

СИСТЕМА ПОДБОРА РАЗРАБОТЧИКОВ ДЛЯ ВЫПОЛНЕНИЯ ЗАДАЧ НА КРУПНЫХ ПРОЕКТАХ

Туманов Д.Е.

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Анисимов В.Я. – к.ф.-м.н., доцент

В докладе рассматриваются вопросы создания системы подбора разработчиков для выполнения задач на крупных проектах, способ обработки текстов задач.

Интерес к профессиям, связанным с информационными технологиями, в последние годы заметно вырос — это видно, например, по тому, как увеличилась доля запросов в поисковых системах по таким ключевым словам как разработчик, разработка, программист или программирование. В 2019 году она была на четверть больше, чем в 2017. Спрос на ИТ-специалистов растёт, за два года доля объявлений о работе для них от всех объявлений на HeadHunter увеличилась на 5,5%. С ростом интереса и количества разработчиков растут и ИТ-компании. За 2019 год белорусский офис EPAM нанял 1600 технических специалистов. IBA Group – 200 технических сотрудников, данный показатель чуть выше, на 100 человек, чем у компании Wargaming. Лидером в процентном приросте штата стала Synesis: по сравнению с 2018 годом — +210%, с 35-й на 11-ую строчку рейтинга сайта dev.by. Но с ростом компаний и их технического штата растут также и затраты на управление, анализ работы, подбор команд из специалистов различных технических направлений. Для этого и было создано рассматриваемое приложение, цель которого – минимизировать расходы на подбор технического специалиста для решения проектных задач.

В основе приложения, а именно в обработке задач, лежат идеи обработки естественного языка и глубоких нейронных сетей.

Направление обработки естественного языка на данный момент является одним из важнейших в исследованиях в области искусственного интеллекта. Это направление зародилось с появлением первых вычислительных машин и с идеей об их использовании для решения задач, связанных с естественным языком, на котором разговаривают и пишут люди. Одной из первых стояла задача машинного перевода – автоматического перевода текста с одного языка на другой с помощью компьютера или вычислительной машины. Для пары «русский-английский» были выполнены первые успешные эксперименты по машинному переводу еще в середине 1950-х годов. Вторая задача – создание диалоговых систем, способных вести с человеком диалог на естественном языке. В те же годы появились первые подобные программы и системы. Третья задача – создание вопросно-ответных систем, способных точно отвечать на вопрос, заданные человеком. Вопросы при этом обычно представляются в форме текста на естественном языке. В последнее время круг задач по обработке естественного языка стал намного шире.

Для обработки текста задачи и подбора разработчиков по данному критерию было принято решение использовать глубокие нейронные сети с рекуррентными и сверточными слоями.

Рекуррентная нейронная сеть представляет отдельный вид нейронных сетей, где связи между элементами образуют направленную последовательность, благодаря чему появляется возможность обрабатывать серии событий во времени или последовательные пространственные цепочки. Главное отличие данного вида нейронных сетей от многослойных перцептронов – рекуррентные сети могут использовать свою внутреннюю память для обработки последовательностей произвольной длины. Это позволяет применять данные сети в таких задачах, как распознавание рукописного текста или распознавание речи, где целостная структура входных данных разбита на части. Применяя сверточные слои в качестве начальных слоев в рекуррентных нейронных сетях при обработке текста, можно добиться достаточного представления входных данных, встраивая его в вектор фиксированной длины, который можно использовать для дальнейших задач. Сегодня всё чаще можно увидеть повсеместное использование сверточных и рекуррентных слоёв в рамках одной нейронной сети.

Пользователь может ввести информацию о задаче несколькими способами:

- ввод текста задачи;
- ввод ссылки на сторонний ресурс с текстом задачи (в данный момент поддерживается только ресурсы из Jira – распространенный инструмент управления проектами);

При обучении модели нейронной сети был достигнут результат в 91%. При попытке использовать для подбора разработчика на основе текста задачи более простые инструменты, например многоклассовую логистическую регрессию с предварительной векторизацией текстов, максимальный из полученных результатов составил 65%, поэтому и было принято решение об использовании глубоких нейронных сетей.

Последующими шагами в развитии системы будет улучшения качества модели, а также написание дополнительных инструментов для более удобного способа ввода задачи в систему (например браузерные расширения и т.д.).

Список использованных источников:

1. Rogerson J. Theory, Concepts and Methods of Recurrent Neural Networks and Soft Computing // Clanrye International – 2015. – P. 254-270.
2. Hasim S. Long Short-Term Memory recurrent neural network architectures for large scale acoustic modeling – 2014. – P. 2–3.
3. Graves A., Liwicki M., Fernandez S., Bertolami R., Bunke H., Schmidhuber J. A novel connectionist system for unconstrained handwriting recognition // IEEE Transactions – 2009. – Vol. 31 – P. 855–868,
4. Dean J., Corrado G., Monga R., Chen K., Devin M., Le Q. V., Mao M. Z., Ranzato M., Senior A. W., Tucker P. A., Yang K., Ng A. Y. Large scale distributed deep networks. // NIPS – 2012. – P. 1232–1240.