

Яндекс

Яндекс на РОМИП-2010: тестирование простой ранжирующей формулы

Александр Сафронов
Разработчик

РОМИП, Казань, 15 октября 2010 года

Кратко о системе

- Используется для экспериментов с текстовыми факторами
- Не используется машинное обучение
- Позволяет быстро, но грубо оценить работоспособность нового фактора

Простая ранжирующая формула

$$Score(q, d) = \sum_i k_i * F_i(q, d)$$

Факторы

- BM25 для всего документа
- BM25 заголовка
- BM25 начальной части документа
- Вес самой длинной непрерывной цепочки
- Кучность
- YMW

Я

Фактор YMW

$$YMW = IWS * WCM$$

Фактор YMW: inverted window size

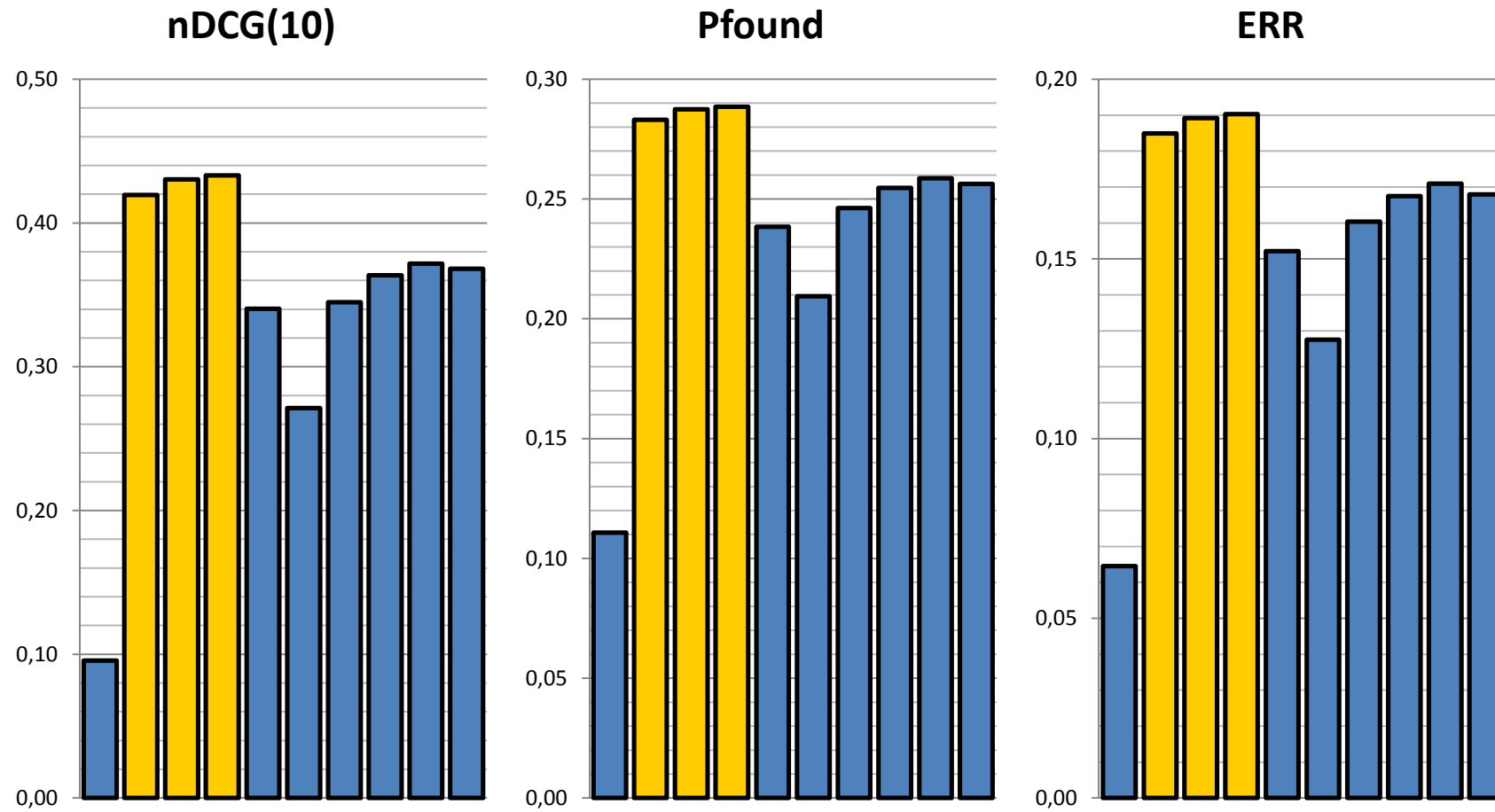
$$IWS = \frac{\log(a)}{\log(mw - |n| + a)}$$

Агеев, Добров, Лукашевич, Сидоров.
Экспериментальные алгоритмы
поиска/классификации и сравнение с
«basic line».

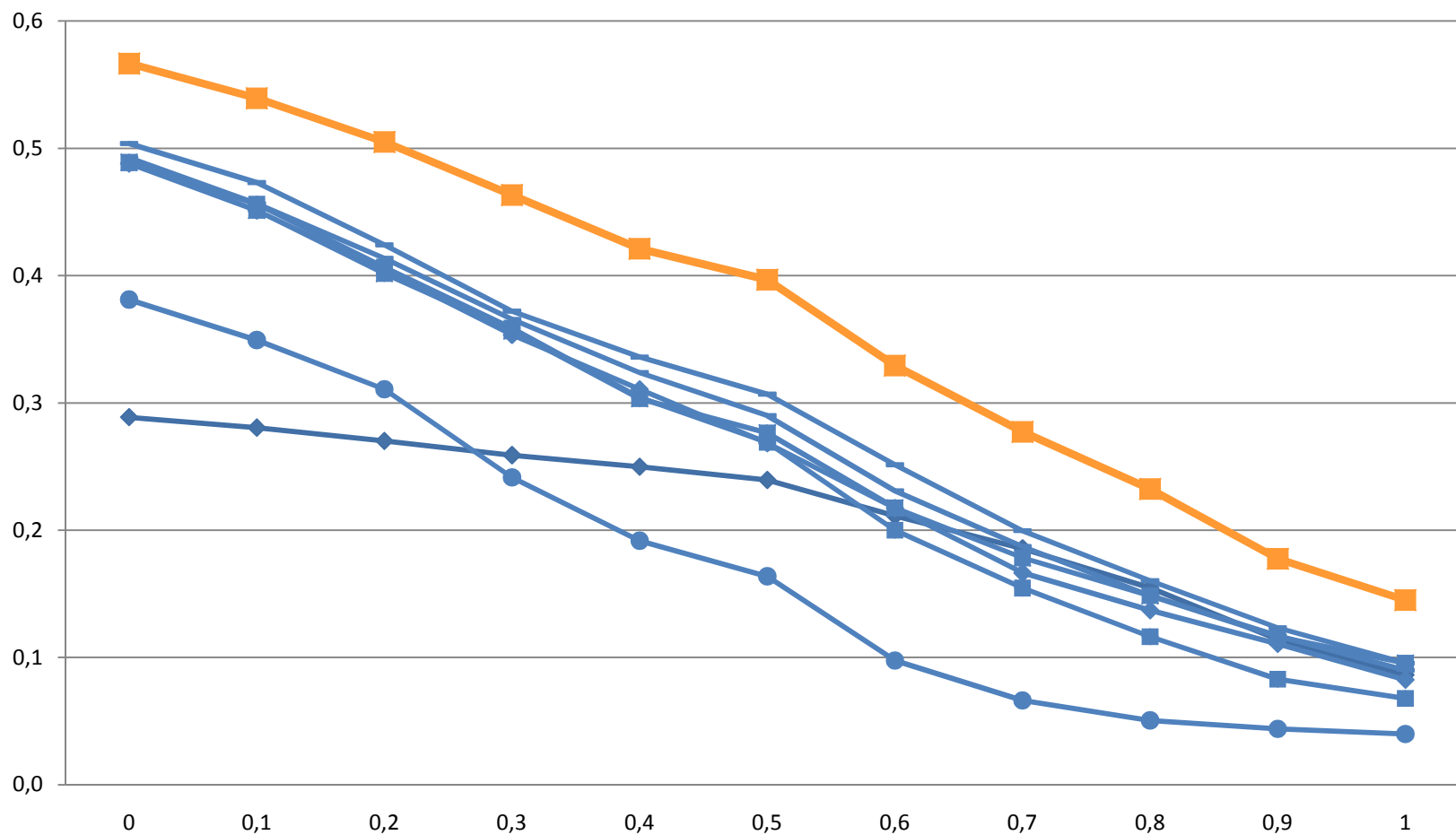
Фактор YMW: weighted coordination match

$$WCM = \frac{\textit{weight}(n)}{\textit{weight}(q) + b * (\textit{weight}(q) - \textit{weight}(n))}$$

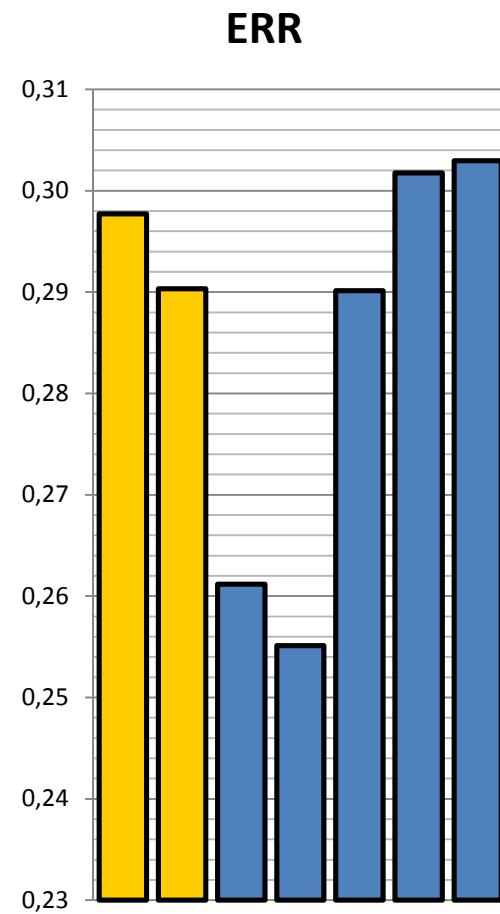
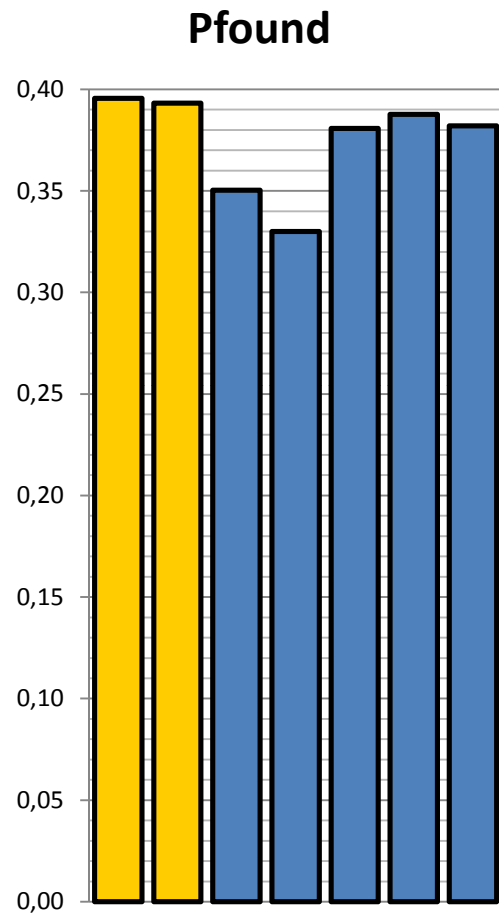
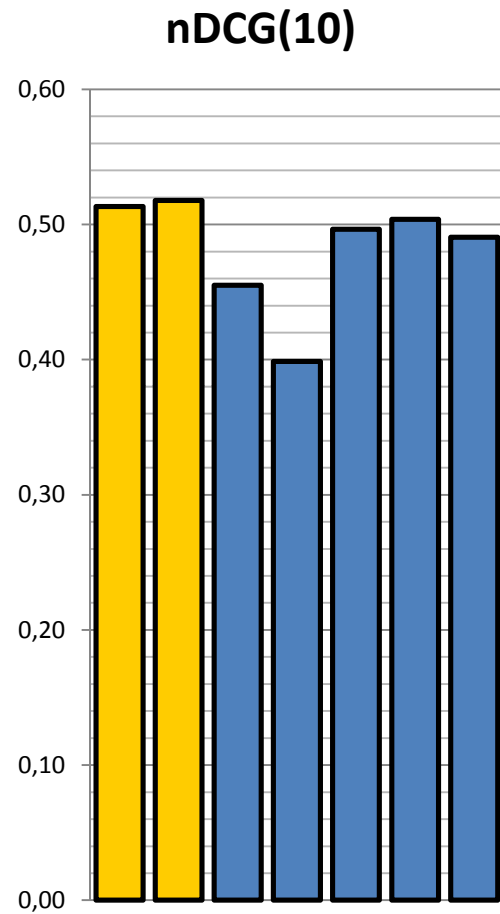
Коллекция ВУ, graded метрики



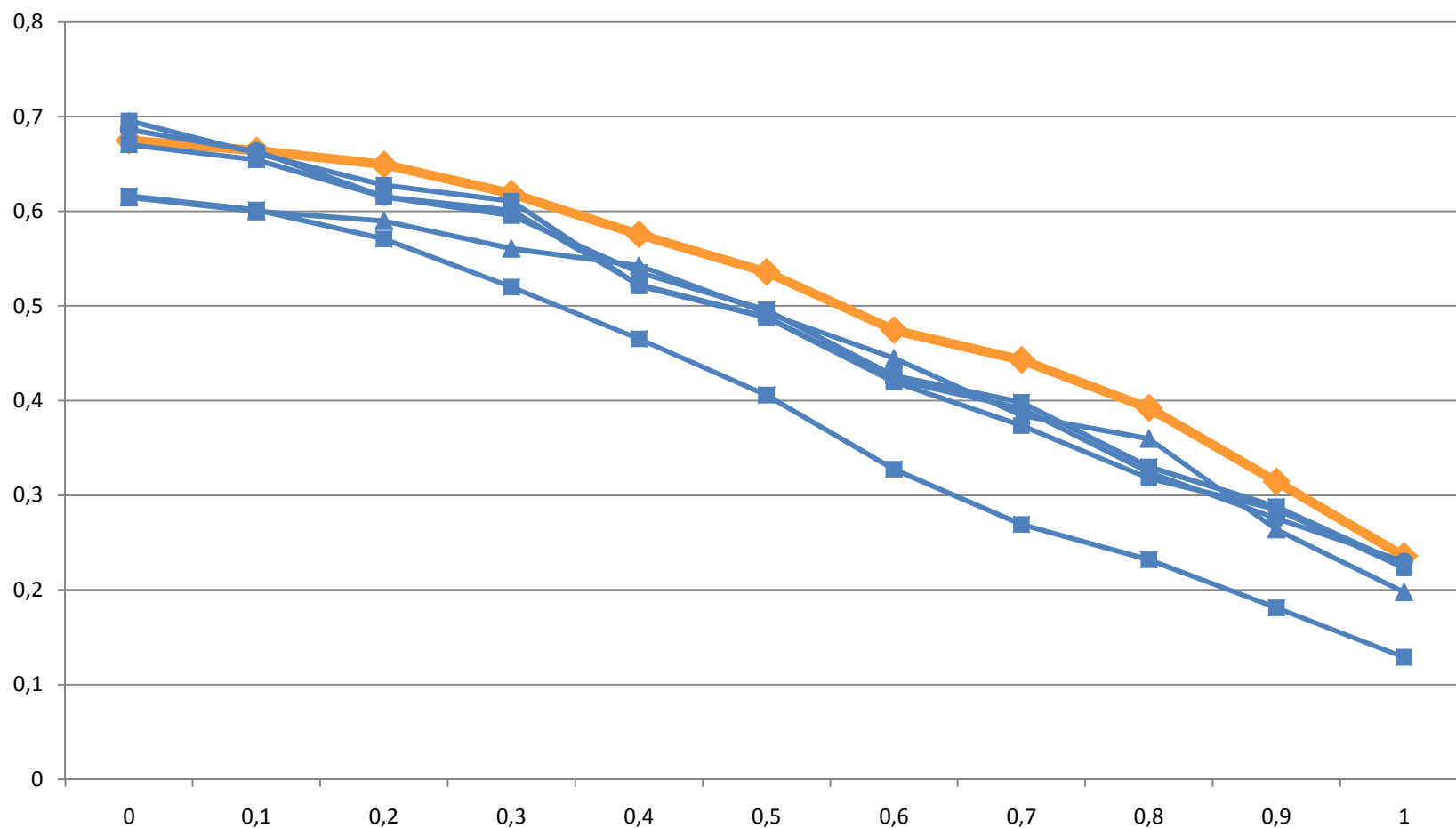
ВУ, график ТРЕС, relevant-plus or



Коллекция КМ, graded метрики



КМ, график ТРЕС, relevant-plus or



Проверка нового фактора в Яндексе

- Машинное обучение
- Перекрестная проверка (cross-validation)
- Т-критерий Стьюдента

Спасибо за внимание!

Я