



OSTIS-2013

(Open Semantic Technologies for Intelligent Systems)

УДК 004.822:514

ОБОБЩЕННАЯ МОДЕЛЬ ЭМОЦИИ И ПРОЦЕССА АНАЛИЗА РЕЧЕВОГО СИГНАЛА В СИСТЕМЕ БЕЗОПАСНОСТИ

Пекарь Д. В. *, Садов В. С. *

**Белорусский Государственный Университет*

Минск, Беларусь

pekar.dima@gmail.com

sadov@bsu.com

Эмоции являются естественным проявлением работы психо-физиологической деятельности нервной системы человека, а речь одним из наиболее естественных способов их проявления, что представляет большой интерес для своего изучения и использования в речевых системах. Эффективная автоматическая обработка речевых сигналов предполагает под собой строгую формализацию предмета анализа, которая может поддаваться не только качественной оценке, но и количественной параметризации.

Ключевые слова: модель, эмоциональная окраска, семантический анализ, обработка речи.

ВВЕДЕНИЕ

Эмоции являются неотъемлемой частью жизнедеятельности человека, которые формируются вследствие психо-физиологической деятельности его нервной системы. Принимая данный факт во внимание, можно положить, что эмоции имеют непосредственную связь с его внутренним состоянием. Согласно существующим теориям, эмоции имеют сложное происхождение, а именно "эмоции – это психофизиологический феномен, поэтому о возникновении переживания человека можно судить как по самоотчету человека о переживаемом им состоянии, так и по характеру изменения вегетативных показателей (частоте сердечных сокращений, артериальному давлению, частоте дыхания и т. д.) и психомоторики: мимике, пантомимике (позе), двигательным реакциям, голосу" [1]. Таким образом, речь является одним из наиболее естественных способов проявления эмоций и отражением эмоционального состояния человека. Анализ эмоциональной окраски речевого сигнала имеет огромное значение для систем распознавания и синтеза речи [2-3], call-центрах [4], диалоговых системах [5-6], системах безопасности, который подразумевает использование некой формализованной модели эмоции.

1. Формализованная модель эмоции в системе анализа эмоциональной окраски речи

Для построения эффективной системы анализа эмоциональной окраски речевого сигнала

необходимо формализовать понятие эмоции – определить ее модель как математический объект, который обладает определенными параметрами и свойствами. Подобный подход позволяет реализовать в дальнейшем процесс анализа эмоционального состояния диктора по его речи в виде последовательности методов обработки и анализа входного речевого образца.

В процессе разработки модели эмоции для анализа эмоциональной окраски речи были выработаны следующие наиболее важные критерии, которым она должна удовлетворять. И так, разрабатываемая формализованная модель должна быть оптимальной в информационном смысле и эффективной с вычислительной точки зрения; должна обладать полнотой, т. е. содержать все необходимые параметры для классификации эмоциональной окраски речи и оценки величины потенциальной опасности; должна быть настраиваемой под те или иные условия использования

Предложенная формализованная модель эмоции, изображенная на рисунке 1.

Основным параметром представленной модели является X – вектор признаков, полученный на основе цифровой обработки речевого сигнала, который содержит информацию о частотном и энергетическом спектре сигнала, значениях таких параметров как частоты формант и основного тона, их статистических параметров, скорости изменения значений характеристик речевого сигнала, а также их модуляции.

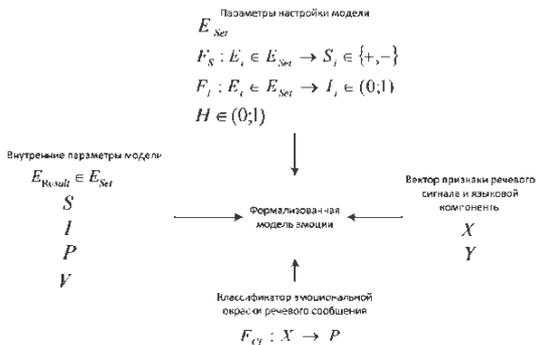


Рис. 1 – Обобщенная формализованная модель эмоции

Параметр Y представляет собой вектор признаков, полученный на основе применения семантического [7] и сентиментного [8,9] анализа языковой информации. С помощью данного параметра возможно извлечение дополнительной информации об эмоциональной окраске сообщения, которая выражена с помощью языко-знаковых систем в общем случае, или тем или иным языком. Таким образом, рассмотренные параметры X и Y предложенной модели эмоции позволяют учитывать при анализе речевого сигнала, как вербальную его составляющую, так и невербальную компоненту, что дает возможность осуществления более полного и комплексного анализа высказывания человека.

Параметр P представляет собой вектор вероятностей такой, что для его элементов выполняется следующее равенство:

$$\sum_{i=1}^{|E_{Set}|} P_i = 1, \quad (1)$$

где элементы вектора P отражают вероятность принадлежности речевого фрагмента к соответствующему типу эмоции из предопределенного множества релевантных эмоций E_{Set} . Подобный подход к описанию эмоции позволяет учесть ее фундаментальную природу – эмоциональная окраска не может быть интерпретирована строго и однозначно. При подобном описании эмоции, элементы вектора P представляют собой координаты данной эмоции в многомерном пространстве, где единичный базисный вектор соответствует отдельной эмоции. Размерность вектора P равна мощности множества возможных эмоций E_{Set} .

Функциональная зависимость $F_{Cl} : X \rightarrow P$ осуществляет отображение значений вектора признаков речевого сигнала X на вектор вероятностей P . В качестве такого функционального отображения применяется классификатор, который строится на основе обучающей выборке данных. Использование построенного классификатора в качестве самостоятельного параметра объясняется тем, что изначально избыточный входной вектор признаков, а так же и параметры используемого

классификатора могут быть оптимизированы под заданного диктора или имеющийся набор речевых образцов для достижения наилучшей точности определения типа эмоции [10]. Подобный подход дает дополнительное преимущество, а именно позволяет придать адаптивность и гибкость предлагаемой модели, т. е. использование различных векторов признаков на стадии определения типа эмоции без изменения структуры входных данных.

Следующим параметром предложенной модели является вес отдельной эмоции I , который отражает степень ее влияния на оценку величины потенциальной опасности. Данный параметр позволяет учитывать тот факт, что различные эмоции могут иметь различный вес и ассоциироваться с различными уровнями потенциальной опасности. Значение параметра I задается априорно для каждой эмоции E_i из допустимого множества E_{Set} , согласно требуемым параметрам работы системы:

$$F_I : E_i \in E_{Set} \rightarrow I_i \in (0; 1), \quad (2)$$

Параметр модели S соответствует полярности эмоции. Согласно современной трактовке эмоций, положительный и отрицательный типы эмоций принимаются как противоположные и человек, в отдельный взятый момент, может находиться либо в радостном, либо в грустном эмоциональном состоянии [1]. Данное утверждение можно считать лишь условно истинным, поскольку деление эмоций на положительные и отрицательные условно, а также наличие сложных эмоций согласно некоторым исследователям [11] противоречит данному высказыванию. Для устранения двойственности интерпретации полярности эмоции полагается, что под положительной эмоцией подразумевается та эмоция, которая способствует повышению конструктивного поведения человека. Соответственно, под отрицательной эмоцией будет подразумеваться та эмоция, которая располагает к деструктивному поведению индивида. Соотнесение каждой допустимой эмоции из множества E_{Set} и ее полярности устанавливается априорно, что отражено соотношением в представленной модели:

$$F_S : E_i \in E_{Set} \rightarrow S_i \in \{+, -\}, \quad (3)$$

Полярность эмоции необходима для оценки величины потенциальной опасности.

Поскольку при анализе эмоциональной окраски речи используется два 'информационного канала' поступления данных – результаты цифровой обработки речевого сигнала и семантической / сентиментной обработки, то это дает возможность варьирования степени информативности каждого из каналов. Для реализации данного свойства модель содержит параметр H , который определяет степень информативности результатов цифровой обработки речевого сигнала при принятии решения о величине

потенциальной опасности. Соответственно, результаты семантической обработки языковой информации имеют вес $1-H$. С помощью параметра H можно 'отключать' влияние семантического и сентиментного анализа языковой информации в случае низкого качества или невозможности распознавания произнесенной речи, или же использовать только семантический / сентиментный анализ в случае анализа текстовой информации, что доставляет дополнительную универсальность и гибкость предлагаемой модели.

Параметр $E_{Result} \in E_{Set}$ соответствует типу эмоции, который присваивается речевому фрагменту после его анализа. Возможный тип эмоции ограничен predetermined множеством E_{Set} релевантных для рассмотрения/выявления типов эмоций.

С каждым входным речевым фрагментом ассоциируется характеристическая величина потенциальной опасности, полученная на основе анализа речевого высказывания, что отражено в соответствующем параметре модели V .

В представленной модели инкапсулирован необходимый набор данных, достаточный для классификации эмоций и оценки величины потенциальной опасности, а также настроечные параметры, которые позволяют настраивать модель под учет конкретных условий ее использования.

2. Модель процесса анализа речевого сообщения в системе безопасности

На основе предложенной модели эмоции разработан следующий процесс обработки речевого сообщения в системе безопасности выявления потенциально опасных ситуаций, модель которого изображена на рисунке 2.

Как видно из схемы, модель процесса анализа речевого сообщения имеет многоуровневую структуру, каждый уровень которой имеет свои задачи по анализу и обработке входных данных.

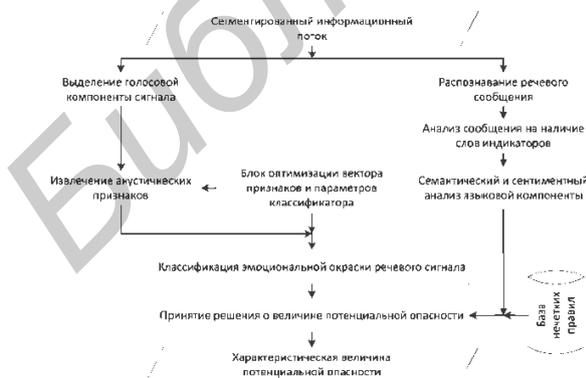


Рис. 2 – Модель процесса анализа речевого сообщения

Первым уровнем представленной модели является уровень предварительной обработки речевого сообщения. На данном уровне осуществляется цифровая обработка

сегментированного сигнала с целью повышения его качества, а также распознавание произнесенной речи и представления ее в текстовом виде для дальнейшей обработки и анализа. Упомянутый уровень находится за рамками рассмотрения представленной работы, поскольку в настоящее время существуют программные средства, которые позволяют решать задачу распознавания речи с приемлемой точностью [12]. Разрабатываемая система анализа эмоциональной окраски речи подразумевает использование исходного речевого сигнала для анализа его эмоциональной окраски и текстового сообщения, которое является распознанной версией произнесенной фразы или выражения для получения языковой компоненты сообщения.

Этап предварительной обработки в предлагаемой модели процесса анализа речевого сообщения позволяет осуществлять 2-х режимный вариант работы. Первым режимом работы является режим сканирования информационного потока путем поиска только априорно заданных слов-индикаторов, которые могут указывать на наличие потенциальной опасности. При обнаружении хотя бы одного predetermined слова-индикатора потенциальной опасности, система генерирует соответствующее сообщение. Подобный режим работы системы целесообразен в случае работы в реальном масштабе времени при анализе высокоинтенсивных потоков речевой информации и наличии ограниченных вычислительных ресурсов для их обработки. Неизбежным следствием снижения вычислительных ресурсов является возможность пропуска целевой информации, если она передается с использованием синонимии языка. Для осуществления детального анализа информационного потока система должна работать во втором режиме с использованием семантического графа для поиска целевого содержания.

Следующим уровнем представленной модели процесса анализа эмоциональной окраски является вычисление вектора признаков. Поскольку разрабатываемая система позволяет анализировать как речевой сигнал, так и языковую составляющую речевого сообщения, то на данном этапе происходит вычисление двух векторов признаков. Первый вектор признаков состоит из акустических признаков – значений различных параметров речевого сигнала, которые вычисляются с помощью методов цифровой обработки сигналов. Второй вектор признаков состоит из семантических признаков речевого сообщения, которые вычисляются с помощью методов семантического анализа. Особенностью данного уровня рассматриваемой модели анализа речевого сообщения является то, что после вычисления вектора акустических признаков происходит удаление некоторых его компонент. Процесс осуществляется на основе данных, которые получены блоком оптимизации входного вектора

признаков, что позволяет исключить малоинформативные элементы вектора признаков и тем самым оптимизировать процесс классификации эмоциональной окраски речевого сообщения путем уменьшения количества используемых признаков.

Третьим уровнем предложенной модели анализа речевого сообщения является уровень классификации эмоциональной окраски. Входными данными для него являются вектора признаков с акустическими и семантическими признаками, которые подаются на вход классификатора эмоциональной окраски. Результатом работы классификатора является вектор вероятностей P . Важным преимуществом предложенной модели сообщения является то, что применяемый классификатор является обучаемым, то есть он проходит настройку с помощью обучающей выборки данных. Применение подобного подхода позволяет реализовать более гибкий процесс анализа речевого сообщения, который оптимизируется под значения входных данных. Дополнительным преимуществом предложенной модели является то, что настроечные параметры классификатора вычислены с учетом входных данных и оптимизированы под использование найденного вектора признаков, при котором используемый классификатор работает наилучшим образом.

Заключительным уровнем рассматриваемой модели является уровень принятия решения об уровне потенциальной опасности, которой включает в себя два этапа. На первом этапе принимается решение о финальном типе эмоции и его знаке, к которому будет отнесен исследуемый речевой фрагмент. Основой выбранного решения является вектор вероятностей P , набор нечетких правил R и конфигурационная настройка, связывающая каждый тип эмоции с соответствующей ей полярностью и заданным весом.

Заключение

Предложенная модель позволяет перейти от качественной характеристики человеческих эмоций к их количественной параметризации и численной оценки.

Подробно рассмотрев модель эмоции и процесса анализа речевого сообщения можно отметить такие преимущества как гибкость и универсальность, т. е. модель может быть настроена под определенные требования; параметры модели могут быть адаптированы под используемые данные для достижения лучших результатов классификации эмоционального состояния.

Предложенный процесс анализа речевого сигнала в системе безопасности позволяет получить количественную оценку уровню потенциальной опасности на основе комплексного анализа, как

речевой компоненты, так и языковой составляющей речевого сообщения.

Библиографический список

- [Ильин, 2001] Ильин Е. П. Эмоции и чувства / Е. П. Ильин. – СПб: Питер, 2001. – 752 с.
- [Schuller, 2007] Towards More Reality In The Recognition of Emotional Speech : proceedings of the 32nd IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). – 2007. – P. 941-944.
- [Schröder, 2001] Emotional Speech Synthesis: A Review: proceedings of the 7th European Conference on Speech Communication and Technology. – 2001.
- [Petrushin, 1999] Emotion in speech: Recognition and application to call centers : proceedings of the conference Artificial Neural Networks in Engineering (ANNIE'99). – 1999.
- [Fragopanagos, 2005] Fragopanagos N. Emotion recognition in human-computer interaction / N. Fragopanagos, J.G. Taylor // Neural Networks. – 2005. – No. 18. – P. 389-405
- [Yacoub, 2003] Recognition of Emotions in Interactive Voice Response Systems : proceedings of the 8th European Conference on Speech Communication and Technology. – 2003. – p. 1-4.
- [Yih, 2012] Polarity Inducing Latent Semantic Analysis: proceedings of the conference on empirical methods in natural language processing and computational natural language learning. – 2012.
- [Wilson, 2005] Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis: proceedings of HLT-EMNLP. – 2005. – P. 347-354.
- [Hernández, 2009] Recognizing Polarity and Attitude of Words in Text: proceedings of the 14th Portuguese Conference on Artificial Intelligence. – 2009.
- [Пекарь, 2012] Пекарь Д. В. Алгоритм оптимизации входных признаков и параметров классификатора эмоционального состояния человека по его речи / Д. В. Пекарь, С. Г. Тихоненко, А. Ф. Чернявский, В. С. Садов // Информатика. – 2012. – № 3(35).
- [Розалиев, 2008] Розалиев В. Л. Предпосылки, возможности, перспективы создания автоматизированной системы распознавания эмоциональности речи // Известия Волгоградского государственного технического университета: межвузовский сборник научных статей. – 2008. – №2(40). – С.58-61.
- [Walker, 2004] Walker W. Sphinx-4: a flexible open source framework for speech recognition / W. Walker, P. Lamere, P. Kwok, B. Raj, R. Singh, E. Gouvea, P. Wolf, J. Woelfel // Sun Microsystems. Inc. – 2004.

GENERALIZED MODEL OF HUMAN EMOTION AND PROCESS FLOW OF SPEECH SIGNAL ANALYSIS IN SECURITY SYSTEM

Pekar D. V. *, Sadov V. S. *

* *Belarusian State University,
Minsk, Republic of Belarus*

pekar.dima@gmail.com

sadov@bsu.com

In presented paper formalized model of human emotion and process flow of speech signal analysis is described. Proposed model can formalize human emotion and parameterize its values. Also it handles all required attributes for further automatic processing. Presented model has advantages such as ability to be customized to fit specific needs, could be adapted to incoming data without changes in input interface. Described process flow allows to estimate level of potential danger based on analysis of voice and lingual component of human speech.