



Е.М. Пудикова

АНАЛИЗ ПОТОКА ОШИБОК В ПРОМЫШЛЕННОЙ СИСТЕМЕ С ПОМОЩЬЮ НЕЙРОНЕЧЁТКОГО КЛАССИФИКАТОРА И БАЙЕСОВСКОЙ СЕТИ ВЫВОДА

(ФГБОУ ВПО «Самарский государственный аэрокосмический университет
им. академика С.П. Королева (национальный исследовательский университет)»)

Современные промышленные системы требуют постоянного мониторинга, так как в сегодняшних условиях веб фреймворк не просто приложение, работающее на единственном сервере – инстансе, но комплексная распределенная система, работающая в неоднородных условиях. В таких условиях системы могут подвергаться нагрузкам, превышающие предусмотренные разработчиками. Промышленная система состоит из нескольких интегрированных продуктов, которые должны работать согласованно, однако часто подобные интеграции дают сбой.

В описанной ситуации ошибки в системе, могут возникать не только как единичные примеры, но образовывать потоки ошибок, которые нуждаются в анализе и корректировке, для того, чтобы вернуть систему в работоспособное состояние. Ввиду этого, данную проблему можно решить при помощи разработки системы, которая будет способна классифицировать ошибки, поступающие в неё на анализ, и на основе результатов классификации определять вероятное решение проблемы, подходящее для данной ошибки.

В качестве входных данных система должна использовать набор сообщений на английском языке, содержащий описание произошедшего ошибочного события, и требующий мониторинга. В качестве выходного результата система должна определять к какому из 5 основных классов сообщений: NCDefect; OtherDefect; ManualTask; AutoTask; InfoTask; относится данная ошибка.

Таким образом, основной задачей системы является определение семантики сообщения, при этом, его семантическим смыслом будет являться класс сообщения. В данной работе для решения подобной задачи исследуются два подхода: байесовская сеть и нейронная нечёткая продукционная сеть. В качестве входных данных использовался реальный поток ошибок, сгенерированный фреймворком в процессе его работы. Этот набор данных был проанализирован экспертами, и на основе их выводов были получены следующие результаты:

1. Отобраны сообщения, которые содержат семантические признаки, на основе которых можно определить соответствующий класс сообщения. Генерируемая выборка сообщений была очищена от нечетких сообщений, которые эксперты не могли корректно классифицировать.

2. Построены семантические правила, позволяющие точно классифицировать сообщения. Эти семантические правила будут использоваться в модели байесовской сети классификации.



3. Определена опытным путем стратегия обучения, позволяющая быстро формировать семантические правила. Шаг обучения должен составлять не более трех примеров. Для наискорейшего обучения требуется подавать на вход первыми сообщения, наиболее точно подходящие данному классу, и лишь затем подавать все более и более нечеткие примеры. Это нужно для того, чтобы центры классов при машинном обучении были определены как можно быстрее.

4. Определено, что полнота семантических правил зависит от статистики представляемых примеров. Все эксперты, при анализе сначала применяют сформированное ими семантическое правило, и, если оно не дало корректного результата, обращаются к правилам, основанным на статистике, а затем корректирует семантическое правило до тех пор, пока оно не даст корректный результат.

В данной работе байесовская сеть используется в качестве системы, осуществляющей вывод на основе сформированных экспертных правил, которые будут отражены в структуре сети в виде вероятностных оценок. Примером такого правила является следующее: если сообщение содержит любой синоним слов «ошибка» или «поломка», и объектами действия или контекста являются сущности, принадлежащие NC, то это NCDefect. В терминах байесовской сети описанное правило будет представляться узлами предпосылок (ошибки и объекты NC), их совместное появление будет означать увеличение вероятности следствия в виде класса NCDefect. Результаты работы байесовской сети будут сравниваться с работой нечеткого нейроклассификатора для проверки полноты правил, формируемых нейронной сетью. В качестве используемой нейронечеткой модели будет использоваться модификация нечеткой сети Ванга-Менделя. Ее схема представлена на рисунке 1.

Выбор данной модели связан с тем, что в процессе обучения такой сети, формируются нечеткие правила, основанные на группе нечетких предпосылок. Каждый слой этой сети будет иметь несколько выходных нейронов, связанных с соответствующим классом. В ней будет два нечетких слоя. Первый слой содержит выходные нейроны, которые будут представлять синонимы слов, входящих на вход сети. По сути, синоним – это класс слова, приводящий к одному и тому же выводу, при применении к нему семантического правила. Вторым слоем в качестве входных нейронов будет использовать выходные нейроны предыдущего слоя. А выходами сети будут являться искомые 5 классов. Алгоритм обучения для данной сети будет гибридным и состоящим из двух фаз:

- 1) определения положения центров нечетких правил и их числа;
- 2) уточнения нечетких правил с помощью алгоритма обратного распространения ошибки.

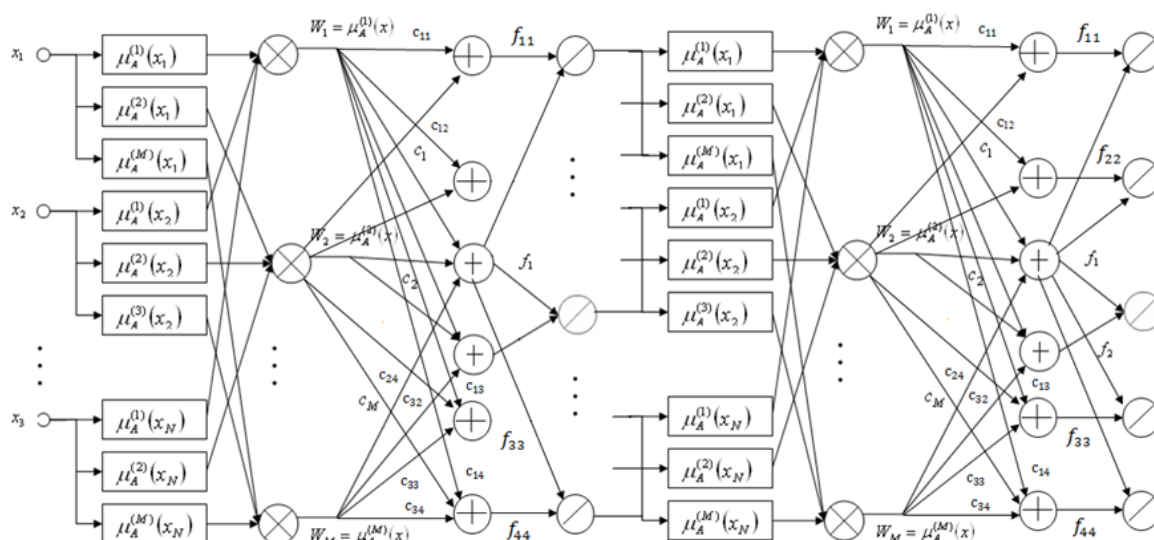


Рис. 1. Схема модели сети Ванга-Менделя

Оба алгоритма, участвующие в фазах обучения такой сети, модифицированы в соответствии с построенной моделью. В качестве функции нечеткой предпосылки используется функция Гаусса, так как она подходит для решения задачи семантического анализа в соответствии с мерой закона Ципфа, который говорит, что распределение семантически значимых слов по отношению к определенному семантическому классу определено по нормальному закону распределения.

На вход нейронечеткому классификатору подаются сообщения, отображенные в числовое пространство с помощью хэш-функции. При этом, на вход каждого нейрона приходит единственное преобразованное слово. Все слова очищены от морфологических признаков, для того, чтобы облегчить процесс анализа сети.

Литература

1. Круглов В.В., Дли М.И., Голунов Р.Ю. Нечеткая логика и искусственные нейронные сети. – М.:ФИЗМАТЛИТ, 2001. – 224 с.
2. Новак В., Перфильева И., Мочкорж И. Математические принципы нечеткой логики. – М.:ФИЗМАТЛИТ, 2006. – 347 с.
3. Андреев А.М., Березкин Д.В., Морозов В.В. Автоматическая классификация текстовых документов с использованием нейросетевых алгоритмов и семантического анализа. – Материалы НПП «Интеллект плюс»