

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Томский государственный университет
Горно-Алтайский государственный университет
Институт оптики атмосферы им. В.Е. Зуева СО РАН

НОВЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ИССЛЕДОВАНИИ СЛОЖНЫХ СТРУКТУР

**МАТЕРИАЛЫ ДЕСЯТОЙ РОССИЙСКОЙ КОНФЕРЕНЦИИ
С МЕЖДУНАРОДНЫМ УЧАСТИЕМ**

Томск
Издательский Дом Томского государственного университета
2014

- решать задачу оптимизации экспертных затрат на проверку портфеля И-признаков/И-событий в системе событийной безопасности;
- вычислять информационную емкость И-признака безопасности данного предприятия, значения которого, оцениваемые экспертом, определяют выполнение/невыполнение нормативного требования безопасности.

Интерпретация информационной характеристики И-события как информационная емкость И-признака безопасности заключается в том, что если вероятность И-события близка к 0 или к 1, то экспертная проверка его наступления дает мало дополнительной информации, и его включение в экспертный портфель не слишком оправдано. Критерием отбора И-признака является значительная величина его информационной емкости, превосходящая пороговое значение, которое определяется требуемой точностью оценки риска опасного события.

Литература

1. Воробьев О.Ю. Эвентология. Красноярск : Сибирский федеральный университет, 2007. 435 с.
2. Воробьев О.Ю. Совокупная система и совокупность систем: эвентологическое сходство и различие // Труды XI международной конференции ФАМЭБ'2012 конференции. Красноярск : СФУ, 2012. С. 131–138.
3. Воробьев О.Ю. Эвентологический системный анализ безопасности // Труды XI международной конференции ФАМЭБ'2012 конференции. Красноярск : СФУ, 2012. С. 113–125.
4. Воробьев О.Ю. Эвентологический анализ систем: система событий во внесистемных обстоятельствах // Труды XI международной конференции ФАМЭБ'2012 конференции. Красноярск : СФУ, 2012. С. 126–130.

ОЦЕНИВАНИЕ КАЧЕСТВА ОБУЧАЕМЫХ КЛАССИФИКАТОРОВ ПО F -МЕРЕ В УСЛОВИЯХ ОТСУТСТВИЯ КОНТРОЛЬНЫХ ВЫБОРОК¹

А.С. Пехтерев, В.В. Поддубный

Национальный исследовательский Томский государственный университет, Томск, Россия
alexandr.pehterev@gmail.com, vvpoddubny@gmail.com

Рассматривается проблема оценки качества работы обучаемых классификаторов произвольного типа, работающих с объектами, описываемыми набором признаков как в непрерывном метрическом пространстве, так и в дихотомическом пространстве. Если сверх обучающей выборки объектов, принадлежность которых к рассматриваемым классам известна, имеется достаточное количество объектов, необходимых для оценки качества классификации (контрольная выборка), можно провести эмпирический анализ вероятностей ошибок классификации и вычислить, например, F -меру качества классификации (среднее гармоническое между полнотой – долей объектов, правильно приписываемых классу из всех объектов этого класса, и точностью – долей объектов, правильно приписываемых классу из всех объектов, приписываемых этому классу). Проблема возникает в случае существенного дефицита или полного отсутствия контрольных выборок объектов из заданных классов, необходимых для оценки качества работы классификатора. Выход может быть найден на пути компьютерного моделирования (имитации) объектов, обладающих статистическими свойствами объектов обучающей выборки.

В работе [1] рассмотрен один из подходов к решению этой задачи для объектов с признаками из непрерывного метрического пространства. Суть подхода состоит в использовании гауссовых копула-функций, позволяющих преобразовать эмпирические маргинальные распределения признаков обучающей выборки каждого класса в многомерные нормальные распределения, благодаря чему путём обратного преобразования из гауссовых многомерных векторов, генерируемых датчиком случайных чисел с математическими ожиданиями, дисперсиями и корреляционными матрицами, соответствующими обучающим выборкам каждого класса, можно получить многомерные векторы признаков объектов тех же классов, имитирующие объекты этих классов и допускающие использование их в качестве недостающих контрольных выборок.

В данной работе эта идея используется для имитации объектов, описываемых в дихотомическом (в том числе бинарном) признаковом пространстве. Проблема здесь состоит в том, что дихотомические признаки относятся к номинальной шкале, и наблюдениям поддаются только частоты их появления в

¹ Работа поддержана грантом РФФИ № 13-07-00373-а.

выборках. Поэтому вопросы сходства и различия объектов, корреляционной связи между их признаками и т.п. могут решаться по-разному [2]. Мы предлагаем следующий подход. Сопоставляя номинальным дихотомическим признакам числовые упорядоченные значения (например, 0 и 1), можно привести номинальные дихотомические признаки к дискретному числовому виду, для которого уже будут иметь смысл интегральные функции распределения и их числовые характеристики (математические ожидания, дисперсии, корреляционные матрицы и т.д.). Статистические оценки первых двух моментов векторов признаков объектов получаются из эмпирических частот появления в классах нулей и единиц. Применяя к эмпирическим распределениям таких признаков объектов обучающей выборки идею построения гауссовых копула-функций аналогично тому, как это делалось в случае непрерывных признаков, можно построить алгоритм генерации бинарных дихотомических объектов со статистическими свойствами обучающих выборок требуемых классов и провести по ним оценку качества классификации по F -мере. Соответствующий алгоритм оценки качества работы классификаторов реализован программно и проверен на ряде примеров решения задач тестового распознавания бинарных объектов.

Литература

1. Поддубный В.В., Пехтерев А.С. Копулы сглаженных эмпирических распределений при наличии связей (совпадений) и их применение в имитационном моделировании // Труды XII международной ФАМЭБ'2013 конференции / под ред. О. Воробьева. Красноярск : НИИППБ, СФУ, 2013. С. 312–321.
2. Ким Дж.-О., Мьюллер Ч.У., Клекка У.Р. и др. Факторный, дискриминантный и кластерный анализ. М. : Финансы и статистика, 1989. 215 с.

НЕПРЕРЫВНАЯ СТОХАСТИЧЕСКАЯ ДИНАМИЧЕСКАЯ МОДЕЛЬ РАЗВИТИЯ ПОЛИСЕМИИ АНСАМБЛЯ ЗНАКОВ ЕСТЕСТВЕННОГО ЯЗЫКА И ЕЁ ИДЕНТИФИКАЦИЯ

В.В. Поддубный¹, А.А. Поликарпов²

¹ Национальный исследовательский Томский государственный университет, Томск, Россия

² Московский государственный университет им. М.В. Ломоносова, Москва, Россия

vpoddubny@gmail.com; anatpoli@mail.ru

Рассматривается непрерывная стохастическая математическая модель развития полисемии языковых знаков. В соответствии с современными лингвистическими представлениями об эволюции и жизненном цикле языкового знака естественного языка языковой знак (слово или словосочетание) возникает в языке первоначально в некотором единственном, обычно весьма конкретном смысловом значении и с течением времени приобретает дополнительные, обычно всё более абстрактные значения (полисемию), пока не исчерпает свой так называемый ассоциативно-семантический потенциал (АСП) и не достигнет своей максимально возможной многозначности. Этот процесс идёт с постоянным замедлением. Одновременно, но с некоторой задержкой во времени, протекает аналогичный, но значительно более медленный процесс выпадения из употребления первоначальных, относительно более конкретных значений знака. Этот процесс продолжается до тех пор, пока не выйдет из употребления последнее (наиболее абстрактное) значение, и тогда жизненный цикл языкового знака заканчивается, знак выходит из употребления. В координатах «полисемия – время (возраст)» жизненный цикл языкового знака представляется, таким образом, унимодалной кривой с максимумом, смещённым к началу процесса эволюции знака.

В работе [1] нами предложена и исследована дискретная диссипативная стохастическая динамическая математическая модель развития ансамбля языковых знаков с целочисленными значениями полисемии. Поток знаков предполагался стационарным пуассоновским, развитие полисемии знаков описывалось конечными нестационарными дважды стохастическими потоками с постоянно убывающими интенсивностями, пропорциональными остаточному целочисленному АСП, флуктуирующему от знака к знаку по показательному закону распределения. Задержка начала процесса выпадения из употребления значений знака принималась случайной с экспоненциальным распределением вероятностей. Путём имитационного моделирования процесса развития полисемии ансамбля знаков по рассмотренной схеме определялось одномоментное эмпирическое (модельное) распределение полисемии, которое сравнивалось с распределением полисемии, полученным из представительных толковых словарей русского и английского языков. Для согласования этих распределений постоянные времени процессов роста и спа-