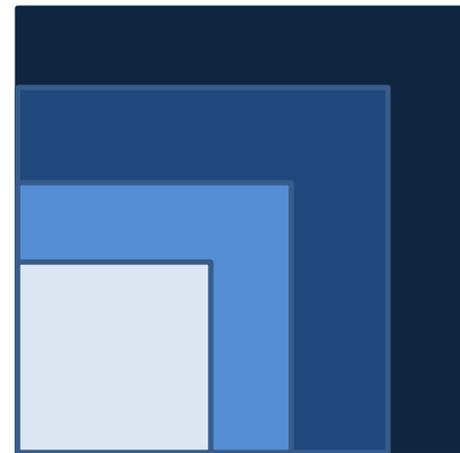
Реализация бинарных операций модели. Синтез



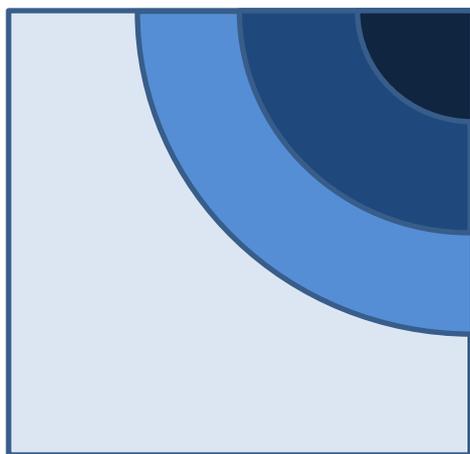
intersect



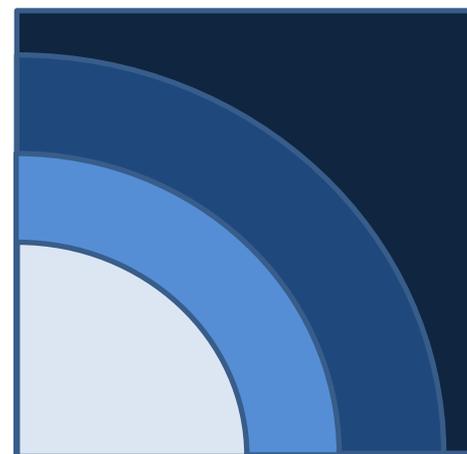
CombMNZ



union



super-intersect

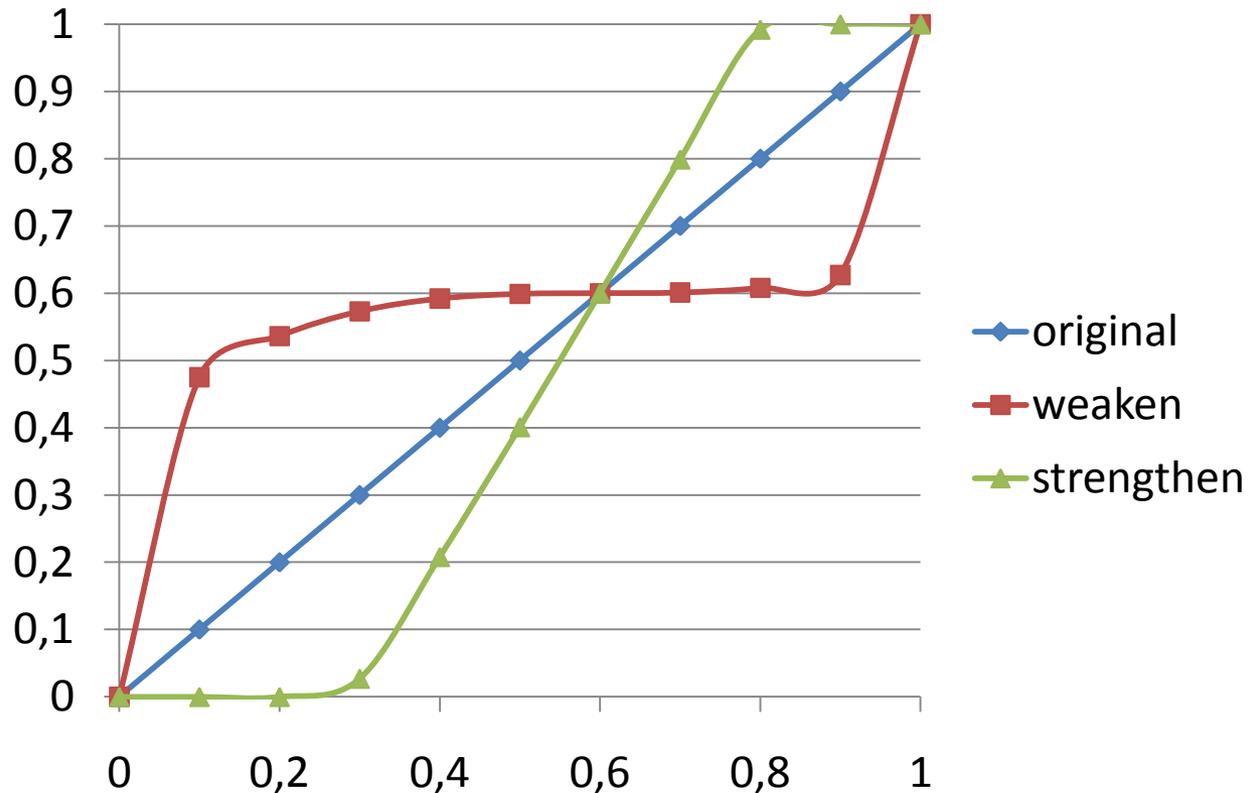


super-union

Реализация унарных операций модели. Нормировка

Операция	Формула	Чувствительность к выбросам	Скос
Norm-maxmin	$\frac{score(e) - \min(score(x))}{\max(score(x)) - \min(score(x))}$	-	-
Norm-avg	$similaring\left(\frac{distancing(score(e))}{avg(distancing(score(x)))}\right)$	+	-
Normalize-dist _p	$similaring(distancing(score(e)) * A)$	+	+
Normalize-score _p	$score(e) * B$	+	+

Реализация унарных операций модели. Усиление и ослабление



$$\text{similaring}((\text{distancing}(\text{score}(e))/M)^n)$$

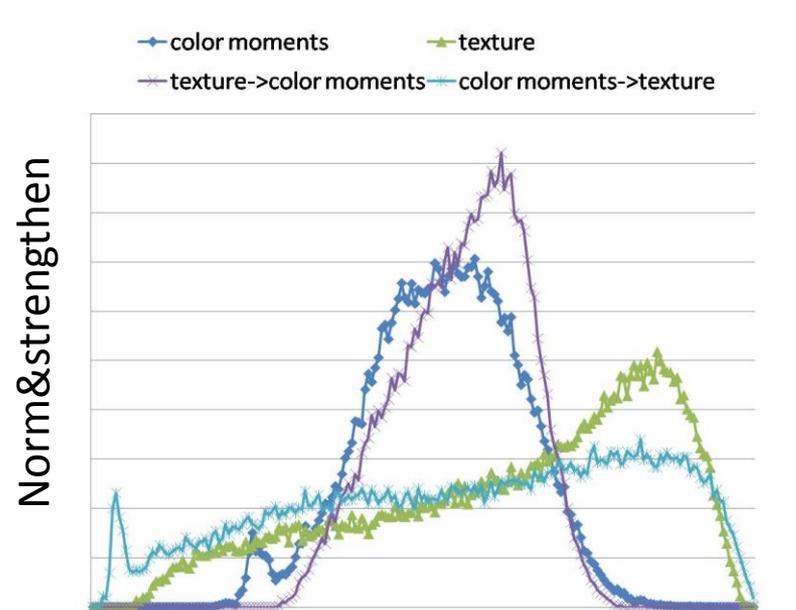
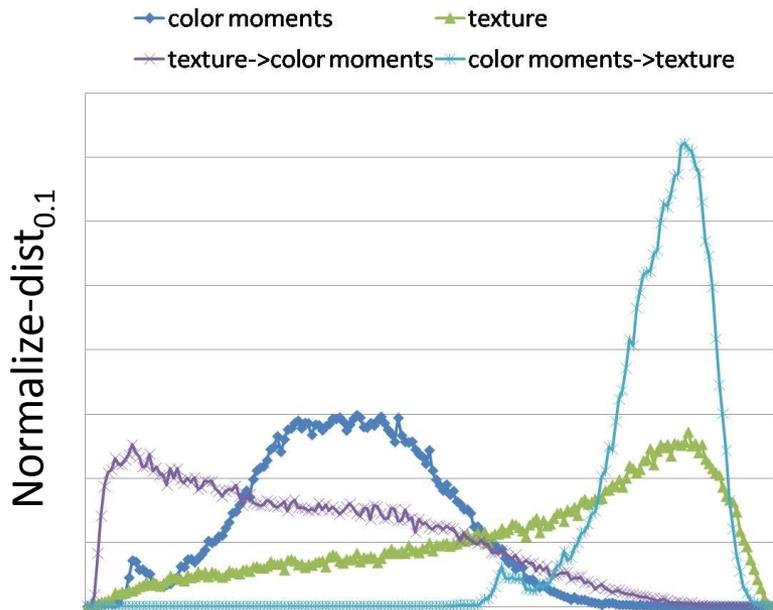
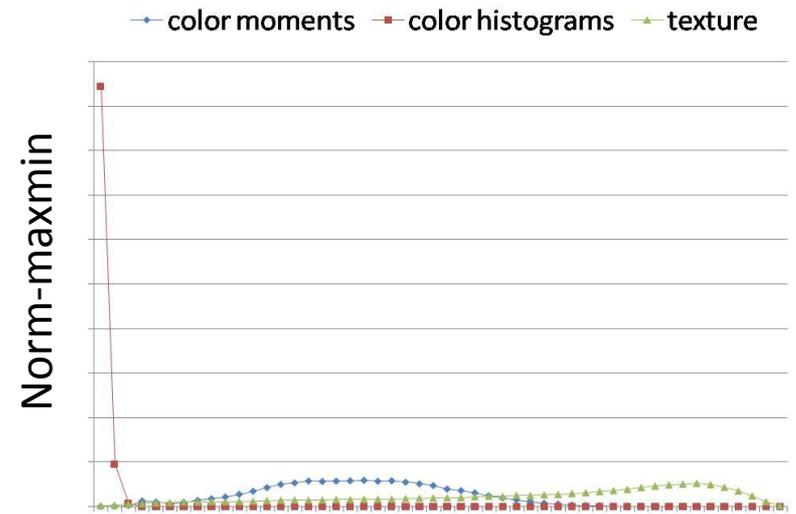
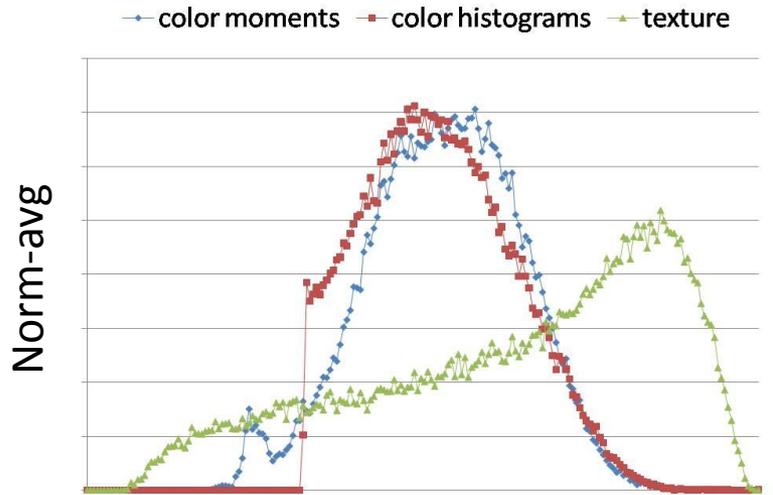
Пример

Intersect(
Union(
Intersect(
не дальше 500 метров,
уютный),
Strengthen(
похожий на тот, что мне понравился)),
Complement(
похожий на тот, что мне не понравился))

Экспериментальное окружение

- Коллекция изображений
 - 1087 изображений
 - 100 запросов
 - 16 семантических групп
- Цветовые моменты, цветовые гистограммы, текстура
- Точность (R-precision)
- Baseline - CombMNZ

Процедуры калибровки



R-precision

texture and color moments

	without norm	norm-avg	norm-maxmin	norm-dist(0.1)	norm&strengthen
CombMNZ	0,45	0,44	0,46	0,38	0,45
super-intersect	0,46	0,45	0,47	0,38	0,45
super-union	0,32	0,33	0,31	0,37	0,46

texture and color histograms

	without norm	norm-avg	norm-maxmin	norm-dist(0.1)	norm&strengthen
CombMNZ	0,20	0,41	0,20	0,37	0,43
super-intersect	0,45	0,41	0,45	0,37	0,43
super-union	0,16	0,35	0,16	0,37	0,44

color moments and histograms

	without norm	norm-avg	norm-maxmin	norm-dist(0.1)	norm&strengthen
CombMNZ	0,49	0,53	0,49	0,53	0,53
super-intersect	0,50	0,53	0,50	0,53	0,53
super-union	0,49	0,52	0,50	0,53	0,52

Результаты работы

- Единообразный подход к построению сложных запросов на основе подобия
- Методы калибровки и синтеза приближенных запросов, описанных функцией оценки
- Качество предложенных методов сопоставимо с известными подходами к решению задачи синтеза данных