

Тема научно-исследовательской работы: «Разработка алгоритмов выявления подозрительных аккаунтов путем анализа данных социальных сетей»

Цель работы: разработать алгоритм обнаружения потенциально опасных страниц в социальных сетях, как можно более эффективно использующий доступные цифровые и текстовые данные пользователя.

1 Основные положения выбранной математической модели

Дискриминантный анализ – метод статистического характера, позволяющий изучать различия между несколькими группами объектов (классами) по нескольким различным переменным (признакам). Кроме того, для построения модели дискриминантного анализа необходимо:

- наличие двух или более классов;
- существование ни менее двух объектов в каждом классе;
- наличие хотя бы одной дискриминантной переменной.

Введем обозначение канонической дискриминантной функции F_{km} следующим образом:

$$F_{km} = u_0 + \sum_{i=1}^p u_i X_{ikm},$$

где X_{ikm} – значение i -го признака m -го объекта в классе k , а u_i – набор коэффициентов, обеспечивающий выполнение следующего требования: средние значения канонической дискриминантной функции для различных классов должны как можно сильнее отличаться друг от друга, причем значения разных канонических дискриминантных функций должны быть некоррелированными между собой.

Для того, чтобы оценить, насколько некоторый признак способен разделить объекты на классы, обозначим критерии разделимости, основывающиеся на отношении матриц рассеяния. Матрица внутригруппового рассеяния (рассеяния внутри классов) показывает разброс

векторов объектов относительно векторов математических ожиданий классов. Обозначим ее за W и будем вычислять ее элементы по формуле:

$$w_{ij} = \sum_{k=1}^g \sum_{m=1}^{n_k} (x_{ikm} - \bar{x}_{ik})(x_{jkm} - \bar{x}_{jk}), \quad i, j = \overline{1, p},$$

где n – количество элементов, n_k – количество элементов в классе k , g – количество классов разбиения, p – количество признаков, x_{ikm} – значение i -го признака для m -го элемента в классе k , а \bar{x}_{ik} – среднее значение i -го признака в классе k , которое вычисляется по формуле:

$$\bar{x}_{ik} = \frac{1}{n_k} \sum_{m=1}^{n_k} n_k \bar{x}_{ikm}.$$

Матрица межгруппового рассеяния (рассеяния между классами), может быть определена разными способами, обозначим ее за B , и будем вычислять ее элементы по формуле:

$$b_{ij} = \sum_{k=1}^g n_k (\bar{x}_{ik} - \bar{x}_i)(\bar{x}_{jk} - \bar{x}_j), \quad i, j = \overline{1, p},$$

где \bar{x}_i – среднее значение i -го признака по всем классам, вычисляемое по формуле:

$$\bar{x}_i = \frac{1}{n} \sum_{k=1}^g n_k \bar{x}_{ik}.$$

Для того, чтобы получить меру разделимости классов, определим критерий разделимости $J(Q)$, где Q – пространство признаков. Критерий $J(Q)$ увеличивается при увеличении рассеяния между классами или уменьшении внутригруппового рассеяния. Эти условия выполняются при вычислении его по различным формулам, выражающим требуемое отношение. Воспользуемся наиболее универсальной формулой, обеспечивающей инвариантность относительно любого невырожденного линейного преобразования:

$$J_1(Q) = tr(S_2^{-1}S_1),$$

Для вычисления критерия J_1 , обычно используют матрицы $S_1 = B$, $S_2 = T = W + B$, где T – матрица рассеяния смеси, элементы которой можно вычислять по формуле:

$$t_{ij} = \sum_{k=1}^g \sum_{m=1}^n (x_{ikm} - \bar{x}_i) (x_{jkm} - \bar{x}_j), \quad i, j = \overline{1, p}.$$

Эта матрица показывает рассеяние точек относительно среднего значения, но не внутри класса, а среди всех векторов, и, чтобы не производить лишних вычислений, ее можно вычислять, с помощью формулы $T = B + W$. Кроме того, можно преобразовать матрицу T в матрицу коэффициентов корреляции R путем деления каждого элемента на корень произведения соответствующих диагональных элементов. Стоит отметить, что если совпадают центры различных классов, тогда соответственно среднее внутриклассовое будет равняться среднему общему, и соответственно $T = W$.

2 Признаки

Для исходных данных воспользуемся страницами из социальной сети “ВКонтакте”. Выберем 50 страниц пользователей, таким образом, чтобы был хотя бы один объект на каждый признак определения потенциально опасной страницы. Множество всех используемых объектов обозначим, как Ω - 50 страниц. Разобьем множество Ω на 2 класса посредством разбиения $\Omega = \{\Omega_1, \Omega_2\}$. Где Ω_1 - предположительно безопасные страницы, а Ω_2 - потенциально опасные.

Поскольку это разбиение, то соответственно, можно отметить выполнение двух свойств:

– $\Omega_1 \cup \Omega_2 = \Omega$ – все объекты принадлежат либо одному, либо другому классу;

– $\Omega_1 \cap \Omega_2 = \emptyset$ - если объект принадлежит одному классу, то он не принадлежит другому.

Исходный набор признаков, получим, используя программные средства python для расчета значений каждой дискриминантной функции для каждого исследуемого объекта. Выберем следующие признаки для расчета:

1. Имя пользователя - входит или нет в пакет русских имен.
2. Возраст пользователя.
3. Длительность регистрации. Количество дней, прошедших с момента регистрации.
4. Уникальность дат загрузки пользовательских данных. Отношение суммы количества загруженных фотографий и постов на стене к количеству уникальных дат загрузки данных.
5. Количество использований фотографии на различных ресурсах. Определяется с помощью поиска по изображениям в системе Google.
6. Наличие ссылки на сторонний ресурс в верхней записи на стене.
7. Наличие ссылки на сторонний ресурс в статусе пользователя.

Данный набор признаков является предварительным и может легко расширяться.

3 Оценка эффективности признаков

Для определения эффективности вклада признаков в классификацию - рассчитаем значение критерия J для классификации по каждому признаку. Выведем информацию о признаках в порядке убывания их индивидуальной значимости: среднее по классу, разброс по классу, значение критерия и вероятность ошибки классификации. Классификацию будем производить с помощью линейного дискриминантного анализатора.

Таким образом, получим информацию о значимости признаков и произведем подтверждение или опровержение выдвинутых гипотез о правильности использования данных для определения потенциально опасных страниц.

Для оценки вероятности ошибки классификации будем пользоваться методом исключения одного объекта.

Пусть, имеются N объектов. Для каждого объекта произведем следующие действия:

1. Исключим объект из множества.
2. По имеющимся $N - 1$ объектам синтезируем классификатор.
3. Классифицируем исключенный объект.

Допустим, верно классифицировано K объектов, тогда оценка вероятности ошибки классификации будет вычисляться с помощью формулы:

$$\varepsilon = \frac{K}{N} 100\%,$$

а эффективность классификации (точность) с помощью формулы:

$$E = (1 - \varepsilon) 100\% = \left(1 - \frac{K}{N}\right) 100\%.$$

Произведем оценку вероятности ошибки классификации для исходного множества объектов.

После произведения вычислений на исходном множестве, можно заменить исходные признаки новыми, путем замены численных значений признаков на бинарные. Таким образом, например, при подтверждении гипотезы о том, что подозрительный пользователь произвел регистрацию в

социальной сети в течение последних 5 дней - можно заменить численное значение количества дней с момента регистрации на логические 0 и 1 (страница старше 5 дней или нет соответственно). Таким образом, заменив численные значения на бинарные, и соответственно, запомнив границы разделения классов на исходном множестве объектов, произведем вычисления оценки вероятности ошибки классификации на новом множестве объектов.

В том случае, если на новом множестве объектов, по установленным границам делимости классов, оценка вероятности ошибки классификации близка к оценке вероятности ошибки классификации на исходном множестве объектов, можно сделать вывод о том, что данные социальных сетей анализируются эффективно. Соотношение оценок вероятности ошибок классификации на исходном и новом множествах покажет эффективность анализа данных.