

УДК 004.414.3

ФОРМИРОВАНИЕ АКАДЕМИЧЕСКИХ ГРУПП И ПРОЕКТНЫХ КОМАНД НА ОСНОВЕ СБОРА ДАННЫХ ОБ ОБУЧАЮЩИХСЯ

Н. А. Коргутлова¹, С. Ю. Басаргина², М. М. Абрамский³, М. А. Солнцев⁴,
Т. С. Бузукина⁵

¹⁻⁵Высшая школа информационных технологий и интеллектуальных систем
Казанского (Приволжского) федерального университета

¹korgutlova97@gmail.com, ²basargina.svetlana@gmail.com, ³ma@it.kfu.ru,
⁴mrt.solncev@gmail.com, ⁵taya.buzukina@gmail.com

Аннотация

Обсуждены вопросы использования данных об обучающихся, представленных в электронном виде, в задачах генерации распределений обучающихся по академическим группам, элективам и проектным командам. Проиллюстрировано применение алгоритмов машинного обучения для этих задач. Показана возможность использования данных, собранных из социальных сетей.

Ключевые слова: *личностный портрет студента, кластеризация, распределение по компетенциям, анализ социальных сетей*

ВВЕДЕНИЕ

Согласно Указу Президента РФ от 07.05.2018 г. №204 [1], основными задачами в сфере образования в ближайшие годы являются: «внедрение на уровнях основного общего и среднего общего образования [...] образовательных технологий, обеспечивающих освоение обучающимися базовых навыков и умений, повышение их [...] вовлеченности в образовательный процесс, а также модернизация профессионального образования, в том числе посредством внедрения адаптивных, практико-ориентированных и гибких образовательных программ».

Одним из приоритетных направлений модернизации образования является формирование новых моделей организации учебной деятельности, учитывающей вариативность и индивидуальность высшего образования. Главным достоинством индивидуального обучения [2] является то, что оно позволяет пол-

ностью адаптировать содержание, методы и темпы учебной деятельности студента к его особенностям, следить за каждым его действием при решении конкретных задач.

Одним из видов реализации индивидуального обучения является обучение в сотрудничестве или малых группах, однако сразу же встает вопрос о методе распределения по группам, которое должно быть не случайным, а основано на личностном портрете студента – его характеристиках, компетенциях, интересах и др. Выполнить такое распределение вручную достаточно сложно, при этом все равно могут быть не учтены способности студентов, так как этот процесс является достаточно трудоемким и длительным. Ручное распределение большого объема данных может привести к случайным ошибкам из-за человеческого фактора. Вследствие этого возникает необходимость автоматизации данного процесса. Таким образом, в рамках настоящей работы поставлен следующий вопрос: как быстро и функционально можно распределить студентов по группам, имея их личностные портреты?

Статья построена следующим образом. В разделе 1 приведены ключевые характеристики личностного портрета студента. В разделе 2 раскрыта возможность уточнения цифрового портрета с помощью анализа данных социальных сетей. В разделах 3 и 4 представлены разработанные методы и инструменты распределения студентов по академическим группам и курсам по выбору. В разделе 5 рассмотрена возможность построения проектных команд.

1. ЛИЧНОСТНЫЙ ПОРТРЕТ ОБУЧАЮЩЕГОСЯ

В течение всего времени обучения в образовательном учреждении происходит сбор большого объема информации о каждом обучающемся. К этой информации относятся:

- данные анкеты – ФИО, дата и место рождения, семейное положение, статус семьи и др.;
- физические особенности, группа здоровья;
- перечень и оценки по предметам за предыдущие семестры, а также (если необходимо) за предметы старших классов средней школы;
- информация о профильной лаборатории/кафедре, куда обучающийся распределен;
- выполненные курсовые работы и проекты за время учебы;

- иные полученные результаты, достижения, освоенные компетенции (сюда могут относиться сертификаты онлайн-курсов, выигранные конкурсы и т. п.).

Собираются эти данные могут из анкеты и личного кабинета обучающегося во внутреннем портале образовательного учреждения, часть данных может собираться вручную администрацией учебного заведения, некоторую часть данных обучающийся может представлять о себе сам. В рамках данной работы нет необходимости четко описывать весь объем извлекаемых данных, поскольку решается только определенная задача распределения.

2. СБОР ДАННЫХ ИЗ СОЦИАЛЬНОЙ СЕТИ

Анализ, проведенный в рамках исследования, показал отсутствие открытых инструментов, которые дают возможность высшему учебному заведению получать информацию об абитуриентах и студентах, размещенную на их страницах в социальных сетях, которые могут быть источниками дополнительной информации для формирования личностного портрета студента [3]. Сотрудники высших учебных заведений имеют возможность собирать эту информацию вручную, но для этого требуется большое количество времени, напрямую зависящее от количества абитуриентов и студентов. Для автоматизации этого процесса нами было принято решение разработать инструмент сбора информации из социальных сетей.

Такой инструмент был разработан для социальной сети ВКонтакте; была использована библиотека VK API с открытым исходным кодом [4], представляемая разработчиками ВКонтакте. VK API – это программный интерфейс, который позволяет получать информацию из базы данных vk.com с помощью HTTP-запросов к специальному серверу. Для использования VK API не нужно подробно знать, как строится база данных, из каких типов таблиц и полей она состоит, – достаточно, чтобы API-запрос «знал» об этом. Синтаксис запросов и тип возвращаемых данных строго определены на стороне службы.

На данный момент инструмент позволяет найти следующую информацию: «имя» и «фамилия», информацию из разделов «о себе», «интересы», «образование», «карьера», «любимые книги». Также есть возможность узнать информацию о записях, размещенных студентами и абитуриентами на их страницах, и о группах и сообществах, членами которых они являются.

Пример запроса с использованием VK API на языке Java:

```
List<UserXtrCounters> students = vk.users().get(actor).userIds(studentIds).fields(UserField.ABOUT,  
    UserField.BDATE,  
    UserField.INTERESTS,  
    UserField.SEX,  
    UserField.HOME_TOWN,  
    UserField.SCHOOLS,  
    UserField.CAREER,  
    UserField.BOOKS)  
    .execute();
```

Пример результата запроса VK API, приходящий в формате JSON¹ (звездочками скрыты персональные данные):

```
{  
  "id": "*****",  
  "name": "***** *****",  
  "birthDate": "**.*.*.*****",  
  "interests": "",  
  "sex": "MALE",  
  "hometown": "Казань",  
  "school": "Лицей №***;",  
  "books": "Э. Хемингуэй, Э. М. Ремарк",  
  "groups": [  
    {"name": "че",  
     "activity": "Юмор"},  
    {"name": "Новые Альбомы",  
     "activity": "СМИ"}  
  ]  
}
```

Для работы с библиотекой VK API необходимо зарегистрировать разработанное приложение в социальной сети «ВКонтакте». Инструмент возвращает данные в формате JSON, поэтому их можно использовать в других инструментах.

3. РАСПРЕДЕЛЕНИЕ СТУДЕНТОВ ПО АКАДЕМИЧЕСКИМ ГРУППАМ

Вопросы эффективного распределение студентов по академическим группам в рамках высшего образования ставятся довольно часто. Необходимо учитывать целую серию факторов, при этом разные образовательные организации могут по-разному формировать критерии хорошего распределения. В данной работе рас-

¹ JavaScript Object Notation

смаатриваются требования Высшей школы информационных технологий и интеллектуальных систем Казанского (Приволжского) федерального университета (ИТИС КФУ), где:

- группы должны делиться по двум потокам для двух языков программирования;
- в каждой группе должен быть примерно одинаковый уровень по ЕГЭ и английскому языку;
- должны быть по возможности учтены все пожелания студентов к совместной учебе, если это не приведет к серьезному нарушению ограничений, названных выше.

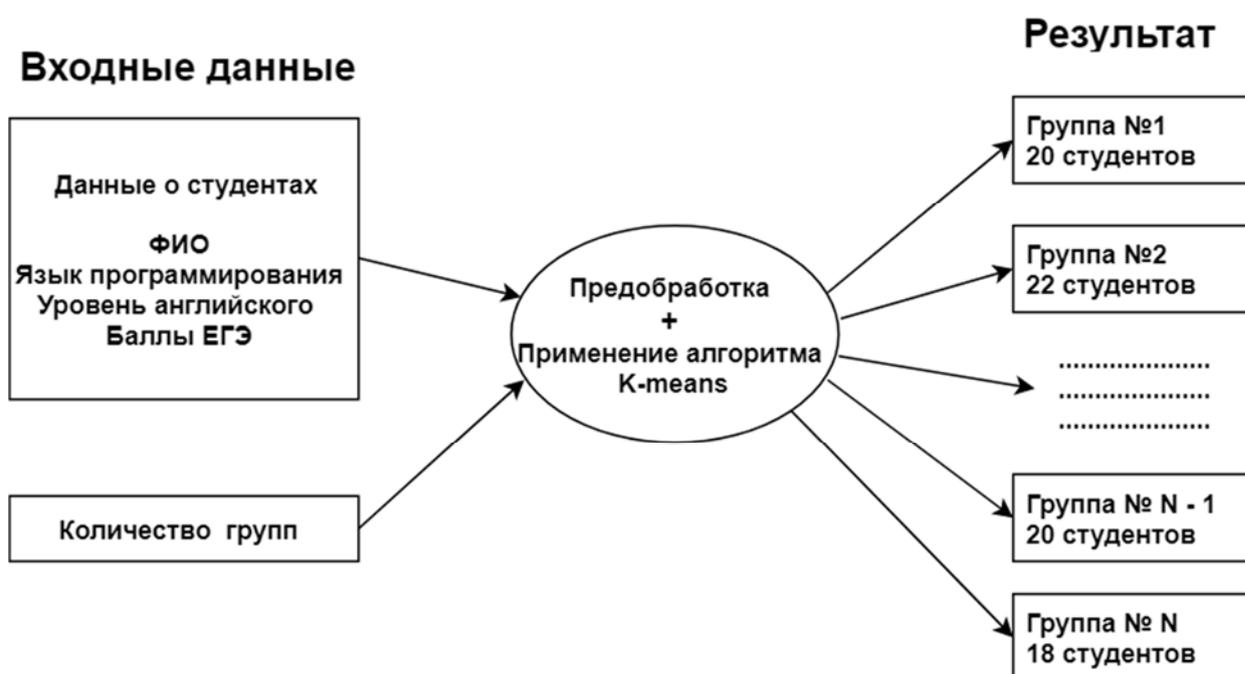


Рис. 1. Распределение студентов на академические группы

Для решения данной задачи был создан инструмент, где в качестве входных параметров о студенте указываются: ФИО, баллы по ЕГЭ, уровень английского языка и выбранный язык программирования (рис. 1). Алгоритм распределения использует кластеризацию [5] методом *K-means* [6] (k-средних), основная идея которого заключается в разбиении множества элементов на несколько кластеров и минимизации среднеквадратического отклонения элементов кластера от его «центра масс».

4. РАСПРЕДЕЛЕНИЕ СТУДЕНТОВ ПО ЭЛЕКТИВАМ

Еще одним видом распределения является разбиение студентов на курсы по выбору или элективы – предметы, которые студенты выбирают самостоятельно для собственного развития. Выбор должен быть последовательным, т. е. выбранный предмет должен сочетаться с текущими компетенциями студента, иначе освоение предмета может быть неэффективным.

В качестве входных данных для данного модуля используются данные как о студентах, так и о курсах по выбору. Основой созданного алгоритма распределения является кластеризация C-means [6], которая позволяет определить вероятность отношения к каждому классу вместо однозначного позиционирования объекта в один кластер. Также преимуществами данного алгоритма являются распределение объекта, основанное на нескольких его признаках, и возможность определения количества и параметров центров кластеров [7] (в качестве центров используются курсы по выбору и информация о них: название, преподаватель, описание курса, данные о студентах, которые окончили данный курс). После окончания работы алгоритма пользователю возвращается матрица принадлежности [8] объектов к каждому кластеру. Схема работы данного алгоритма представлена на рис. 2.

На основе полученной матрицы принадлежности для каждого студента выбираются наиболее подходящий курс, а также список «возможных» курсов, т. е. тех курсов, которые также с достаточно большой вероятностью должны быть полезны данному студенту.

Следующим этапом является формирование групп студентов на курсы по выбору. Существует несколько вариантов разбиения:

- для каждого студента выбирается наиболее подходящий курс;
- помимо наиболее подходящих курсов используется список «возможных» курсов.

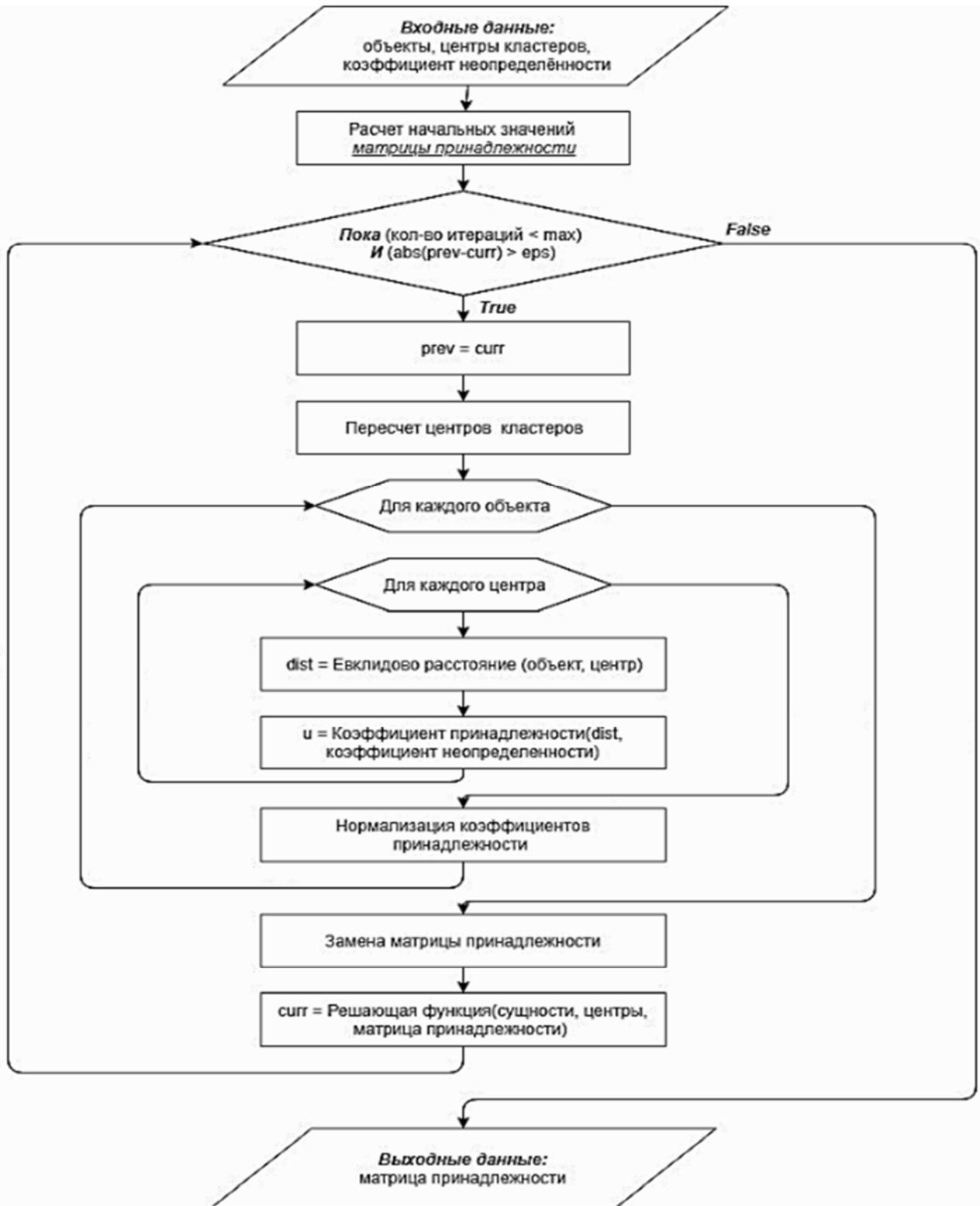


Рис. 2. Кластеризация данных для распределения студентов по элективам

При первом варианте разбивки группы получаются неравномерными, на некоторые курсы попадает большое количество студентов, а на других курсах

студентов достаточно мало. Второй вариант предоставляет возможность сделать группы более равными по количеству, при этом личностные параметры студентов также учитываются. Данный алгоритм распределения предполагает перемещение студентов, для которых существует несколько подходящих курсов, из групп с наибольшим количеством студентов в другие группы из списка «возможных», пока такое перераспределение выполнимо.

5. ПОСТРОЕНИЕ ПРОЕКТНОЙ РАБОТЫ

Частью личностного портрета студента являются набор компетенций, которыми он владеет, и их уровень. На основе этого набора компетенций может быть поставлена задача распределения студентов по проектным командам для выполнения учебного или научного проекта. При этом распределении нужно учитывать как компетенции, которыми участник проекта уже должен владеть, так и компетенции, которые ему необходимо развивать.

На рис. 3 представлена схема проектной работы, где работа выстраивается вокруг ролей, в которых нужно обладать определенным уровнем компетенций и развивать его.

Для решения задачи распределения по проектным командам был реализован специальный инструмент. Входными данными для него являются параметры личностных портретов студентов. Нужно подать на вход csv файл, содержащий информацию о студентах в формате: группа, ФИО, список компетенций. Преподаватель может распределить группу студентов на набор проектных команд. При создании одной проектной команды преподаватель имеет возможность задать ему параметры:

- название проекта;
- курс;
- список ролей в проекте;
- список компетенций, требуемых для выполнения этих ролей;
- требуемые и выходные уровни владения компетенциями.

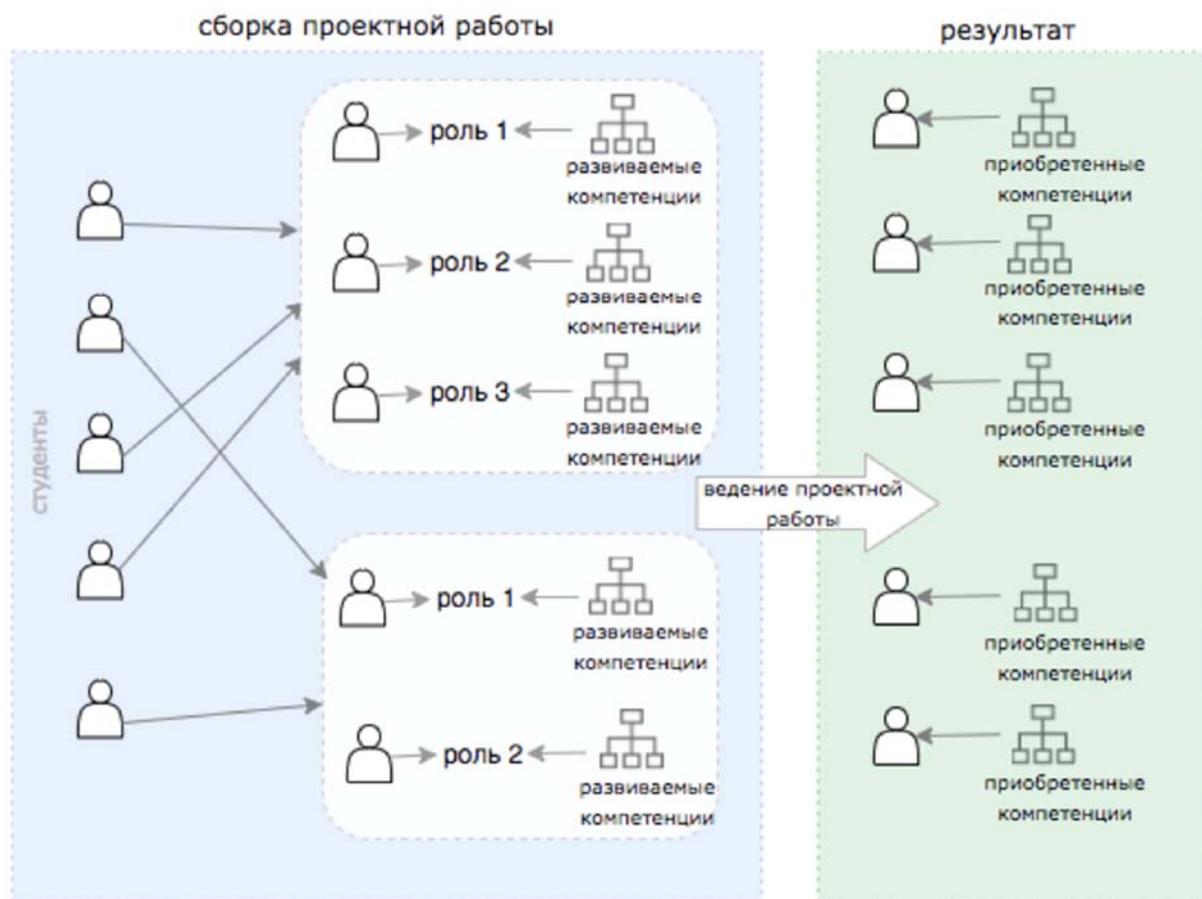


Рис. 3. Схема построения проектной работы

В силу разных целей и критериев распределения инструмент был спроектирован с некоторой долей адаптивности. Было реализованы несколько типов распределения:

- *Распределение студентов по схожим компетенциям.* Это распределение предполагает создание команд, в каждой из которых все студенты должны обладать как минимум одной схожей компетенцией. Для этого необходимо подсчитать, какое количество команд будет выделяться под ту или иную компетенцию путем вычисления пропорции. Затем происходит добавление студентов в команды, компетенциями которых они обладают. Схема работы данного алгоритма представлена на рис. 4.

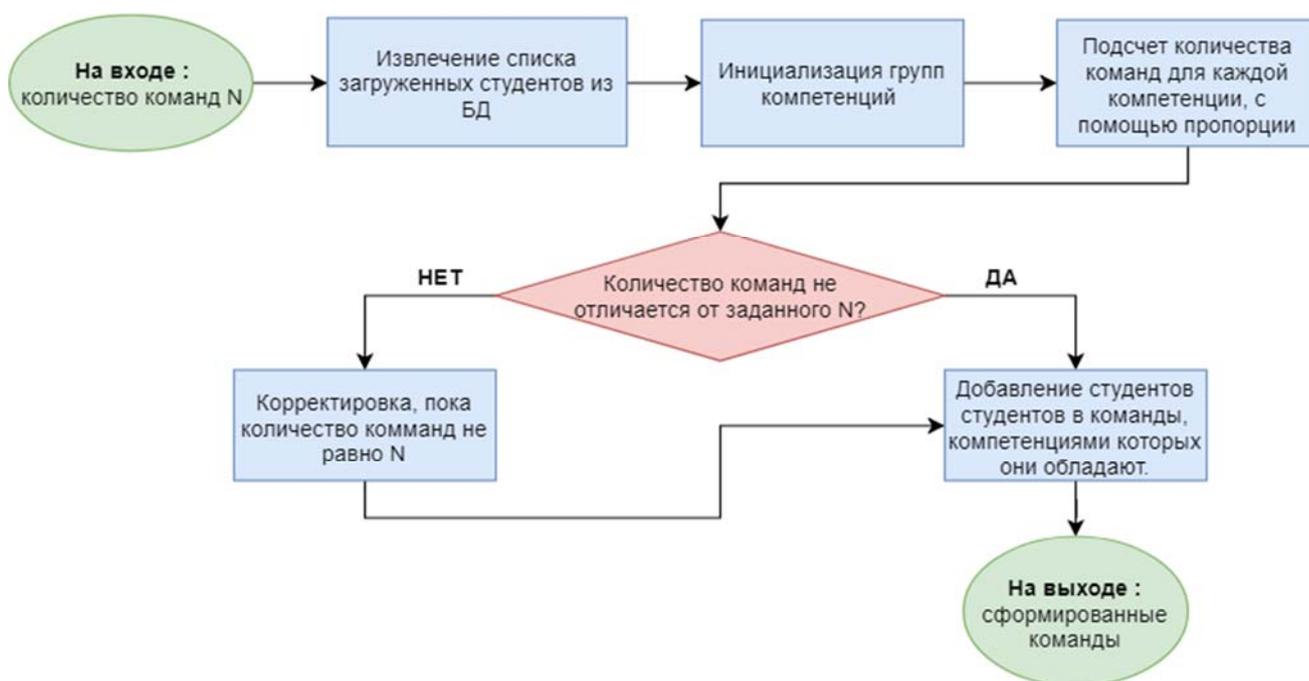


Рис. 4. Схема алгоритма распределения студентов по схожим компетенциям

- *Распределение студентов по различным компетенциям.* В данном распределении предоставляется возможность задать параметры для каждой компетенции или выбрать автоматическую настройку. Параметр для одной компетенции принимает на вход число, которое обозначает количество людей, обладающей данной компетенцией, как минимум входящих в состав каждой команды.

Начальный этап включает инициализацию групп компетенций и добавление в каждую из них студентов, обладающих единственной компетенцией.

Второй этап подразумевает добавление студентов в группы компетенций, пока не будет достигнута *квота* данной компетенции. Квота вычисляется умножением количества команд на параметр пользователя для данной компетенции. Если все квоты, то есть минимальные требования пользователя, достигнуты, то идет переход к последнему этапу, в противном случае необходимо выполнить обмен студентов между группами компетенций. На рис. 5. представлен пример обмена, таким образом можно выполнить обмен не только между двумя студентами, но и составить цепочки из некоторого числа студентов.

До обмена



После обмена

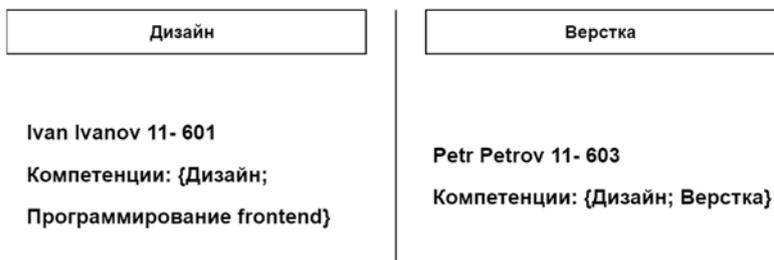


Рис. 5. Пример обмена между студентами

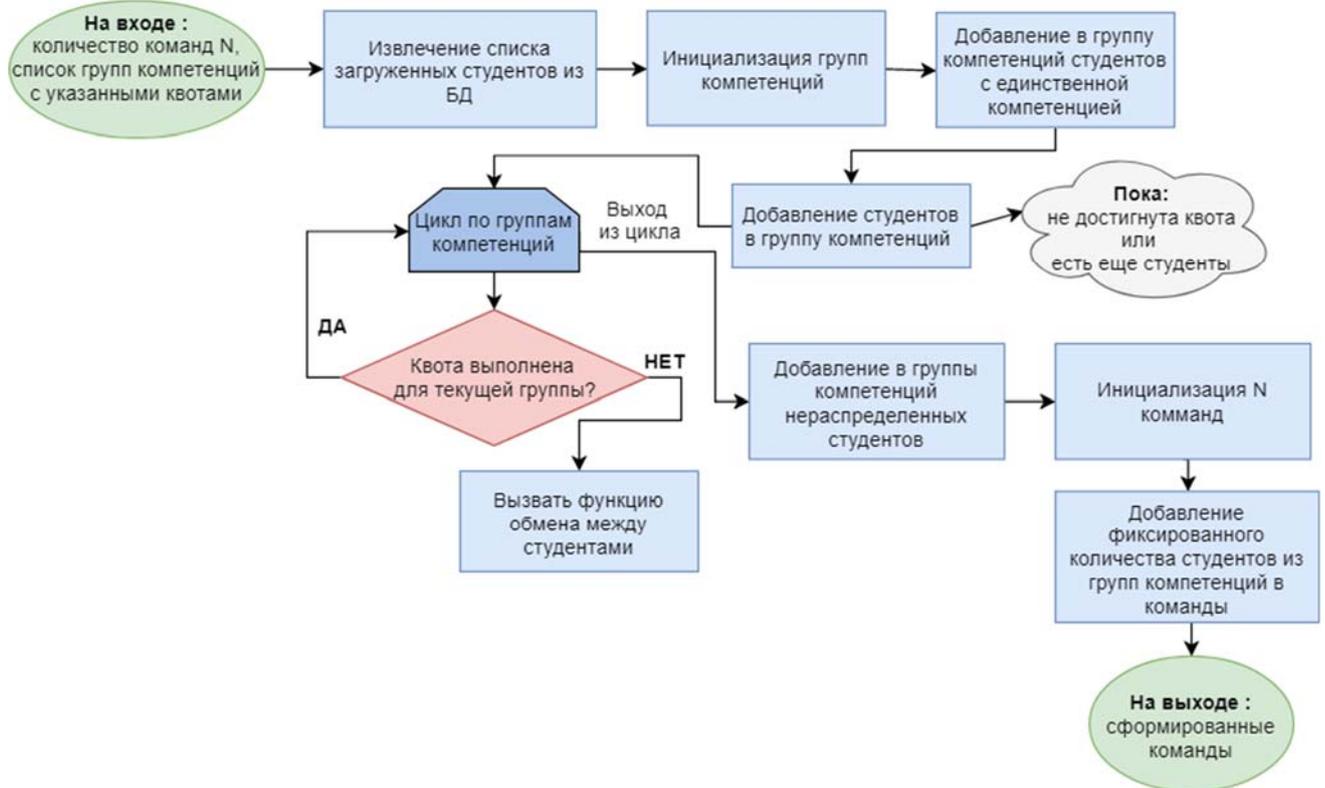


Рис. 6. Блок-схема алгоритма распределения студентов по различным компетенциям

Последний этап включает добавление оставшихся студентов в группы компетенций и далее генерацию команд на основе данных групп. Вся вышеописанная схема алгоритма представлена на рис. 6.

- *Распределение случайным образом.* В зависимости от того, какой набор требуемых компетенций и уровней указал преподаватель, генерируется команда из студентов, наиболее подходящих по заданным параметрам. Для каждой роли указываются компетенции, которые будут развиваться в рамках проектной работы. Также для каждой компетенции вносится значение минимального уровня владения, без которого участие в проекте бессмысленно.

При формировании сущности «проекта» может учитываться также и иерархичность используемых в нем компетенций: преподаватель может указать требуемые входные уровни для всех компетенций, являющихся родительскими по отношению к выбранной. По итогам проектной работы, в случае успешной сдачи проекта, студенты расширяют перечень освоенных компетенций, что может быть отмечено в их личностном портрете.

ЗАКЛЮЧЕНИЕ

Выполнен анализ данных, необходимых для различных видов распределения по группам. Исследованы методы сбора данных и их разбивки, разработаны модели и инструменты для распределения студентов по академическим группам и проектным командам. Созданный инструмент апробирован при распределении студентов по академическим группам, поступивших в 2017 году в Высшую школу ИТИС КФУ.

В дальнейшем планируется расширение возможностей данного инструмента под конкретные академические задачи. Планируется апробировать полученный инструмент в рамках проектной работы студентов.

СПИСОК ЛИТЕРАТУРЫ

1. Указ Президента Российской Федерации № 204 от 07.05.2018 г. «О национальных целях и стратегических задачах развития Российской Федерации на период до 2024 года». URL: <http://kremlin.ru/acts/news/57425>
2. *Селевко Г.К.* Энциклопедия образовательных технологий. Том 1. М.: Народное образование, 2005. С. 224–225.

3. Семенов А.В., Веретенник Е.В., Пронин А.С. Формирование учебных групп в университете с помощью анализа социальных сетей // Вопросы образования. 2014. № 3. С. 18–21.

4. Официальный сайт проекта VK API. URL: <https://vk.com/dev>

5. Королев М.А. Статистический словарь. М.: Финансы и статистика, 1989. 623 с.

6. Котов А., Красильников Н. Кластеризация данных, 2006. URL: <https://logic.pdmi.ras.ru/~yura/internet/02ia-seminar-note.pdf>

7. Нейский И.М. Классификация и сравнение методов кластеризации // Интеллектуальные технологии и системы. Сб. учебно-методических работ и статей аспирантов и студентов. М.: НОК «CLAIM», 2006. Вып. 8. С. 130–142.

8. Тараскина А.С. Нечеткая кластеризация по модифицированному методу С-средних и ее применение для обработки микрочиповых данных // Проблемы интеллектуализации и качества систем автоматизации. 2013. № 13. С. 217–228.

GENERATION OF ACADEMIC GROUPS AND PROJECT TEAMS BASED ON LEARNERS DATA ACQUISITION

N. A. Korgutlova¹, S. Y. Basargina², M. M. Abramskiy³, M. A. Solncev⁴,
T. S. Buzukina⁵

¹⁻⁵ Higher School for Information Technologies and Intelligent Systems of Kazan (Volga Region) Federal University

¹korgutlova97@gmail.com, ²basargina.svetlana@gmail.com, ³ma@it.kfu.ru,
⁴mrt.solncev@gmail.com, ⁵taya.buzukina@gmail.com

Abstract

The questions of usage of the learners' data in the solutions for generating student academic groups, electives and project teams are considered. The applications of Machine Learning clustering algorithms for these tasks are illustrated. The opportunity of usage of social network data is shown.

Keywords: *personal portrait of student, clustering, competence distribution, social networking analysis*

REFERENCES

1. Decree of President of Russian Federation # 204 from 07.05.2018. "O natsional'nykh tselyakh i strategicheskikh zadachakh razvitiya Rossijskoj Federatsii na period do 2024 goda". URL: <http://kremlin.ru/acts/news/57425>
2. *Selevko G.K.* Entsiklopediya obrazovatel'nykh tekhnologij. Tom 1. M.: Narodnoe obrazovanie, 2005. S. 224–225.
3. *Semenov A.V., Veretennik E.V., Pronin A.S.* Formirovanie uchebnykh grupp v universitete s pomoshh'yu analiza sotsial'nykh setej // Voprosy obrazovaniya. 2014. № 3. S. 54–57.
4. Ofitsial'nyj sajt proekta VK API. URL: <https://vk.com/dev>
5. *Korolev M.A.* Statisticheskij slovar'. M.: Finansy i statistika, 1989. 623 s.
6. *Kotov A., Krasil'nikov N.* Klasterizatsiya dannykh, 2006. URL: <https://logic.pdmi.ras.ru/~yura/internet/02ia-seminar-note.pdf>
7. *Nejskij I.M.* Klassifikatsiya i sravnenie metodov klasterizatsii // Intel'ktual'nye tekhnologii i sistemy. Sbornik uchebno-metodicheskikh rabot i statej aspirantov i studentov. M.: NOK «CLAIM», 2006. Vypusk 8. S. 130–142.

8. *Taraskina A.S.* Nechetkaya klasterizatsiya po modifitsirovannomu metodu S-srednikh i ee primenenie dlya obrabotki mikrochipovykh dannyykh // Problemy intellektualizatsii i kachestva sistem avtomatizatsii. 2013. № 13. S. 217–228.

СВЕДЕНИЯ ОБ АВТОРАХ



КОРГУТЛОВА Наталья Александровна – студентка Высшей школы информационных технологий и интеллектуальных систем Казанского (Приволжского) федерального университета.

Natalya Aleksandrovna KORGUTLOVA – student of the Higher School of ITIS KFU.

email: korgutlova97@gmail.com



БАСАРГИНА Светлана Юрьевна – студентка Высшей Школы Высшей школы информационных технологий и интеллектуальных систем Казанского (Приволжского) федерального университета.

Svetlana Yurievna BASARGINA – student of the Higher School of ITIS KFU.

email: basargina.svetlana@gmail.com



АБРАМСКИЙ Михаил Михайлович – старший преподаватель кафедры программной инженерии Высшей школы информационных технологий и интеллектуальных систем Казанского (Приволжского) федерального университета.

Mikhail Mikhailovich ABRAMSKIY – senior lecturer at Department of Software Engineering of Higher School of ITIS KFU

email: ma@it.kfu.ru



СОЛНЦЕВ Марат Альбертович – студент Высшей школы информационных технологий и интеллектуальных систем Казанского (Приволжского) федерального университета.

Marat Albertovich SOLNCEV – student of the Higher School of ITIS KFU.

email: mrt.solncev@gmail.com



БУЗУКИНА Таисия Сергеевна – студентка Высшей школы информационных технологий и интеллектуальных систем Казанского (Приволжского) федерального университета.

Taisiya Sergeevna BUZUKINA – student of the Higher School of ITIS KFU.

email: taya.buzukina@gmail.com

Материал поступил в редакцию 31 июля 2018 года