Learning Sentence Reduction Rules for Brazilian Portuguese

Daniel Kawamoto and Thiago Alexandre Salgueiro Pardo

Núcleo Interinstitucional de Linguística Computacional (NILC)
Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo
Avenida Trabalhador São-Carlense, 400 - Centro
P.O.Box 668. 13560-970 - São Carlos, SP, Brazil

Abstract. We present in this paper a method for sentence reduction with summarization purposes. The task is modeled as a machine learning problem, relying on shallow and linguistic features, in order to automatically learn symbolic patterns/rules that produce good sentence reductions. We evaluate our results with Brazilian Portuguese texts and show that we achieve high accuracy and produce better results than the existing solution for this language.

1 Introduction

Text summarization is the task of producing a shorter version of a source text, its summary [12]. Summaries are useful in several circumstances, from decisions and actions that people make in their day by day lives (e.g., renting a movie, buying a book, or simply getting news updates) to machine applications (e.g., information retrieval and extraction).

Extractive summaries (or simply extracts) are traditionally defined as summaries composed of fragments from the source text, without rewriting [13]. Such approach is the most common one in text summarization research. In this line, two paradigms may be observed. The dominant one is the intersentential summarization, in which a summary is produced by selecting and juxtaposing important sentences from the original text. The other paradigm, referred by intrasentential summarization, or simply sentence reduction/compression, consists in shortening the sentences of a text by removing uninteresting and/or irrelevant parts of it. As an example of sentence reduction, we show below a source sentence (in Portuguese, the language for which we have worked) and its corresponding reduced version (by removal of words):

Source Sentence: *O assessor de Relações Institucionais da Transierra, Hugo Muñoz, afirmou nesta quinta-feira que o envio de gás natural da Bolívia para o Brasil está quase totalmente recuperado.*

Reduced Sentence: *O assessor afirmou que o envio de gás natural está quase recuperado.*

Such paradigms need not to compete. Ideally, they should work concurrently or interleaved, for instance, by reducing the sentences selected to compose a summary. This schema follows closely the human strategy for producing summaries [6], [8].

Kawamoto D. and Alexandre Salgueiro Pardo T.
Learning Sentence Reduction Rules for Brazilian Portuguese.
DOI: 10.5220/0003030300900099
In *Proceedings of the 7th International Workshop on Natural Language Processing and Cognitive Science (ICEIS 2010)*, page
ISBN: 978-989-8425-13-3

Sentence reduction is also useful for several other tasks, e.g., subtitle generation (where sentence size is limited by market standards and human ability to read what is being shown at some specific speed), mobile applications (where screen size is generally very small), and sentence simplification (for poor literacy readers and people with medical limitations – aphasia and dyslexia, for example).

Although sentence reduction with summarization purposes is a very interesting and important task, few works have dealt with it. Its challenge resides in the fact that such process must observe (i) the grammaticality of the resulting sentence, (ii) the text focus, guaranteeing that the important concepts of the text are not omitted, and (iii) the sentence context, in order to produce coherent and cohesive summaries. For Portuguese, the scenario is even worse, since there is only one known work in the area [19].

Sentence reduction for summarization is the focus of this paper. We model the task as a machine learning problem and apply it for Brazilian Portuguese texts. We rely on shallow (positional and distributional) and linguistic (morphosyntactic, syntactic and semantic) features to capture the relevance of the words in a sentence. Our purpose is to automatically learn symbolic patterns/rules that produce good sentence reductions. We evaluate our results and show that we achieve high accuracy and produce better results than the existing solution for Brazilian Portuguese.

The rest of the paper is organized as follows. Section 2 presents some relevant related work. The method we propose for sentence reduction and its development are presented and discussed in Section 3. We report our evaluation and the obtained results in Section 4. Some final remarks are shown in Section 5.

2 Related Work

Jing [7] presents one of the pioneering and most important works on sentence reduction. She uses several knowledge sources to determine whether a part of the sentence (words, phrases, and clauses) must be removed. The decision relies on syntactic and semantic clues (determining which syntactic components and verb arguments are grammatically obligatory), contextual information (how important each part of the text is in relation to the topic of the text, which is computed based on the relations that exist among each part and the rest of the text – using Wordnet and morphological relations), and corpus evidence (which gives probabilities of removing some part of the text given its surrounding context). Lin and McKeown [10] continue the above work by incorporating sentence combination functionality to their system. Given two reduced sentences, they implement combination operators that work over the syntactic trees of the sentences, producing a single tree that must be linguistically realized.

Knight and Marcu [9] use corpora composed of original texts and their abstracts to train a statistical model and a decision-based model to perform sentence reduction. The statistical model follows the recent machine translation trend, i.e., it represents the task of sentence reduction as a noisy-channel framework, which codifies how an original sentence may generate a reduced sentence. The decision-based model use shift-reduce parsing paradigm to perform the same task. Syntactical information guides both models. Daumé III and Marcu [4] extend the above work by enabling the models to process complete documents instead of only sentences. They incorporate in the model discourse relations from RST (Rhetorical Structure Theory) [14] over text segments to help determining which segments are more important and, therefore, which ones (already reduced or not) may be deleted from the text. More recently, Turner and Charniak [21] try to improve Knight and Marcu noisy-channel model by revisiting some model decisions. Nguyen et al. [16], on the other hand, use SVMs [23] to codify a set of features and the decision on which operations of the shift-reduce parsing to apply. Their feature set includes the syntactical information used by Knight and Marcu and semantic features: the named entities types of each word, whether each word is a head word or not, and whether each word has relationships with other words. Unno et al. [22] also extend Knight and Marcu work by considering maximum entropy models, using other features as depth of words in the syntactic tree and the own words.

In another line, Clarke and Lapata [2] model the problem of sentence reduction as an optimization problem. They encode decision variables and constraints in the model that try to guarantee the grammaticality of the reduced sentence and the removal of unimportant parts of it. The problem is solved by an integer programming approach. Still in a different line, Cordeiro et al. [3] model the problem through Inductive Logic Programming, where the alignment between paraphrases of the similar sentences is the basis for extracting relevant information and training the model.

Most of the above works are based on Jing and McKeown previous work [8] on corpus annotation and study of the phenomenon of summary decomposition, i.e., how the summary parts come from the corresponding source text. They identify several rewriting operations, including those that account for sentence reduction. Similar studies were conducted at [5] and [15]. All these works try to automate the task of aligning parts of the summary with the original parts in the corresponding text by using Hidden Markov Models and statistical alignment models.

To the best of our knowledge, GistSumm (GIST SUMMarizer) is the only approach for the task of sentence reduction for Brazilian Portuguese language [19]. In this system, sentence reduction is carried out by simply removing all the stopwords from the sentences, which consists in an overly simplistic solution.

Differently from previous works, in this paper we aim at learning symbolic word-based patterns/rules for sentence reduction. We present our approach in the next section.

3 Our Method

In the next subsection we report the annotation of a corpus of news texts in Brazilian Portuguese, which is the basis for this work. Then we present how we modeled the task of sentence reduction as a machine learning problem.

3.1 Corpus Annotation

Initially, in order to have available data for our proposal, we proceeded to a corpus annotation task.

We randomly selected 18 texts from TeMário corpus [18], which is a corpus consisting of news texts and their corresponding summaries in Brazilian Portuguese. The selection of this number of texts is due to the amount of necessary effort to annotate them. As it will be shown later, such amount showed to be enough for learning interesting sentence reduction rules.

Nine computational linguists were asked to read 2 texts each, to judge each sentence, and to annotate possible sentence parts that could be removed. Each text was annotated by only one judger. The instruction for the annotation was simply to annotate parts of the sentences (words, phrases, or entire clauses) that could be removed, without loosing grammaticality. It was not necessary to identify sentence parts to be removed in every sentence, since some sentences are very important and should be kept intact. We also did not establish any compression rate, i.e., how much sentences should be compressed.

Some examples of annotated sentences (in Portuguese) are shown below, where the brackets indicate the parts that could be removed in a sentence reduction process:

Clinton chegou [a Tóquio] na terça-feira.

[Para analistas políticos,] o acordo firmado [ontem] muda substancialmente a relação entre EUA e Japão.

A [maliciosa] canção deve grande parte [de seu impacto] ao produtor Clifton ["Specialist"] Dillon.

According to the instructions given to the annotators, grammaticality should be preserved. By asking the annotators to read the texts, we intended that their annotations kept the text focus and resulted in coherent and cohesive texts.

As a final step, we parsed the sentences with the PALAVRAS parser for Portuguese [1], which is told to be the best one for such language. The parser produces the syntactic tree and a shallow semantic analysis for each sentence. The shallow semantic analysis simply assigns semantic tags to some words in the sentence, e.g., human, local, animal, and organization tags. Notice that the semantic tags are not only named entity tags, since other words besides proper nouns may also be given tags.

3.2 Sentence Reduction as a Machine Learning Problem

We modeled the task of sentence reduction as a machine learning problem as follows. The information unit on which reduction decisions will be applied is the word. Each word of each sentence of a text must be judged in terms of a feature set that encode its relevance in the corresponding sentence and text. As a result, the word must be classified as “must be removed” or “must not be removed” from the sentence it belongs to.

It is important to say that we took into consideration all the tokens in the sentences, i.e., we also considered punctuation marks as words. This is important because, in some cases, punctuation marks are essential for sentence reduction, e.g., commas are one of the main hints for identifying relative clauses and appositions, whose words are good candidates for removal.

Our features codify shallow (positional and distributional) and linguistic (morphosyntactic, syntactic and semantic) aspects of each word. We use 15 features, which are listed in what follows:

- Part of speech tag of the word (e.g., noun, verb, adverb, etc.);
- Main extra morphosyntactic information available for the word: this indicate, for instance, whether a verb is the main verb, and whether an article is definite or not; if such information is not available, the value for the feature is set to "none";
- Other extra morphosyntactic information available for the word: other possibly available information, as the previous feature;
- Syntactic information of the word: this feature specifies to which syntactic component the word belongs to, e.g., subject, direct object, adjuncts, etc.;
- Semantic tag of the word (e.g., human, organization, etc.): if not available, the value for this feature is set to "none";
- Secondary semantic tag of the word: as previous feature (since some words may have more than one semantic tag);
- Tertiary semantic tag of the word: as previous feature[1];
- Part of speech of the preceding word (i.e., the word in the word-1 position): if there is not a preceding word (in the case the word is in the first position in the sentence), the value for this feature is set to "none";
- Part of speech of the other preceding word (i.e., the word in the word-2 position): same behavior of the previous feature;
- Part of speech of the following word (i.e., the word in the word+1 position): if there is not a following word (in the case the word is in the last position in the sentence), the value for this feature is set to "none";
- Part of speech of the other following word (i.e., the word in the word+2 position): same behavior of the previous feature;
- Frequency of the word: its value is the frequency of occurrence of the word in the text to which its sentence belongs;
- Presence in the most important sentence of the text: if the word occurs in the sentence of the text that is judged as the most important, the value of this feature is "gist_word", otherwise it is "not_gist_word"; all stopwords are defined as "not_gist_word", even if they occur in the most important sentence; the most important sentence is assumed to be the one that contains the most frequent words, and therefore, would convey the text main idea;
- Position of the word in the text: this feature stores where the word occurs in the text – in the beginning (if the word is one of the 20% first words), in the end (the 20% last words), or in the middle of the text (every other word);
- Position of the word in the sentence: this feature stores where the word occurs in the sentence it belongs to – in the beginning (if it is the first word of the sentence), in the end (if it is the last word – not counting punctuation in this case), or in the middle of the sentence (every other word).

Our features were built in a way to try to capture the important phenomena in sentence reduction process. Morphosyntactic (including part of speech) and syntactic features capture the functional importance of the words in the sentence and also the

[1] In fact, we verified that a word may have up to 3 semantic tags assigned to it by the PALAVRAS parser, and this is the reason for having 3 semantic features.

local context (given the features for the preceding and following words of each word). Semantic features capture relevant meaning aspects that may eventually interfere in the reduction process. Frequency and presence-in-the-most-important-sentence features try to encode the context and the contribution of the words to the text focus. For computing these two features, the whole text is lemmatized in order to avoid discrepancies. In particular, for computing the feature presence-in-the-most-important-sentence, we used an intersentential summarization system [17]. The position of the words in the sentence and in the text also try to encode word importance, as it is known that some text parts usually contain more important information (for instance, it is widely known that news texts usually present their main information in the first sentences).

We extracted features for each word of the sentences of the 18 texts that were manually annotated. The class assigned to each word and its feature values was the one indicated by the human annotation: if the word was marked as possible to remove, the class was "must be removed"; otherwise it received the class "must not be removed".

Having the problem modeled in this way, we expect to be able to learn symbolic patterns/rules for deciding which words of a sentence to remove. Our experiments are described in the next section.

4 Experiments

We used the tool WEKA [24] for running our experiments. We selected decision trees (the J48 algorithm) for our experiments. As it is known, it is a symbolic representation, which may also be directly mapped into rules, as we wish in this work.

We randomly selected one text for test from the 18 available texts, while the remaining 17 texts were left for training. In general, the 17 texts produced 17.102 learning instances (remember that each word – including punctuation – produces an instance) after balancing the data by duplicating the "must be removed" instances, since there were much more "must not be removed" instances. Data duplication for balancing classes is a usual practice in machine learning, as it is discussed by Prati and Monard [20]. The text for testing produced 462 instances (unbalanced). We tried to use as many texts as possible for training (leaving only 1 text for testing) in order to try to learn interesting rules.

We obtained an error rate of 18.4% in the test set, i.e., a general precision of 81.6%. It is shown below some examples of rules extracted from the decision tree that we obtained:

IF the word is in the beginning of the sentence AND the part of speech tag is pronoun AND there is no extra morphosyntactic information available AND there is no semantic information available THEN the word must be removed.

IF the word is in the middle of the sentence AND it is part of a syntactical component of the type adverbial adjunct THEN the word must be removed.

These rules look intuitive and encode our general knowledge that pronouns in the beginning of sentences are important elements and that adverbs are not. Other rules are not so obvious. See, for instance, the following 2 rules:

IF the word is in the middle of the sentence AND it is part of the syntactic component of the type subject AND its semantic tag is administration THEN the word must not be removed.

IF the word is in the middle of the sentence AND it is part of the syntactic component of the type subject AND its semantic tag is organization THEN the word must be removed.

The semantic tags are the only difference between them, but one may see that the tags are of related nature (in fact, such tags are usually confused in semantic annotation tasks). Other rules are incredible simple and have high coverage (i.e., they correctly account for several instances). See, for instance, the rule below:

IF the word is in the end of a sentence THEN the word must not be removed.

In Table 1 we show the confusion matrix for the decision tree we learned. One can see that the results are quite good. Although the test set has relatively few "must be removed" instances, the decision tree correctly classified most of them, misclassifying only 3 instances (7.8% of them). For the "must not be removed" instances, 82 instances were misclassified (19.3% of them).

Table 1. Confusion matrix.

Predicted class / Actual class	Must not be removed	Must be removed
Must not be removed	342	82
Must be removed	3	35

The next step we performed was the informativeness evaluation of the sentences that our method produced. Informativeness is one of the most important evaluation criteria in summarization [12]. For performing the evaluation, we initially built the reduced sentences of the test set by removing the words indicated by the decision tree. We then collected the reduced sentences built by the human annotators for the same test set, the reduced sentences automatically produced by GistSumm, and the reduced sentences produced by randomly removing words (this process was automatically done). GistSumm and the random method were considered baseline methods in our evaluation, i.e., methods that we must outperform in order to show that our approach is worth of being pursued. The human sentences are our reference sentences, i.e., the ones that we aim at reproducing.

Having the reduced sentences given by the 4 methods above, we used ROUGE (Recall-Oriented Understudy for Gisting Evaluation) [11] for comparing the informativeness of the sentences. ROUGE is an automatic metric that is able to rank summaries (automatically produced or not) by their quality. This metric basically computes the number of n-grams that a summary under evaluation and at least one reference (human) summary have in common. The more n-grams in common they have, the best the summary under evaluation is. ROUGE authors have shown that such measure is as good as humans in ranking summaries, even if we consider only unigrams. For this reason, ROUGE has become a mandatory measure in any summarization evaluation and in international summarization contests (like TAC – Text Analysis Conference[2], which is the most important conference on summarization), frequently replacing human judgment, which is expensive and

---

[2] www.nist.gov/tac

subjective and, therefore, subject to errors and inconsistencies. ROUGE scores fall within the 0 (the worst) to 1 (the best) range.

We used the human summaries as reference summaries. ROUGE average results (for all the sentences in the test set) considering only unigrams are showed in Table 2. We show precision, coverage, and f-measure figures computed by ROUGE, which are measures traditionally used in the area. F-measure is a measure that combines precision and coverage, being a unique indicator of the quality of a system.

Table 2. Informativeness results.

	Precision	*Coverage*	*F-measure*
Random method	0.82782	0.60067	0.69388
GistSumm	0.77676	0.67836	0.72026
Our method	0.97418	0.69074	0.80392

One may see that our method outperforms both the random method and GistSumm in all the measures. Surprisingly, we not only achieved very high ROUGE scores (our precision is close to 1), but also performed much better than GistSumm, which we imagined would be a hard baseline to beat, since it only removes stopwords from the sentences and, therefore, does not run the risk of removing important content words.

As illustration, we show in Figure 1 examples of reduced sentences (in Portuguese) produced by the 3 automatic methods and by the human annotators, with good and bad reductions. It is important to notice that, each time the random method runs, it will produce different results. We believe that the last sentence produced by our method (item l) is not so good because the original sentence is very big and our model might still not be robust enough to deal with more complex structures like this, and/or we might not have enough training data on this particular structure type so that our method could not learn appropriate rules.

Sentences produced by humans

(a) *Aviões da Otan bombardearam ontem posições sérvias ao norte de Sarajevo.*
(b) *Um helicóptero da ONU tentou perseguir os sérvios, que responderam a tiros.*
(c) *O texto da declaração evita qualquer sugestão de que os Estados Unidos estejam pedindo ao Japão para revisar sua Constituição de 1947.*

Sentences produced by the random method

(d) *Aviões da bombardearam sérvias norte de Sarajevo.*
(e) *Um da tentou perseguir os sérvios, tiros.*
(f) *O texto da declaração qualquer sugestão de Estados estejam pedindo ao Japão para Constituição 1947, compreensivelmente pacifista militarista país, exacerbado Segunda.*

Sentences produced by GistSumm

(g) *Aviões Otan bombardearam ontem posições sérvias norte Sarajevo.*
(h) *Helicóptero ONU tentou perseguir sérvios, responderam tiros.*
(i) *Texto declaração evita cuidadosamente sugestão Estados Unidos estejam pedindo Japão revisar Constituição 1947, compreensivelmente pacifista depois passado militarista país, exacerbado durante Segunda Guerra.*

Sentences produced by our method

(j) *Aviões da Otan bombardearam ontem posições sérvias de Sarajevo.*
(k) *Um helicóptero da ONU tentou perseguir os sérvios, tiros.*
(l) *texto evita qualquer sugestão que Estados Unidos estejam pedindo ao Japão para sua Constituição de.*

Fig. 1. Examples of reduced sentences.

We present some final remarks in the next section.

5 Final Remarks

The method we presented in this paper combine features of diverse nature in a symbolic machine learning technique in order to learn word-based patterns/rules for performing sentence reduction. We evaluated our method for Brazilian Portuguese language and showed that we achieved high accuracy and could outperform a random baseline and a system that we believed would be a hard baseline.

An important drawback of our approach is that it is not possible to specify a compression rate for the reduction process in the way we performed it here, since we modeled the task without considering such parameter. If the observation of certain compression rate is necessary, one might iteratively apply the learned reduction rules until the point that the compression rate is achieved.

One source of error that we could detect in sentence reduction comes from processing big sentences. More training data and/or more informative features may solve this (for instance, features encoding information about sentence chunks instead of words only). Another source of error is presumably the parser we use, but we did not undergo a strict verification of this point.

As future work, more robust evaluation should be carried out, including more texts in the training and test data sets, as well as comparing our approach to other methods for English, for instance. Although we tested our method for Portuguese, it looks general enough to be also applied to other languages. We also believe that one interesting research line in this area that is worth of following is to consider user queries (as it happens in query-focused summarization) as an extra parameter to guide sentence reduction.

Acknowledgements

The authors are grateful to FAPESP and Santander for supporting this work.

References

1. Bick, E. (2000). *The Parsing System PALAVRAS: Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. Aarhus University Press.
2. Clark, J. and Lapata, M. (2006). Constraint-based Sentence Compression: An Integer Programming Approach. In the *Proceedings of the COLING/ACL*, pp. 144-151.
3. Cordeiro, J.P.; Dias, G.; Brazdil, P. (2009). Unsupervised Induction of Sentence Compression Rules. In the *Proceedings of the Workshop on Language Generation and Summarisation*, pp. 15-22. Singapore.
4. Daumé III, H. and Marcu, D. (2002). A noisy-channel model for document compression. In the *Proceedings of the Conference of the Association for Computational Linguistics*, pp. 449-456.

5. Daumé III, H. and Marcu, D. (2005). Induction of Word and Phrase Alignments for Automatic Document Summarization. *Computational Linguistics*, V. 31, N. 4, pp. 505-530.
6. Endres-Niggemeyer, B. and Neugebauer, E. (1995). Professional summarising: No cognitive simulation without observation. In the *Proceedings of the International Conference in Cognitive Science*.
7. Jing, H. (2000). Sentence Reduction for Automatic Text Summarization (2000). In the *Proceedings of the 6th Applied Natural Language Processing Conference*, pp. 310-315.
8. Jing, H. and McKeown, K.R. (1999). The Decomposition of Human-Written Summary Sentence. *Research and Development in Information Retrieval*, pp.129-136.
9. Knight, K. and Marcu, D. (2002). Summarization beyond sentence extraction: A probabilistic approach to sentence compression. *Artificial Intelligence*, V. 139, N.1, pp. 91-107.
10. Lin C.-Y. and McKeown K.R. (2000). Cut and Paste Based Text Summarization. In the *Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics*, pp. 178-185.
11. Lin, C.Y. and Hovy, E. (2003). Automatic Evaluation of Summaries Using N-gram Co-occurrence Statistics. In the *Proceedings of 2003 Language Technology Conference* (HLT-NAACL 2003), Edmonton, Canada.
12. Mani, I. (2001). *Automatic Summarization*. John Benjamins Publishing Co. Amsterdam.
13. Mani, I. and Maybury, M.T. (1999). *Advances in Automatic Text Summarization*. MIT Press.
14. Mann, W.C. and Thompson, S.A. (1987). *Rhetorical Structure Theory: A Theory of Text Organization*. Technical Report ISI/RS-87-190.
15. Marcu, D. (1999). The automatic construction of large-scale corpora for summarization research. In the *Proceedings of the 22nd Conference on Research and Development in Information Retrieval*, pp 137–144.
16. Nguyen, M.L.; Shimazu, A.; Horiguchi, S.; Ho, B.T.; Fukushi, M. (2004). Probabilistic Sentence Reduction Using Support Vector Machines. In the *Proceedings of the 20th international conference on Computational Linguistics*.
17. Pardo, T.A.S.; Rino, L.H.M.; Nunes, M.G.V. (2003). GistSumm: *A Summarization Tool Based on a New Extractive Method*. In Lecture Notes in Artificial Intelligence 2721, pp. 210-218. Faro, Portugal. June 26-27.
18. Pardo, T.A.S. e Rino, L.H.M. (2003). *TeMário: Um Corpus para Sumarização Automática de Textos*. Série de Relatórios do NILC. NILC-TR-03-09. São Carlos-SP, Outubro, 13p.
19. Pardo, T.A.S. (2005). *GistSumm - GIST SUMMarizer: Extensões e Novas Funcionalidades*. Série de Relatórios do NILC. NILC-TR-05-05. São Carlos-SP, Fevereiro, 8p.
20. Prati, R.C. e Monard, M.C. (2008). Novas abordagens em aprendizado de máquina para a geração de regras, classes desbalanceadas e ordenação de casos. In the *Proceedings of the VI Best MSc Dissertation/PhD Thesis Contest*.
21. Turner, J. and Charniak, E. (2005). Supervised and Unsupervised Learning for Sentence Compression. In the *Proceedings of the 43rd Annual Meeting of the ACL*, pp. 290-297.
22. Unno, Y.; Ninomiya, T.; Miyao, Y.; Tsujii, J. (2006). Trimming CFG Parse Trees for Sentence Compression Using Machine Learning Approaches. In the *Proceedings of the COLING/ACL*, pp. 850-857.
23. Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. New York: Springer-Verlag.
24. Witten, I.H. and Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.