

УДК 621.397

ОЦЕНКА ЗНАЧИМОСТИ АКУСТИЧЕСКИХ ПРИЗНАКОВ В ЗАДАЧЕ ДЕТЕКТИРОВАНИЯ ГОЛОСОВОЙ АКТИВНОСТИ

Кравцов С.А., аспирант кафедры динамики электронных систем Ярославского государственного университета им. П.Г. Демидова, e-mail: sk860@outlook.com;

Топников А.И., к.т.н., ассистент кафедры динамики электронных систем Ярославского государственного университета им. П.Г. Демидова, e-mail: sk860@outlook.com;

Приоров А.Л., д.т.н., доцент кафедры динамики электронных систем Ярославского государственного университета им. П.Г. Демидова, e-mail: andcat@yandex.ru.

THE ASSESSMENT OF THE SIGNIFICANCE OF ACOUSTIC FEATURES IN THE TASK OF DETECTING VOICE ACTIVITY

Kravtsov S.A., Topnikov A.I., Priorov A.L.

The significance value of the spectral flatness measure, mel-frequency cepstral coefficients and gammatone-frequency cepstral coefficients for the voice activity detection was evaluated. To evaluate the significance of features we used a method which based on decision trees and comparing the values of the areas under ROC-curves (AUC) for support vector machine (SVM), the perceptron and the method of nearest neighbors classifiers (k -nearest neighbors). The accuracy of the classification in the case of all analyzed coefficients and the most significant features was evaluated.

Keywords: speech signal, feature importance, voice activity detector, noise.

Ключевые слова речевой сигнал, значимость признака, детектор голосовой активности, шум.

Введение

Сегодня сложно представить современную систему цифровой обработки речевых сигналов без детектора голосовой активности (ДГА, Voice Activity Detector, VAD). Также в литературе ДГА встречается под названием «детектор речевой активности» (ДРА). Его применяют для определения наличия речи и выделения границ речесодержащих фрагментов в обрабатываемом сигнале. Благодаря применению ДГА, увеличивается точность работы алгоритмов идентификации и верификации диктора [1-4], распознавания голосовых команд [5]. При этом задача разработки устойчивого алгоритма ДГА в случае наличия внешних шумов является актуальной и в настоящее время. В процессе разработки такого алгоритма необходимо выбрать как классификатор, принимающий окончательное решение о наличии или отсутствии речи, так и наиболее важные акустические признаки сигнала, т.е. такие признаки, используя которые достигалось бы наибольшее значение верной классификации фрагмента сигнала – речи или паузы.

Целью работы является разработка и тестирование методики выделения наиболее значимых акустических признаков речевого сигнала в задаче детектирования голосовой активности.

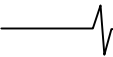
Оценка значимости признаков

Для оценки значимости выбраны следующие акустические признаки: мера спектральной плоскостности (МСП, Spectral Flatness Measure, SFM) [6], мел-частот-

Проведена оценка значимости меры спектральной плоскостности, мел-частотных кепстральных коэффициентов и гамматон-частотных кепстральных коэффициентов в задаче детектирования голосовой активности. Для оценки значимости признаков использовался метод на основе решающих деревьев и сравнение значений площадей под ROC-кривыми для классификаторов, использующих метод опорных векторов, перцептрон и метод ближайших соседей. Оценена точность классификации как в случае использования всех анализируемых коэффициентов, так и для выделенных наиболее значимых признаков.

ные кепстральные коэффициенты (МЧКК, Mel-Frequency Cepstral Coefficients, MFCC) и гамматон-частотные кепстральные коэффициенты (ГЧКК, Gammatone-Frequency Cepstral Coefficients, GFCC). Важно отметить, что МЧКК и ГЧКК широко применяются не только в алгоритмах детектирования голосовой активности [3, 7-10], но и в алгоритмах идентификации диктора [11-12] и определения языка [13], а также при решении ряда других задач в области речевой тематики. Эффективность использования меры спектральной плоскостности в качестве одной из характеристик сигнала в задаче детектирования голосовой активности подтверждена рядом исследований [8, 9, 14].

Для классификации сигналов использовались классификаторы на основе метода опорных векторов (Support Vector Machine, SVM), перцептрона и метода k ближайших соседей (k -nearest neighbors) [9]. Указанные классификаторы обучаются с учителем, т.е. требуют размеченную обучающую выборку, вследствие чего возможно переобучение. Оно заключается в том, что полученные результаты будут хорошо работать в условиях, близких к условиям базы обучения, но плохо – с тестовой выборкой. Для предотвращения переобучения



используют перекрестную проверку (cross-validation). Для этого имеющаяся выборка данных разбивается на n частей. Затем на $n-1$ частях данных производится обучение моделей классификатора, а оставшаяся часть используется для тестирования. Данная процедура повторяется n раз, таким образом каждая из n частей данных будет использована для тестирования. Для приведенных результатов использовалось $n=4$.

Применяемые в работе классификаторы на основе перцептрона и метода опорных векторов с линейным ядром являются линейными, а их качество зависит от некоторых свойств данных. В частности, признаки должны быть нормализованы, то есть иметь одинаковый масштаб. Если это не так, и масштаб одного признака превосходит масштаб других, то качество классификации может резко упасть. Один из способов нормализации заключается в стандартизации признаков. По результатам проведенных исследований, представленных в [9], принято решение не использовать нормализацию признаков, т.к. это не дало ожидаемого повышения точности классификации в рассматриваемой задаче, а в некоторых случаях уменьшило это значение.

Для оценки важности признаков использовался метод на основе решающих деревьев, который относится к классу логических методов. Их основная идея состоит в объединении простых решающих правил. Одна из особенностей решающих деревьев заключается в том, что они позволяют определить важность всех используемых признаков. Важность признака можно оценить на основе того, как сильно улучшился критерий Джини благодаря использованию этого признака в вершине дерева [15].

При построении дерева необходимо задать критерий важности $Q(X, j, s)$, на основе которого осуществляется разбиение выборки на каждом шаге. Пусть R_m – множество объектов обучающей выборки, попавших в вершину m . Обозначим через p_{mk} долю объектов класса $k (k \in \{1, \dots, K\})$, попавших в вершину m :

$$p_{mk} = \frac{1}{N_m} \sum_{x_i \in R_m} [y_i = k],$$

где $N_m = |R_m|$. Через k_m обозначим класс, чьих представителей оказалось больше всего среди объектов, попавших в вершину m : $k_m = \arg[\max_k(p_{mk})]$. Критерий важности при ветвлении вершины m определяется как:

$$Q_G(R_m, j, s) = F_G(R_m) - \frac{N_l}{N_m} F_G(R_l) - \frac{N_r}{N_m} F_G(R_r),$$

где l и r – индексы левой и правой дочерних вершин, а

$$F_G(R_m) = \sum_{k \neq k'} p_{mk} p_{mk'}$$

Важность признаков была оценена для МСП, 12-мерных векторов МЧКК и ГЧКК. Для этого в работе использовалась собственная русскоязычная речевая база продолжительностью 44 минуты с частотой дискретизации 8000 Гц и точностью представления отсчетов 16 бит. Суммарная длительность речи совпадает с суммарной длительностью пауз и равна 22 минутам. Перед проведением исследования тестовые сигналы зашумлялись аддитивным белым гауссовским шумом (АБГШ) с отношением сигнал/шум (ОСШ) от -15 дБ до 25 дБ и шагом 5 дБ. После этого полученный сигнал разбивался на окна длительностью 16 мс с перекрытием 4 мс, и для каждого окна рассчитывался вектор признаков. Таким образом получена матрица коэффициентов размерностью $M \times W$, где M – количество рассчитанных коэффициентов ($M = 25$), а W – количество окон ($W = 220000$).

В табл. 1 приведены признаки, имеющие наибольшее суммарное значение важности для всего исследуемого диапазона ОСШ.

В результате выделены 7 признаков: мера спектральной плоскостности, 1, 2, 5-й мел-частотные кепстральные коэффициенты и 3, 4, 5-й гамматон-частотные кепстральные коэффициенты, как имеющие наибольшее значение важности по критерию Джини.

Установлено, что значения важности признаков при ОСШ от -15 дБ до -5 дБ отличаются незначительно, а при ОСШ от 0 дБ и выше важность данных коэффициентов значительно возрастает по сравнению с другими. Так, суммарное значение важности для данных 7 признаков более 0,55 и достигает значения 0,84 при ОСШ 20 дБ. Сумма значений важности для всех признаков при определенном ОСШ равна 1.

Детектирование голосовой активности

По принятой классификации детектирование голосовой активности является задачей бинарной классификации. Для оценки ее качества в случае использования указанных ранее коэффициентов построены ROC-кривые и рассчитаны площади под ними (AUC, area under ROC curve). Расчет AUC схематично изображен на рис. 1, где ИПЗ – истинно-положительное значение (true positive rate), ЛПЗ – ложно-положительное значение (false positive rate).

Таблица 1. Зависимость значения важности признака от ОСШ

Коэффициент	ОСШ, дБ								
	-15	-10	-5	0	5	10	15	20	25
МСП	0,04	0,04	0,03	0,04	0,09	0,12	0,15	0,05	0,14
МЧКК-1	0,05	0,05	0,08	0,07	0,05	0,21	0,25	0,25	0,33
МЧКК-2	0,04	0,05	0,04	0,10	0,05	0,08	0,02	0,19	0,01
МЧКК-5	0,04	0,05	0,05	0,03	0,05	0,05	0,14	0,01	0,12
ГЧКК-3	0,04	0,04	0,05	0,06	0,09	0,03	0,04	0,12	0,06
ГЧКК-4	0,07	0,11	0,11	0,15	0,15	0,11	0,10	0,08	0,01
ГЧКК-5	0,07	0,07	0,11	0,13	0,14	0,14	0,09	0,14	0,11

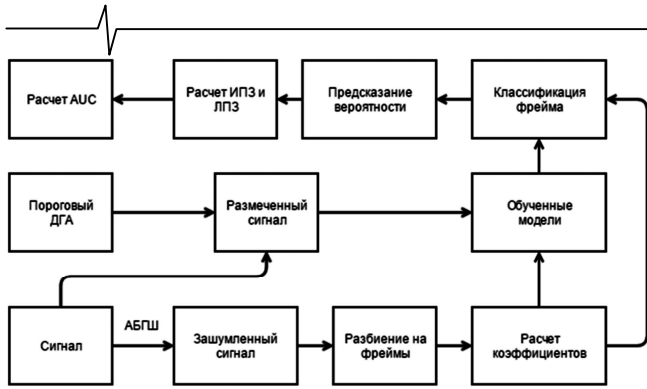


Рис. 1. Блок-схема расчета площади под ROC-кривой (AUC)

В качестве порогового ДГА использовался алгоритм, описанный в работе [2] и принимающий решение на основе сравнения энергии фрейма с заданным пороговым значением. Данный алгоритм выбран для разметки тестовых сигналов, так как он прост в реализации, не требует обучения и имеет высокую точность выделения голосовой активности в незашумленных речевых сигналах.

В табл. 2 приведены усредненные значения AUC для каждого классификатора при использовании конкретного коэффициента.

Выделенные коэффициенты определены как самые значимые признаки на предыдущем этапе с помощью

решающих деревьев. Полученные значения площади под ROC-кривой также подтвердили эффективность использования данных признаков в задаче детектирования голосовой активности для всех исследуемых в работе классификаторов.

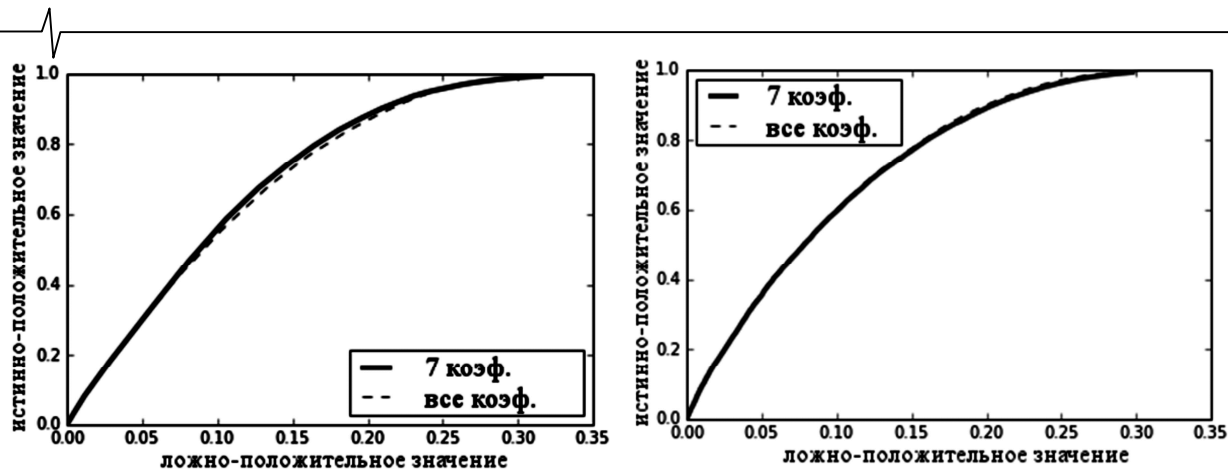
Благодаря выделению наиболее значимых признаков можно добиться не только уменьшения размерности обрабатываемых матриц коэффициентов, но и возможного при этом увеличения точности классификации ДГА.

На рис. 2 показаны ROC-кривые для ОСШ=10 дБ.

В табл. 3 отражена точность работы классификаторов как с полным набором акустических признаков (МСП, МЧКК и ГЧКК) и их комбинациями (МСП и МЧКК, МСП и ГЧКК), так и с выделенными на предыдущем этапе 7 коэффициентами. Видно, что при использовании различных классификаторов эффективность использования анализируемых акустических признаков различна. Так, при классификации с помощью перцептрона, площадь под ROC-кривой наибольшая в случае использования МСП и ГЧКК для всего диапазона ОСШ. Для классификатора на основе метода ближайших соседей точность выше при использовании выделенных на предыдущем этапе 7 коэффициентов. Для метода опорных векторов наибольшие значения AUC достигаются как

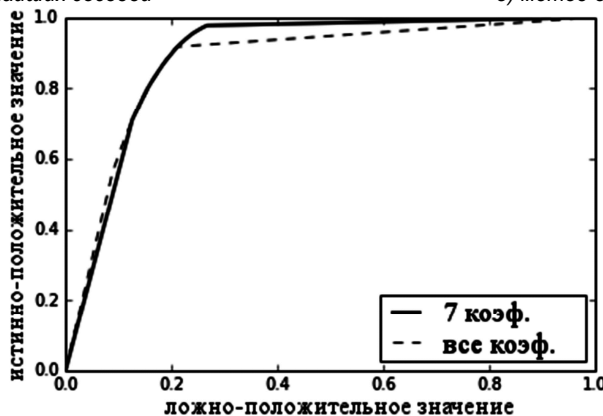
Таблица 2. Среднее значение площади под ROC-кривой (AUC)

Коэффициент	Классификатор		
	Метод опорных векторов	Перцептрон	Метод ближайших соседей
МСП	0,752	0,777	0,800
МЧКК-1	0,786	0,779	0,832
МЧКК-2	0,734	0,763	0,808
МЧКК-3	0,495	0,495	0,691
МЧКК-4	0,628	0,642	0,703
МЧКК-5	0,737	0,659	0,785
МЧКК-6	0,715	0,624	0,777
МЧКК-7	0,645	0,586	0,722
МЧКК-8	0,521	0,513	0,643
МЧКК-9	0,473	0,533	0,613
МЧКК-10	0,490	0,510	0,612
МЧКК-11	0,504	0,506	0,603
МЧКК-12	0,497	0,506	0,597
ГЧКК-1	0,537	0,548	0,612
ГЧКК-2	0,608	0,639	0,683
ГЧКК-3	0,734	0,751	0,776
ГЧКК-4	0,812	0,812	0,830
ГЧКК-5	0,797	0,797	0,817
ГЧКК-6	0,721	0,721	0,751
ГЧКК-7	0,669	0,669	0,707
ГЧКК-8	0,643	0,646	0,688
ГЧКК-9	0,628	0,648	0,689
ГЧКК-10	0,605	0,622	0,669
ГЧКК-11	0,604	0,614	0,663
ГЧКК-12	0,579	0,582	0,637



а) метод ближайших соседей

б) метод опорных векторов



в) перцептрон

Рис. 2. ROC-кривые

Таблица 3. Зависимость AUC от ОСШ при использовании различных наборов признаков сигнала

Классификатор	Коэффициенты	ОСШ, дБ								
		-15	-10	-5	0	5	10	15	20	25
Метод опорных векторов	Все коэффициенты	0,62	0,68	0,75	0,81	0,86	0,91	0,95	0,99	1,00
	7 коэффициентов	0,62	0,68	0,74	0,81	0,86	0,91	0,96	0,99	1,00
	МСП+МЧКК	0,57	0,68	0,75	0,81	0,86	0,91	0,95	0,98	1,00
	МСП+ГЧКК	0,62	0,70	0,68	0,73	0,80	0,86	0,91	0,96	0,99
Перцептрон	Все коэффициенты	0,61	0,66	0,75	0,79	0,84	0,88	0,90	0,78	0,98
	7 коэффициентов	0,51	0,62	0,76	0,81	0,86	0,89	0,85	0,95	0,98
	МСП+МЧКК	0,59	0,65	0,72	0,80	0,84	0,86	0,90	0,93	0,98
	МСП+ГЧКК	0,63	0,71	0,77	0,83	0,88	0,90	0,93	0,96	0,99
Метод ближайших соседей	Все коэффициенты	0,57	0,66	0,73	0,79	0,85	0,90	0,94	0,98	1,00
	7 коэффициентов	0,57	0,66	0,73	0,79	0,85	0,90	0,95	0,98	1,00
	МСП+МЧКК	0,57	0,66	0,73	0,79	0,85	0,90	0,94	0,98	0,99
	МСП+ГЧКК	0,56	0,65	0,72	0,78	0,82	0,86	0,91	0,95	0,98

при использовании всех 25 коэффициентов, так и при использовании выбранных 7 коэффициентов.

Если усреднить полученные значения AUC, можно выделить следующие зависимости: при использовании всех коэффициентов и метода опорных векторов достигается наибольшее значение $\overline{AUC} = 0,84$, то же значение достигается и при использовании выделенных 7 коэффициентов; в случае использования МСП, ГЧКК и перцептрона \overline{AUC} также равно 0,84; а при работе классификатора на основе метода ближайших соседей

наибольшее значение $\overline{AUC} = 0,83$ получено для 7 коэффициентов.

На рис. 3 отображена зависимость значения \overline{AUC} как для всех анализируемых ОСШ, так и для конкретных диапазонов.

Если сравнить классификаторы при использовании всех анализируемых коэффициентов и 7 коэффициентов, имеющих наибольшее значение важности, то можно отметить следующее: при ОСШ от -15 дБ до -5 дБ все классификаторы имеют большее значение AUC в слу-

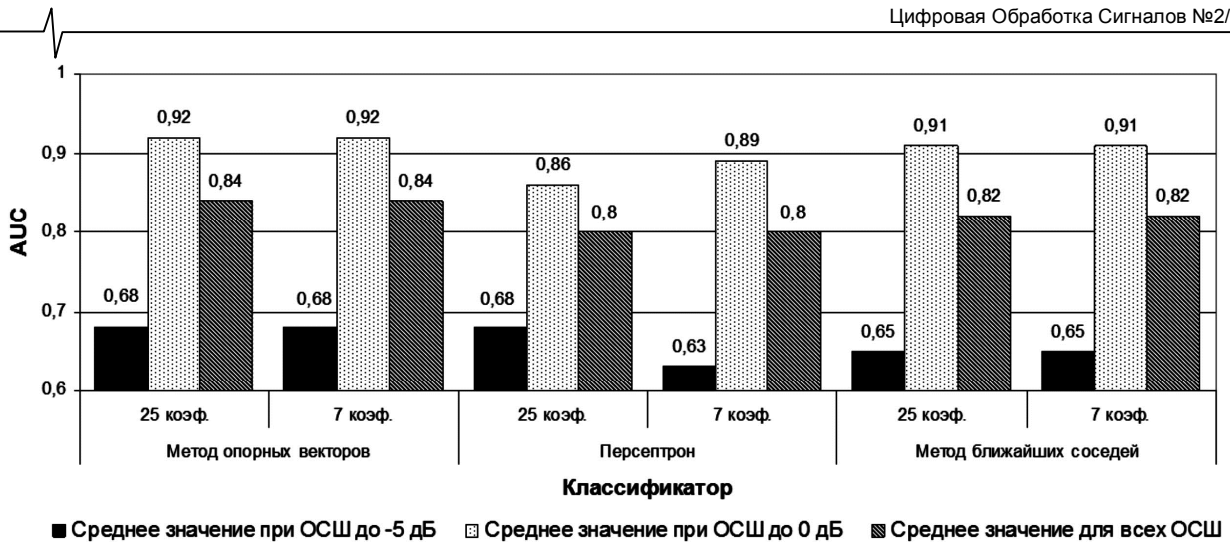


Рис. 3. Усредненные значения AUC для разных классификаторов и наборов используемых признаков

чае использования всех 25 коэффициентов (МСП, 12 МЧКК и 12 ГЧКК), а при ОСШ от 0 дБ и выше AUC в основном выше для выбранных 7 коэффициентов (МСП, 1, 2, 5-го МЧКК и 3, 4, 5-го ГЧКК). Подобная зависимость может быть объяснена полученными высокими значениями важности этих признаков при ОСШ более 0 дБ.

Заключение

Таким образом, использование метода расчета важности признаков на основе решающих деревьев эффективно в задаче определения наиболее важных акустических признаков для ДГА. Как наиболее значимые коэффициенты из числа анализируемых можно выделить: меру спектральной плоскостности, 1, 2, 5-й мел-частотные кепстральные коэффициенты и 3, 4, 5-й гамматон-частотные кепстральные коэффициенты. Полученные результаты исследований дают основания полагать, что использование данных коэффициентов при небольшом уровне зашумления увеличит скорость и точность работы алгоритмов детектирования голосовой активности.

Работа выполнена при поддержке гранта РФФИ № 16-37-00308.

Литература

1. Кравцов С.А., Тупицин Г.С., Топников А.И., Сагацян М.В., Приоров А.Л. Исследование работы детектора речевой активности в задаче идентификации диктора // Радиотехнические и телекоммуникационные системы. 2015. № 4 (20). С. 61–68.
2. Kinnunen T., Rajan P. A practical, self-adaptive voice activity detector for speaker verification with noisy telephone and microphone data // ICASSP. 2013. P. 7229–7233.
3. Tazi E.B., Benabbou A., Harti M. Voice activity detection for robust speaker identification system // Spec. Issue Int. J. Comput. Appl. Softw. Eng. Databases Expert Syst. – SEDEXS. 2012. № 9. P. 35–39.
4. Матвеев Ю. Технологии биометрической идентификации личности по голосу и другим модальностям // Вест-

ник МГТУ им. Н.Э. Баумана. Сер. «Приборостроение». 2012. № 3. С. 46-61.

5. Pasanen A. Voice activity detection in noise robust speech recognition // Theses, Tampere Univ. Technol. 2002.

6. Johnston J. Transform coding of audio signals using perceptual noise criteria // Sel. Areas Commun. IEEE. 1988. Vol. 6, № 2. P. 314-323.

7. Kinnunen T., Chernenko E. Voice activity detection using MFCC features and support vector machine // Int. Conf. Speech. 2007. Vol. 2. P. 555-561.

8. Кравцов С.А., Топников А.И., Приоров А.Л. Детектор речевой активности на основе голосующих моделей гауссовских смесей // Электромагнитные волны и электронные системы. 2015. Vol. 20, № 8. С. 29-34.

9. Кравцов С.А., Топников А.И. Анализ работы линейных классификаторов в задаче детектирования речевой активности // Цифровая обработка сигналов и ее применение. Докл. 18-й междунар. конф. 2016. Т. 1. С. 403-409.

10. Cooper D. Speech detection using gammatone features and one-class support vector machine // Dis. University of Central Florida Orlando. 2013.

11. Gudnason J., Brookes M. Voice source cepstrum coefficients for speaker identification // Acoust. Speech Signal Process. 2008. ICASSP 2008. IEEE. 2008. P. 4821–4824.

12. Tazi E., Benabbou A., Harti M. Efficient text independent speaker identification based on GFCC and CMN methods // Multimed. Comput. Syst. (ICMCS). 2012 Int. Conf. on. IEEE. 2012. P. 90–95.

13. Wong E., Sridharan S. Comparison of linear prediction cepstrum coefficients and mel-frequency cepstrum coefficients for language identification // Intell. Multimedia, Video Speech Process. 2001. P. 95-98.

14. Moattar M., Homayounpour M. A simple but efficient real-time voice activity detection algorithm // Signal Process. Conf. Eur. IEEE. 2009. P. 2549-2553.

15. Соколов Е. Семинары по решающим деревьям // machinelearning.ru.