

Рецензия на доклад “Методы определения геолокации мобильных устройств” аспиранта 2-го года обучения Шестакова Андрея.

В докладе Андреем освещались различные способы определения местоположения мобильных устройств. Обсуждались достоинства и недостатки каждого из них. Основное внимание было уделено методу “FingerPrinting”, позволяющему по степени интенсивности сигнала базовых станций достаточно точно и быстро определить местоположение устройства. Для вычисления местоположения метод использует предварительно собранные данные о значениях силы сигнала в каждой территориальной единице. Каждой территориальной единице соответствует m -мерный вектор ($m = 20$) вещественных чисел (FingerPrint) — значений силы сигнала соответствующих m базовых станций. При определении местоположения устройства пользователя с устройства снимается значение силы сигнала m станций, строится m -мерный вектор q , после чего в базе производится поиск k наиболее близких векторов к вектору q . Местоположение определяется как усредненное значение координат территориальных единиц, соответствующих k ближайшим векторам к q .

В качестве недостатка данного метода докладчиком отмечалась высокая вычислительная сложность процесса поиска ближайших m -мерных векторов в базе, содержащей порядка 85 миллионов векторов.

Андреем был предложен алгоритм, позволяющий ускорить метод FingerPrint и увеличить его точность. Предложенный алгоритм заключается в том, чтобы на этапе поиска ближайших векторов к q поиск производить, сначала используя только первую компоненту: отбирается множество кандидатов со значением первой компоненты, равной первой компоненте q . Далее поиск производится среди множества кандидатов, но уже учитывая вторую компоненту. Процесс повторяется до тех пор, пока множество кандидатов не сократится до размера k . Этот алгоритм был заявлен в качестве научного результата.

Главным недостатком данного доклада и в целом проделанной работы является то, что метод, предложенный автором является одним из методов поиска k ближайших соседей. Проблема поиска k ближайших, точнее, проблема построения эффективного метода для поиска k ближайших, в литературе известна с 70-х годов XX века и активно изучается в настоящее время. Однако докладчиком данный момент был полностью упущен. Не было произведено никакого сравнения с современными методами. Вполне вероятно, что применение одного из современных методов к данной проблеме могло бы существенно повысить скорость работы даже в сравнении с предложенным методом автора. В заключение советую автору ознакомиться в качестве «отправного пункта» со статьей в википедии https://en.wikipedia.org/wiki/Nearest_neighbor_search, а также обратить внимание на библиотеку методов поиска ближайших соседей <https://github.com/searchivarius/nmslib>, которая содержит state-of-the-art методы и наборы данных для сравнения эффективности методов.

Александр Пономаренко 30.06.2016