**A Case for Ordinal Peer-evaluation in MOOCs**

АННОТАЦИЯ

MOOCs были весьма успешными из-за легкости распространения информации: Кто-нибудь с подключением к Интернету можно смотреть видео лекции и загрузить учебные материалы. Тем не менее, они по-прежнему отстают от обычного классы в одном из важнейших аспектов-обратной связи и оценки студентов-из-за серьезных расхождений в количестве студентов, обучающихся и количество имеющихся экспертов. Одним из способов проведения оценки и обратная связь по экспертной оценке, в котором ответы, представленные студентом являются анонимными и при условии, к набору других студентов, чтобы оценить. В течении всмотреться методы оценки, эти оценщики сверстников назначить количественные оценки в данные решения. В этой статье мы рассмотрим альтернативный подход к равному оценка на основе парных сравнений. Мы представляем доказательства того, что такое Порядковый подход может быть значительно более устойчивым к недостаток опыта оценщики, по сравнению с традиционными подходами количественных. Эта работа является первым шагом в понимании компромисс между точностью количественные оценки и надежность порядковых оценок для экспертной классификации.

1 Введение

Появление массовых открытых онлайн-курсов (или MOOC) через платформы, такие как Coursera и EDX позволила миллионам людей во всем мире получить доступ к качественному образованию в дешевом и удобным для Вас способом. Эти курсы, как правило, учатся несколько тысяч студентов, и эти число растет в быстром темпе. Самое главное преимущество MOOC является легкость распространения информация: любой человек с подключением к Интернету можете смотреть видео лекции и скачать исследование материала. Однако, MOOC до сих пор отстают от обычных классов в одном из важнейших аспектов: отзывы и оценки студентов. Из-за масштабности этих курсов, невозможно у инструктора или ассистента оценить тысячи ответов. Кроме того, поскольку эти курсы предлагаются бесплатно (и, надеюсь, продолжит быть), нанять платных специалистов для этой задачи не является практически осуществимым вариантом.

Одними из наиболее перспективных альтернативных методов оценки оценка авто и коллегиальная оценка, которые были опробованы с разной степенью успеха. Оценка авто использует программное обеспечение, чтобы оценить решений студентов для домашних заданий или экзаменов. В то время как методик оценки авто хорошо подходят для вопросы, которые являются достаточно объективным, как, например, вопросы множественного выбора или написания компьютер программы, они не подходят для субъективных вопросов. Рассмотрим, например, эссе из урок литературы или смартфон приложение от класса на взаимодействие человека и компьютера. Эти задания не может быть оценена также по современным машинным обучением, требуя мнение человека [1]. Действительно, многие области исследования, по сути, являются субъективными. Альтернативные способы проведения оценки и обратная связь экспертных оценок. В системах, использующих экспертной оценки, ответы студенту выдается и предоставляется набор



*Рисунок 1: пример отзывы, которые ученик получит в соответствии с планом Порядковый номер оценок. На рисунке рассматривается вопрос, в котором студенты должны спроектировать приложение часы для мобильные телефоны.*

других студенты, чтобы оценить. Рубрики, предлагаемые инструктором служит ориентиром для студентов в процессе оценки. Например, в Coursera работает экспертной оценки для человека и компьютера взаимодействие (hci) на поле [2,3], в котором Домашняя работа каждого ученика оценивается по 3-5 студентов, и итоговая оценка студента рассчитывается как медиана этих оценок.

Несмотря на свой огромный потенциал, экспертная оценка имеет лишь ограниченное признание в MOOC. Только несколько курсов уже попыталась прибегнуть к экспертной оценке. На этих курсах, окончательные оценки полученных студентами в рамках процесса взаимной оценки были зачастую предвзятыми. Такие нарушения приводят к серьезным оппозиции к трудовой экспертной оценки в MOOC [4]. Причины довольно очевидны. Значительная часть студентов, обучающихся в MOOC не серьезно относятся к курсу, не говоря уже об оценке работы своих коллег. Высокий процент отсева студентов также привести к сложной динамики в системе. Даже когда ученики имею в виду ну, отсутствие опыта и присутствие погрешностей приводит к значительному шуму в своих оценках. В система также сталкивается с дополнительными мягкими ограничениями, такими как ограничения на количество решений, что студенту может быть предложено оценить, или наличие очень небольшого числа экспертов-оценщиков, которые должны быть использовать с умом.

В данной статье рассматривается проблема проектирования схем экспертной оценки. Обычные оценки схемы, как правило, следуют количественный подход присвоения баллов каждому ученику (например, в HCl курс на coursera). В этой статье, однако, мы спорим на случай альтернативы порядковому подходу (сравнительный) подход к экспертной оценке. При таком подходе каждый студент-оценщик будет дано некоторые пары решений, и в каждой паре он/она должны выбрать решение "лучше". Пар может включает решения, предлагаемые другими студентами, или решения, которые были предварительно откалиброваны эксперты. Тогда эти сравнения будут агрегированы по системе, которая присваивает окончательную оценку к каждому студенту. В дальнейшем мы подробно остановимся на этом подходе, и обсуждать его возможные заслуги и недостатки по сравнению с традиционным количественный подход.

Мы считаем, что количественные и порядковые подходы, оба имеют значение, но отличаются по эффективности в зависимости от на проблему оценки. Мы покажем простую задачу, на которой люди гораздо больше способный в порядковой оценке, однако мы также понимаем, что простые попарные сравнения игнорировать информацию представить в количественных результатов. Наша работа является первым шагом в понимании Порядковый коллегиальная оценка. В будущем, комбинированный подход может сбалансировать надежность порядковые оценки точность количественных результатов.

Остальная часть статьи организована следующим образом. Раздел 2 сравнивает количественные и порядковые подходы к экспертной оценке. В разделе 3 рассматриваются начальная работа по моделирование порядковых экспертных оценок а также представлены некоторые предварительные результаты. Раздел 4 содержит Заключительные замечания.

**2 Количественные и порядковые оценки**

Традиционный метод оценки был кардиналом в природе, но такие оценки были как правило, выполняется специалистами. Неясно, является ли этот подход является эффективным в отсутствие экспертизы. В этом разделе мы выступаем за альтернативу Порядковый средства оценки и обеспечения качественное сравнение этих двух методов.



Таблица 1: Скриншоты и без ошибок скорости двух экспериментов, сравнивающих количественных и порядковых задачи по Amazon Mechanical Turk. (Скриншоты показывают только порядковое часть эксперимента.)

**2.1 Вопросы калибровки**

Каждый студент имеет присущие уклоны, которые могут меняться со временем или зависеть от качества другого ответы оценивали. Студент может быть консервативным и всегда назначать умеренные оценки, в то время как другой студент может иметь тенденцию завышать оценки в крайности. Эти несоответствия трудно чтобы узнать, с размерами выборки размером до 5 или 10, которые можно было бы получить, как правило, в установке по экспертной оценке.

Кроме того, эти несогласованности изменяются со временем, что делает их труднее изучать и модель. На с другой стороны, такие несоответствия устраняются автоматически (по дизайну) в порядковом установки.

Как показано ниже, с помощью практических экспериментов, люди могут быть значительно более компетентен при выполнении порядковые оценки, чем те количественные.

**2.2 Экспериментальные результаты по точности в Количественный против порядковых оценок**

Можно было бы получить оценки порядковые из количественных из них: после сбора оценок количественных, любой пара решений можно сравнить с использованием количественных баллов. Такой аргумент предполагает, что порядковое подход не предусматривает каких-либо дополнительных данных, а на самом деле приводит к потере информации.

В этом разделе мы представляем результаты некоторых экспериментов, проведенных нами с помощью Amazon Mechanical Turk краудсорсинга платформу, предложить совершенно противоположное. Эксперименты показывают, что когда оценки выполняются людьми, порядковые оценки содержат значительно меньше шума, чем количественные оценки.

(А) Оценка возраста людей из фотографий: Каждая задача раздать работникам нужно было получить расчетные возраст 10 человек, чьи фотографии были показаны. В порядкового установки, рабочие должны были выбрать пожилой человек из пар картинок, в то время как установка количественное необходимых рабочих, чтобы войти в Рассчитываемый возраст. В общей сложности 100 человек выполнили задачи.

(Б) оценка площади кругов: каждая задача данного работникам состоит из 25 вопросов. В Количественный вариант задач, для каждого вопроса, работник был показан круг в ограничивающий прямоугольник, и работник был обязан определить процент площади бокса, что круг, занимаемого. В порядковое версия, работник был показан два круга в отдельных идентичных коробках ограничивающими, и Работник должен был определить круг, занимаемое больший процент площади в соответствующем коробка. Ограничивающая коробка была 200 пикселей в ширину и 200 пикселей в высоту, а радиус окружности был выбран в качестве 30? Бета (15, 3), где бета обозначает бета-распределения. Количественные ответы были затем преобразуется в порядковой форме, выбирая пары вопросов и глядя на какой круг был учитывая более высокое значение. В общей сложности 50 человек выполнили задачи.

В таблице 1 показаны результаты. Преобразование количественные ответы на порядковых ответов приводит к значительно выше, чем частота ошибок напрямую просят порядковых оценок.

**2.3 Удобство оценки**

Проблема независимой оценки в MOOCs требует решения не только из-за несоответствий отсутствие опыта, но и с дефицитом ресурсов. Это практически неосуществимым, чтобы спросить каждого студента уделять слишком много его / ее времени и усилий в решении задач по экспертной оценке. Таким образом, интерес к максимизировать количество информации, которая может быть собрана в соответствии с ограниченным усилием студентами, и разработать процесс оценки сверстников, не усталость студентов-оценщики. С этой целью его достаточно хорошо известно [6, 7], что люди часто находят значительно проще, чем сравнивать счет. *В Настройку независимой оценки, оценщик может найти задачу обеспечения простого сравнения быть гораздо проще, чем предоставление точной балльной оценки.* Порядковое подход позволит затем для сбор большего количества оценок для того же уровня усилий, по сравнению с оценками количественных.

**2.4 Конкретная Конструктивная обратная связь**

Учитывая, что оценщики не являются экспертами, мы выдвигаем, что порядковое обратной связи, как показано на рис. 1, может часто может быть более желательным для студента по сравнению с комментариями по абсолютной шкале. Так как люди лучше в сравнении предметов, оценщик может быть в состоянии предложить более проницательные замечания по положительные и отрицательные аспекты решения студента, и сделать эти пункты более конкретным путем сравнения с другой solution.1 Такая точечная оценка будет также направлять студента в понимании как именно он / она может улучшить его / ее решение.

**4. Выводы**

Мы постулировать порядковое подход к Peer-оценки в MOOCs, которые мы утверждаем, является устойчивым к отсутствием экспертизы среди грейдеров. Мы также представляем первоначальную работу по моделированию и анализу такого порядковое установки, и получить обнадеживающие результаты от предварительных экспериментов на реальных и синтетических данные.