

*Интервью с Миколой Печенижским – соорганизатором (PC Chair)
конференции Educational Data Mining 2011
(Технический университет Эйнховена, Нидерланды)*

Вопросы

1. На 4-ой конференции по анализу данных в образовании (EDM 2011) присутствовало около 100 человек из США, Европы и других стран, среди них много участников из такого известного в области компьютерных наук университета как Карнеги Мелон. Почему в последние годы эта тематика стала такой актуальной и привлекательной для специалистов в области Computer Science и особенно машинного обучения?
2. Какие ключевые направления конференции Вы бы могли отметить?
3. Насколько сильно Ваши интересы пересекаются тематикой конференции? Какие задачи для Вас наиболее интересны? Интересна ли Вам когнитивная психология и представлена ли она в рамках конференции?
4. Участники конференции – и маститые профессора, и молодые ученые довольно открыто общались и обменивались идеями, ни те, ни другие не стеснялись представлять доклады в форме постера. Почему сессии постерных докладов так важны? И что такое хороший постерный доклад в двух словах?
5. Среди работ, представленных на конференции были выявлены победители. В чем вклад этих людей для сообщества?
6. Каким на Ваш взгляд может оказаться влияние конференции на развитие будущих образовательных технологий?
7. В следующем году очередная конференция EDM пройдет на Крите, хотелось бы Вам (сообществу) увидеть ученых России и стран СНГ с интересными докладами? Что бы Вы могли им пожелать?

Стенограмма интервью

1) *На 4-ой конференции по анализу данных в образовании (EDM 2011) присутствовало около 100 человек из США, Европы и других стран, среди них много участников из такого известного в области компьютерных наук университета как Карнеги Мелон. Почему в последние годы эта тематика стала такой актуальной и привлекательной для специалистов в области Computer Science и особенно машинного обучения?*

Д. Почему в последние дни тематике стала такой актуальной?

М. Честно говоря, я бы сказал, что в Computer Science пока она не очень актуальна, более актуальна в смежных областях, в том числе в психометрике и образовании, применении искусственного интеллекта в образовании. На протяжении нескольких последних лет организовывались workshop'ы на смежных конференциях по User Modeling, Adaptive Hypermedia и Intelligence Tutoring Systems. И вот четыре года назад впервые организовали конференцию по анализу данных из образовательных доменов.

Д. В чем специфика, в чем интерес, почему сейчас?

М. Собственно, сейчас больше гораздо данных собрано и собирается. Т.е. тут нужно отметить, что образование понимается достаточно широко, это может быть как образование университетское, так и образование в школьных учреждениях.

Д. Джек Мостов, например, учит детей читать.

М. Да, это, как правило, довольно серьезные большие программы как для обычных детей в дошкольном и младшем школьном возрасте в Штатах, также и в развивающихся странах, в Африке, например, в Скандинавии, в частности, в Финляндии тоже очень серьезно занимаются вопросам как использовать компьютеры в обучении детей чтению, арифметике, базовой математике и тому подобное. Но в последнее время все больше и больше используются Technology Enhance Learning. ITS технологии как в университете, особенно для дистанционного обучения, но все больше и больше технологии используются в обычном университетском образовании, т.н. Blended Learning, т.е. когда дополнительные средства используются для поддержания традиционного обучения с помощью лекций и семинаров. Это и видеолекции, материалы выкладываются онлайн. Можно собирать самые различные данные, то как учащийся использует ресурсы, проходит тесты, какие оценки получает. И поэтому можно рассматривать самые различные задачи, в которых методы статистики, машинного обучения и data mining очень полезны как для учащихся, так и для преподавателя, так и для людей ответственных за весь образовательный процесс в целом.

Д. Понятно, а вот системы, это Learning Management Systems (LMS) и есть? Да?

М. LMS сейчас активно используются в различных университетах для поддержки традиционного образования. Существуют университеты, которые специализируются на дистанционном обучении. Они тоже используют LMS и различные дополнительные средства. В этих LMS, Я думаю, Вы знаете, что очень много данных собирать о том, что где, когда студенты делают, как они переписывают.

Д. В Высшей школе экономики внедряют один из продуктов. Пока ситуация находится на уровне пробного внедрения.

М. Много людей было представлено из Carnegie Mellon. Да это действительно так, CM это один из крупнейших университетов, где активно разрабатываются системы для создания т.н. Intelligent Tutoring Systems и они очень активно совмещают как базовые исследования, так и их коммерциализацию, т.е. они выпускают реальные продукты, которые используются в школах. Например, Algebra Tutor. В основном они используются в Штатах, но есть и специальные проекты, как дети из Африки обучаются, например. В принципе, похожие системы разрабатывается в Германии тоже для обучения математике, но в Европе данная система не носит такой массовый характер как в США.

2) Какие ключевые направления конференции Вы бы могли отметить?

Д. Какие ключевые направления конференции Вы могли бы отметить? Например, три наиболее важных.

М. Есть несколько значимых направлений, которые очень легко заметить любому, кто посещал конференцию в этом году и в предыдущие годы. Т.е. это использование Data Mining для поддержки Intelligence Tutoring Systems. Т.е. как обрабатывать данные о том... Идея в том, что в интеллектуальных тьюторах используется огромное количество заданий, которые учащиеся могут выполнять и все эти интеракции и, собственно, результаты учащихся записываются и в последующем могут анализироваться с помощью методов Data Mining. Для построения т.н. моделей пользователей или моделей студентов, которые можно использовать для адаптации, персонализации и направления учащихся в ту или иную область знаний, в которой им нужно подтянуться.

Д. Применение Data Mining для Intelligent Tutoring Systems – это одно направление.

М. Я бы даже выделил Student Modelling как одно из больших направлений. Т.е. оно <DM> используется сейчас в основном для Intelligence Tutoring Systems, а так же есть еще параллельное направление Adaptive Educational Hypermedia Systems. Эти системы отличаются от Intelligent Tutoring Systems тем, что у пользователя гораздо больше свободы в том, какие именно ресурсы они могут просматривать в тот или иной момент, а так же системы, такие как LMS, там примерно такие же подходы используются для определения какие типы пользователей есть, как они взаимодействуют с системой, как им помочь найти наиболее подходящие ресурсы для обучения и тому подобное. Второе, как мне кажется, значимое направление, которое ещё не очень широко представлено, но с каждым годом все больше и больше статей появляется это т.н. Process Mining in Educational Domains. В данном случае мы заинтересованы не в каких-то отдельных паттернах, либо модели предсказания, а мы хотим понять как собственно устроен процесс обучения, например, в университете, т.е. какие курсы в какой последовательности большинство студентов берет и сдает.

Д. Используются ли для этого какие-либо UML средства?

М. Действительно, можно использовать различные UML средства для описания этих процессов, но исторически так сложилось что в Process Mining Community использовались сети Петри и другие модели представления данных, моделирования процессов и, в частности, на нашей кафедре тоже широко поддерживается это направление – Process Modelling и Process Mining в различных областях. И собственно недавно мы применяли эти методы для образовательных данных. Также важным направлением является визуализация данных, визуализация паттернов. На сегодняшний день мы не можем сказать, что очень много работ представлено в этой области, точнее не так легко показать новизну и значимость с точки зрения экономической ценности работы. Но все время появляются либо короткие статьи, либо постеры, которые представляют какие-то интересные идеи, как визуализировать обучающие данные.

Обсуждение тем по презентации на открытии конференции.

Generic frameworks methods это всё сюда подходит, поэтому она очень большая. Mining Assessment Data, в принципе, это очень похоже на Student Modelling, т.е. это такое общее Interaction Data, Browsing Data и Assessment Data – в принципе, три этих компоненты и служат для Student Modeling, поэтому этот класс такой

большой... Интегрированный Data Mining и педагогика – использование экспертных знаний, интегрирование их в учебный процесс. И задача адаптации и персонализации опять же тесно связана со Student Modelling.

Д. А адаптация потребностям студентов?

М. Да, адаптация на основе модели студентов, подбираются наиболее подходящие материалы, либо разбиение этого материала на предметные последовательности. Также наиболее оптимальное представление материала, т.е., во многих системах один и тот же материал представлен в различных вариантах, например, преобладает теория, или то, что больше нравится, подходит для данного студента. Process Mining уже упомянут.

М. Последнее название можно расшифровать, Replication Studies, что имеется в виду?

М. RS имеется в виду, что, собственно, уже какая-то теория была построена и был задел, экспериментальное подтверждение этой теории и люди хотят реплицировать, повторить такое исследование в своей стране. Опять же такую работу очень сложно опубликовать как полную статью, поэтому её публикация возможна как постера или короткой статьи.

Д. Спасибо. Мы обсудили ключевые направления, перейдем к вопросу 3.

3) А насколько сильно Ваши интересы пересекаются тематикой конференции? Какие задачи для Вас наиболее интересны? Интересна ли Вам когнитивная психология и представлена ли она в рамках конференции?

М. На нашей кафедре развиваются несколько направлений серьезно, это Data Mining, Process Mining и Adaptation и Personalisation, т.е. Adaptive Information Systems. Т.е. мы занимаемся этими направлениями в различных областях. В области образования эти направления продвигаются аналогичным образом. Мы разрабатываем, либо применяем существующие методы Data Mining и Process Mining для образовательных данных, либо смотрим в чем их особенности и какие новые методы необходимы либо для анализа interaction data между студентами и различными образовательными системами, либо assessment data, т.е. насколько хорошо или насколько плохо студенты справляются с заданиями. И как можно использовать средства Data Mining для адаптации и персонализации контента и презентации.

Д. Интересна ли Вам когнитивная психология и представлена ли она в рамках конференции?

М. Как человеку интересна, как исследователю тоже, но пока это не является таким серьезным направлением наших исследований. На конференции также Вы найдете очень много работ в этом направлении. Но часть людей, которые занимаются Intelligent Tutoring Systems, не рассматривают проблемы определения meta-cognitive skills опять же с точки зрения student modeling или learner modelling, точно также как когда психологи и лидеры, ведущие в образовании создания контента для обучения учащихся алгебре, чтению, счету используют свои знания в когнитивной или нейрокогнитивной психологии.

4) *Участники конференции – и маститые профессора, и молодые ученые довольно открыто общались и обменивались идеями, ни те, ни другие не стеснялись представлять доклады в форме постера. Почему сессии постерных докладов так важны? И что такое хороший постерный доклад в двух словах?*

М. Достаточно широкий вопрос. Т.е. с одной стороны постеры можно рассматривать как вид презентации работы. Т.е. на многих конференциях сейчас можно увидеть сейчас постеры. И идея в том, что можно в менее формальной обстановке обсудить работу, какие-то конкретные детали, что редко случается возможным на обычных пленарных сессиях. Т.е. пленарная сессия, как правило, это презентация и несколько минут отводится на вопросы и ответы. Во время постеров вы можете обойти каждую из нескольких групп людей и получить ответы на большее количество вопросов. Как Вы правильно заметили, гораздо проще молодым ученым общаться с профессорами и спросить либо совет, либо получить какие-то замечания по поводу их работ и аналогичных исследований проводимых в других университетах. На многих конференциях постеры рассматриваются просто как другой формат представления работы. Однако на большинстве конференций постеры подразумевают, что представляется работа, которая ещё не доведена до конца, т.е. какие-то наработки есть, предварительные результаты получены, есть какое-то интересное начинание, но работа ещё не закончена и её еще рано принимать как полноценную статью, однако она достаточно интересна для большинства людей, которые пришли на конференцию и могут обсудить работу на данном этапе и получить совет в каком направлении им двигаться дальше.

5) *Среди работ, представленных на конференции, были выявлены победители. В чем вклад этих людей для сообщества?*

М. В этом году был приз за лучшую работу в общем зачете, а также приз за лучшую студенческую работу, т.е. студент должен быть либо единственным автором, либо первым автором статьи. А также приз за лучший постер. А если кратко, то все статьи находятся в открытом доступе, все, кто заинтересован, пожалуйста, приходите, читайте, смотрите. Как правило, приз дается за работу, которая внесла наибольший вклад на данном в развитие области. EDM достаточно молодая дисциплина и мы стараемся поощрять наиболее интересные доклады и работы. В этом году две студенческие работы были очень близки друг к другу по показателю новизны и значимости и было принято решение дать две студенческие награды. Что касается постеров это больше как приз зрительских симпатий с одной стороны, с другой стороны важно мотивировать больше и больше людей присылать свои работы в качестве постеров, чтобы конференция расширялась и все больше и больше новых направлений было представлено в рамках конференций. Мы стараемся поощрять наиболее интересные.

6) *Каким на Ваш взгляд может оказаться влияние конференции на развитие будущих образовательных технологий?*

М. Мне кажется, что стоит выделить два основных направления, первое – это, собственно, создание технологий в которых будут использоваться в Intelligent Tutoring Systems, Adaptive Hypermedia и других средствах, которые становятся все более популярными. С другой стороны, мне кажется, что Educational Data

Mining позволит многим образовательным процессам стать более прозрачными и понятными как для учащихся, так и преподавателей и людей, ответственных за процесс образования. Что имеется в виду, сейчас стал возможным сбор самых различных данных об образовательных процессах, как студенты обучаются, как сдают курсы, либо какие-то конкретные задания. Т.е. если Educational Data Mining поможет в каждом образовательном учреждении, в каждом университете или школе предоставлять информацию о том, что собственно происходит, это позволит лучше понимать эти процессы и вносить корректировки, для преподавателей это будет большой задел для адаптации и улучшения их курсов, для директоров учреждений и людей, ответственных за образовательные процессы, т.е. они смогут пересмотреть данные программы, посмотреть где узкие места, почему много студентов не справляются с тем или иным курсом, придумать наиболее правильную последовательность как студенты должны изучать тот или иной материал. Т.е. возможно давать советы студентам почему именно важно пройти этот курс и не просто пройти, а получить хорошую оценку, на что это будет влиять и т.п., все эти советы будут не голословными, а поддержанными Educational Data Mining.

7) В следующем году очередная конференция EDM пройдет на Крите, хотелось бы Вам (сообществу) увидеть ученых России и стран СНГ с интересными докладами? Что бы Вы могли им пожелать?

М. Да, конечно, на конференции представлены участники многих стран. К счастью, в этом году были также участники из России, что, к сожалению, происходит не так часто, т.е. русскоговорящих участников, как правило, много не только на конференциях по Educational Data Mining, но и Computer Science конференциях в целом. Как правило, эти участники приезжают из Германии, США и других стран, т.е. конечно хотелось бы видеть больше участников из России и стран СНГ. В этом нет ничего невозможного, присылайте интересные работы. Если для вас эта область новая приезжайте просто пообщаться с коллегами.

Д. Хотелось бы добавить: не стесняйтесь делать постерные доклады.

М. Да, да, не стесняйтесь делать постерные доклады – это очень хорошее начало, хороший задел для того чтобы показаться в сообществе, завязать отношения с коллегами и продолжить research.

Д. Благодарит за интервью.