The effect of word frequency and position-in-utterance in Mandarin speech errors: A connectionist model of speech production

I-Ping Wan[1], and Marc Allassonnière-Tang[2]

[1] Graduate Institute of Linguistics, Research Center for Mind, Brain and Learning, National Chengchi University, Taipei, Republic of China
[2] Lab Dynamics of Language UMR 5596, CNRS/University of Lyon, France
ipwan@g.nccu.edu.tw, marc.tang@univ-lyon2.fr

Abstract. The connectionist model of speech processing infers that word frequency and position-in-utterance play a major role in the occurrence of speech errors. First, words that are not frequently used are more likely to result in speech errors since they generally receive less activation than frequently occurring words and require more activation to be chosen. Second, speech errors are more likely to occur near the end of utterances since, according to the given-before-new-principle, utterance-final words convey new information that has not yet been activated in the preceding context. The information of word frequency and position-in-utterance is extracted automatically from 382 utterances of a Mandarin speech error corpus and fed to generalized linear mixed models and a decision-tree based classifier. The results show that word frequency and position-in-utterance can predict of the occurrence of speech errors with a performance over (but close to) the majority baseline. Therefore, additional information is required to improve the accuracy of the predictions.

Keywords: Speech errors, Mandarin, Frequency, Position-in-utterance

1 Introduction

This paper focuses on lexical errors, i.e., erroneous selections of lexical items that involve a meaningful morpheme or word. They typically occur when the 'lemmas' of semantically or phonologically appropriate candidates for lexical items are activated. For instance, when a speaker says *glass* instead of *cup*, or *book* instead of *cook*. The other types of errors that result in meaningless strings of phoneme (e.g., when a speaker says *perple* instead of *person* or *people*) and errors that originate from the surrounding context (e.g., when a speaker says *the glass is in the glass* instead of *the glass is in the fridge*) are excluded from the current study for theoretical and practical reasons. First, this study investigates the frequency of the target (i.e., the intended word) and the error word in utterances. Thus, meaningless words must be excluded since they cannot be assigned a frequency in corpora. Second, context-induced errors and purely phonological errors are less relevant to the cognitive representation of

language, since the cause of context-induced errors is due to interference of the surface structure of language rather than its deep inner processing.

The main contributions of this paper are as follows. First, most of the literature focused on speech errors resulting in meaningless strings of phonemes due to their higher occurring frequency and relevance to phonology [1–4]. By focusing on lexical errors, this study provides another type of data that can verify the predictions made by language processing models at the semantic level. Second, previous studies found several tendencies predicted by language-processing models within corpora of lexical errors [5–10]. However, these results were mostly obtained from Germanic and Romance languages and may be subject to Galton's problem [11], i.e., the tendencies observed in speech errors may be language-family-specific rather than universal. This paper thus provides data on speech errors in Mandarin, which enhances linguistic diversity in the results. Last but not least, previous studies investigating speech processing models with speech errors seldom provide quantitative analysis on the predictive power of word frequency and position-in-utterance with regard to speech errors. A few studies known to the authors investigates quantitatively the negative correlation between word frequency and the probability of occurrence of speech errors [12]. Nevertheless, the effect of position-in-utterance is not considered simultaneously with word frequency in these studies. This paper also aims at filling this gap.

2 Hypotheses and research questions

In parallel models of speech production, multiple levels of processing take place simultaneously [13–15]. This approach assumes that speech processing is not a serial motion but a simultaneous activation of distinct units of speech (e.g., phonemes, morphemes) represented as nodes that interact with each other in parallel across the semantic, word, and sound levels. An example is demonstrated in Fig. 1.

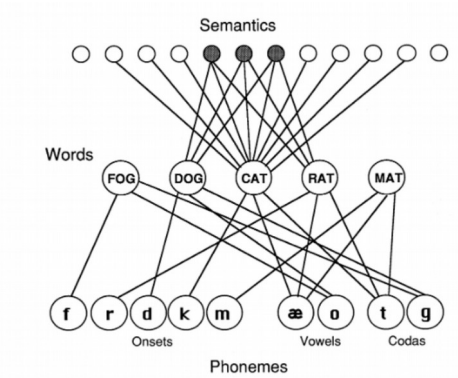

Fig. 1. A simplified overview of the connectionist approach [16]

Assuming that a speaker intends to retrieve and convey the concept of *cat*, the semantic features related to the target word *cat* are activated and spread to neighbor

nodes. For instance, the semantic feature *four-legs* is activated and the activation spreads to *cat* at the word level. However, the semantic feature *four-legs* also activates the words that share this feature, e.g., *dog* and *rat*. The activation of the word *cat* spreads at the same time to the phoneme level, e.g., to the phoneme [t]. The same process applies for other nodes that received activation, e.g., *rat* and *dog*. The activation of nodes is simultaneous and multidirectional. By way of illustration, the activation of *rat* by the feature *four-legs* can also spread back to the semantic level and activate other semantic features that were not activated by *cat,* e.g., *is grey*. The same logic applies between the word level and the sound level. Finally, the most highly activated word of the appropriate grammatical category is chosen.

Several predictions can be made only under models of the connectionist approach. First, word frequency is expected to play a significant role in predicting the occurrence of speech errors. Words that are frequently used are activated more frequently and are thus more likely to reach the required level of activation before potential erroneous candidates. Few previous studies investigated the effect of word frequency on speech errors in Mandarin [12]. However, the results were not analyzed in connection with the predictions of speech processing models. Second, the position-in-utterance of the target word is expected to play a significant role in predicting the occurrence of speech errors. This prediction relates to the given-before-new-principle in discourse analysis [17–22], which states that old information comes first in utterances, while new information comes later. Under such premise, speech errors are less expected at the beginning of an utterance since old information would already have received activation from the preceding context and would thus be less likely to result in erroneous activation. The opposite statement would be assumed for utterance-final words conveying new information that has not yet been activated. No previous studies known to the authors have investigated the predictive effect of position-in-utterance on speech errors.

3 Data

The 382 speech errors investigated in this study are from the same source as [12]. The speech errors are retrieved from a conversational corpus produced in a naturalistic setting by approximately 100 native speakers of Taiwan Mandarin between 1995 and 2009 [3, 23, 24]. Each audio file of conversational speech is automatically transcribed based on a Speech-to-Text software modeled by Taiwan AI Lab with an average accuracy rate of 70%. The entire transcript is then automatically segmented by the Academia Sinica word segmenter [25, 26]. Further editing and correction is made by two research assistants. Speech errors are identified based on repair (self-correction) initiated by the speakers. As an example in (1a), the target *dong4ci2* ‘verb’ is erroneously replaced by *ming2ci2* ‘noun’. However, the speaker immediately initiates repair with the intended target. In (1b), the target *na4* ‘that’ is erroneously replaced by *zhe4* ‘this’, but the speaker also initiates repair directly after the speech error. Separate analyses are made by research assistants working on the corpus and inconsistencies are resolved by analysis of the context and the phonetic realization of the words [7]. This method restricts the sample size of speech errors.

However, it is considered appropriate for theoretical reasons. Errors that occur without repair from the speakers are hard to identify and verify. By way of illustration, if a speaker says *more* instead of *less* in *I have more time for myself* but does not initiate repair, the speech error cannot be identified since the utterance is perfectly grammatical. An analysis of the context could potentially help to identify errors of this type, but without confirmation from the speaker for each individual error (which is practically impossible and still not extremely reliable since it is involves an off-line judgment from the speakers), it would be mere guesses rather than confirmed errors.

(1) Examples of lexical speech errors in the corpus

a.	zhe4xie1 ming2ci2 …	tong1tong1 dong4ci2	yao4	jia1	dan1shu4	
	these noun	all verb	need	plus	singular	
	‘these must be used with singular nouns … singular verbs’					
b.	zhe4 shi2hou4	…	na4 […]	tian1	zai4	bian4lun4	de0
	this time	…	that	day	at	debate	DE
	‘this … that day during the debate […]’						

Each speech error is annotated with the following information: target, error, preceding context, and following context. The corpus has 382 speech errors and includes 3022 words when considering the preceding and following words of each speech error. Due to the small size of the entire corpus (382 sentences in total), information about word frequency is added based on word frequency from the Academia Sinica Corpus [25], which has 11,245,330 words and is the first fully POS-tagged balanced Chinese Corpus [27]. Then, to facilitate comparison between high and low frequencies of words, the logarithm of the raw frequency is used. Information about position-in-utterance is also added by counting how many words separate the speech error from the beginning of the utterance in which it occurred. Each error is assigned a number based on how many words separate it from the beginning of the utterance. By way of illustration, in (1a), the error *ming2ci2* ‘noun’ occurred at the sixth word of the utterance, its position-in-utterance is thus annotated as 5. This value is then normalized by dividing the position-in-utterance by the total number of words in the utterance (not including the error nor the repair of the target), i.e., 5/5 = 1. A value of 1 indicates that the error is found at the end of the utterance. A value of 0 indicates that the error is at the beginning of the utterance.

4 Analysis

In terms of word frequency, previous studies [12] already pointed out that (i) the targets and the errors have higher frequency than most of the other words in the lexicon (ii) the word frequency of the targets and the errors is commonly lower than

the frequency of words in the surrounding context. We thus only provide an overview of the interaction between word frequency and position-in-utterance. In Fig. 2, a PCA analysis of the two variables is shown. Since we only have two variables, the two components refer to word frequency (PC1 on the x axis) and position in utterance (PC2 on the y axis, coded as dist_to_start). First, we can visualize the effect of word frequency attested in previous studies. The errors are more likely to occur on the left side of the plot, in the opposite direction of the loading word_freq, which indicates that errors are are found more frequently with words of small frequency. Second, a similar effect seems to be found with position-in-utterance, as more errors are located at the bottom of the plot. However, the effect is less obvious than with frequency. As a reminder, this is only a visualization of the data. The following paragraphs provide the quantitative analysis for the interaction between word frequency and position-in-utterance.

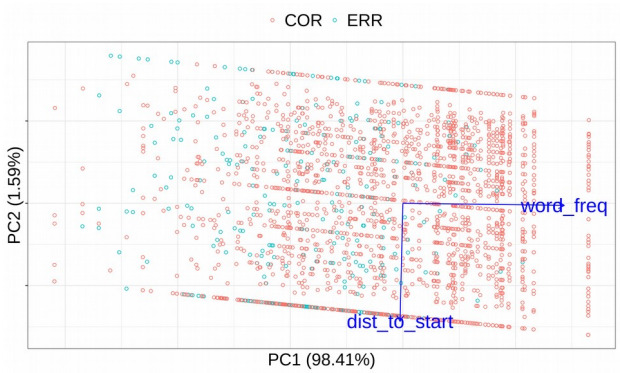

Fig. **2.** PCA visualization of the interaction between word frequency and position-in-utterance with regard to speech errors.

We then feed the information of word frequency and position-in-utterance to two different machine learning methods. First, we use generalized linear mixed models (GLMMs) [28, 29] to interpret the effect of the two variables on the occurrence speech errors. Second, we try to reproduce the results from previous studies [12] by using a decision tree based classifier to predict the occurrence of speech errors based on the two variables.

With regard to the GLMMs, the parameters were set as follows: 4 chains with 500 of iterations each, including 200 iterations as warm-up. The results reported are from the model that does not consider random effects or interaction for the two variables. This choice is made on the theoretical premise that we want to directly assess the effects of the two variables on speech errors. We did test random effects and interaction in other models and compared their fit with leave-one-out cross validation. The model considering word frequency as a random effect gives the best fit, but the divergence across the models is not big. Thus, we consider that the results reported in the current paper are sufficient for identifying the effect of word frequency and position-in-utterance on speech errors. The credible intervals of the output are shown in Figure 3. First of all, neither of the two ‘humps’ cross 0 (Rhat = 1, ESS = 763, 690), which indicates that the two variables have a clear positive/negative effect. In

our case, word frequency has a negative effect (all the values are negative) while position-in-utterance (represented by dist_to_start, i.e., distance to the beginning of a sentence), has a positive effect as all the values are positive. In other words, the model indicates that i) the higher the frequency of a word, the less likely it is to be a speech error (est = -0.32) ii) the bigger distance to the beginning of sentence, i.e., the closer a word is to the end of a sentence, the more likely it is to be a speech error (est = 0.59). Therefore, the output of the model supports the two predictions based on the connectionist model.

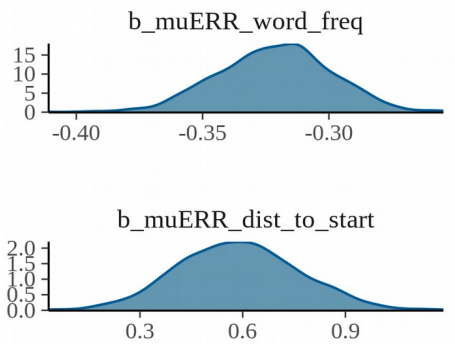

Fig. **3.** Credible intervals for word frequency and position-in-utterance

Then, the computational classifier based on decision trees is used to predict the presence of speech error for each word of each utterance in the speech error corpus. These decision trees are based on binary recursive partitioning [30, 31]. The main functioning of the classifier is explained as follows. Binary splits recursively divide the data into homogeneous or near-homogeneous buckets. The split is considered ideal if the homogeneity of the buckets is improved after the split. To assure a low variation in the output, the model does not use all the variables and observations. Instead, the model uses a bootstrap sample of the original data and also selects a random subset of variables for each split. That is to say, first, the algorithm scans through the variables and selects the strongest association with the response. Then, the data set is divided into two subsets based on the chosen variable. These two steps are repeated for every subset until no variables may split the data with statistical significance. The main advantage of decision trees is their visualization of the interaction between the variables. That is to say, a decision tree is generated and can be read to make a prediction on a specific data point.

To train the classifiers, the data is split into two sets, one for training, the other for testing. The training set contains 70% of the data while the test set has 30%. To avoid biases from a specific combination of tokens between the training and test sets, the same process was conducted ten times with different training and test sets. Since the results did not vary across the sets, we report the output of the tenth set in the current paper. Moreover, to replicate previous studies [12], we also assess the performance of the classifier in different five different window sizes. That is to say, we first ask the classifier to identify the speech error from the target and its preceding and following word. Then, we expand the window to the two preceding and following words, and so

on, until the maximum size of five preceding and following words, which capture the full length of each sentence in the corpus.

The performance of the decision tree is evaluated based on its accuracy, precision, and recall. The accuracy provides an overview of the performance by dividing the correctly predicted tokens with the total of the tokens. The precision evaluates how many tokens are correct among all the output of the classifier and recall quantifies how many tokens are correctly retrieved among all the expected correct output [32]. Finally, since the quantity of correct words and speech errors is unbalanced within the data set, we use the majority rule as a benchmark of accuracy. Taking window size five as an example, i.e., when we consider the entire corpus: The corpus of 3022 words only contains 382 speech errors, the computational classifier may reach an accuracy of 87.4% simply by guessing that all the words do not undergo speech errors. In such case, the computational classifier should have an accuracy higher than 81.8% to be considered as having a good performance. The same logic is applied for the other window size. The performance on the tenth test set is reported for each window size in Table 1.

Table 1. The performance of the decision-tree based classifier

Window	1		2		3		4		5	
Type	cor	err	cor	err	cor	err	cor	err	cor	err
Precision	0.80	0.62	0.94	0.40	0.95	0.38	1	0	1	0
Recall	0.77	0.66	0.82	0.64	0.87	0.61	0.84	0	0.88	0
Accuracy	0.73		0.80		0.84		0.84		0.87	
Baseline	0.62		0.76		0.81		0.84		0.87	

As observed in previous studies [12], the classifier does not perform well with large window sizes. That is to say, speech errors are hard to detect in large corpora due to their scarcity. On the other hand, for window sizes smaller than four, the performance of the classifier increases and exceeds the majority baselines. The performance on individual categories is also analyzed. First, the model performs well at detecting the absence of speech errors (cor). This is not surprising since the majority of the data points are without speech errors. The model therefore has less difficulty to identify this category. Second, starting from window size three, the recall on detecting errors does not increase by much, however, its precision increases. An interesting fact is the increase of accuracy in comparison with the model used in [12], which only considered word frequency. In the current model, the accuracy of window size 1-3 is 3-4% higher (similar results are found when our classifier is fed with information on word frequency and position-in-utterance individually), which suggests that adding the information of position-in-utterance does result in better predictions.

The regularities found by the classifier in window size one are visualized in Fig. 4. The upper nodes refer to the decision path and the buckets at the bottom indicate the ratio of speech errors (ERR) and correctly uttered words (COR). As an example, if the logarithm of word frequency is smaller than 9.2 (node 1 → node 3) and the distance

to the beginning of the sentence is larger or equal to 9.6, i.e., if the word is either the final word or the penultimate word of a sentence (node 3 → node 7), the model predicts that a speech error occurs. Within the entire data set, 77 tokens are categorized under such pathway and 91% (70/77) of them are classified correctly.

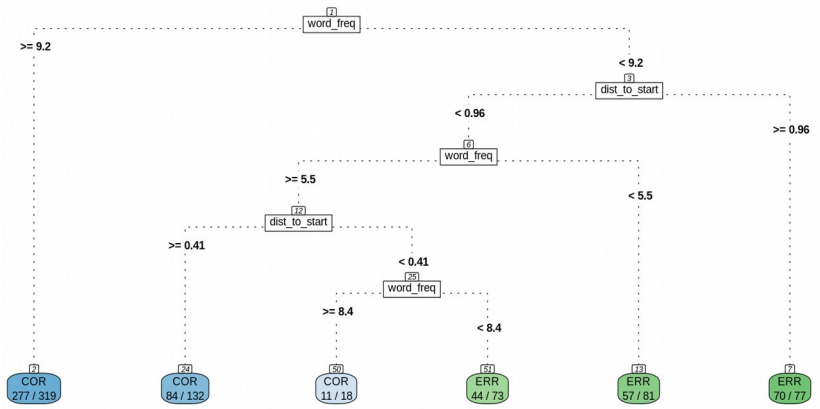

Fig. 4. Decision tree for the occurrence of speech errors in the corpus of 1888 words

The tree shows that position-in-utterance and word frequency both plays a role in predicting speech errors. On the one hand, words with high frequency are less likely to result in speech errors (e.g., node 1 → node 2). On the other hand, words located at the end of an utterance are also extremely likely to result in speech error (node 1 → 3 → 7). Similar tendencies are found further away from these two extremes. As an example, words closer to the beginning of a sentence are more likely to result in speech errors if they have a low frequency (e.g., node 1 → 3 → 6 → 13).

5 Conclusion

The connectionist model predicts that i) Frequently used words are less likely to result in speech errors since they are easily activated ii) Utterance-initial words are less likely to result in speech errors since they convey old information that already has been activated in the preceding context. The results of the current analysis support these two hypotheses, but also indicate that other variables should be added in the model to result in an accurate prediction of speech errors. Potential candidates for additional variables are part-of-speech tags and semantic/phonological distance with the preceding and following contexts.

References

1. Alderete, J., Davies, M.: Investigating Perceptual Biases, Data Reliability, and Data Discovery in a Methodology for Collecting Speech Errors From Audio

Recordings. Lang Speech. 62, 281–317 (2019). https://doi.org/10.1177/0023830918765012.
2. Alderete, J., Tupper, P.: Connectionist approaches to generative phonology*. In: Hannahs, S.J. and Bosch, A.R.K. (eds.) The Routledge Handbook of Phonological Theory. pp. 360–390. Routledge (2017). https://doi.org/10.4324/9781315675428-13.
3. Wan, I.-P.: Mandarin speech errors into phonological patterns. Journal of Chinese Linguistics. 35, 185–224 (2007).
4. Wan I.-P.: Consonant Features in Mandarin Speech Errors. Concentric : Studies in Linguistics. 42, 1–39 (2016). https://doi.org/10.6241/concentric.ling.42.2.01.
5. Fay, D., Cutler, A.: Malapropisms and the structure of the mental lexicon. Linguistic Inquiry. 8, 505–520 (1977).
6. Fromkin, V.: Speech errors as linguistic evidence. Mouton de Gruyter, The Hague (1973).
7. Harley, T.A., MacAndrew, S.B.G.: Constraints upon word substitution speech errors. Journal of Psycholinguistic Research. 30, 395–418 (2001). https://doi.org/10.1023/A:1010421724343.
8. Jaeger, J.J.: Kids' slips: What young children's slips of the tongue reveal about language development. Lawrence Erlbaum Associates, Mahwah (2004).
9. Jaeger, J.J., Wilkins, D.: Semantic relationships in lexical errors. In: Jaeger, J.J. (ed.) Kids' slips: What young children's slips of the tongue reveal about language development. pp. 311–384. Lawrence Erlbaum Associates, Mahwah (1993).
10. Nooteboom, S.G.: The tongue slips into patterns. In: Fromkin, V. (ed.) Speech errors as linguistic evidence. pp. 87–95. Mouton de Gruyter, The Hague (1973).
11. Naroll, R.: Galton's problem. In: Naroll, R. and Cohen, R. (eds.) A handbook of method in cultural anthropology. pp. 973–989. Columbia University Press, New York (1973).
12. Tang, M., Wan, I.-P.: Predicting Speech Errors in Mandarin Based on Word Frequency. In: Su, Q. and Zhan, W. (eds.) From Minimal Contrast to Meaning Construct. pp. 289–303. Springer Singapore, Singapore (2019). https://doi.org/10.1007/978-981-32-9240-6_20.
13. Dell, G.S.: Representation of serial order in speech: Evidence from the repeated phoneme effect in speech errors. Journal of Experimental Psychology: Learning, Memory, and Cognition. 10, 222–233 (1984). https://doi.org/10.1037/0278-7393.10.2.222.
14. Dell, G.S.: A spreading-activation theory of retrieval in sentence production. Psychological Review. 93, 283–321 (1986). https://doi.org/10.1037/0033-295X.93.3.283.
15. Dell, G.S.: The retrieval of phonological forms in production: Tests of predictions from a connectionist model. Journal of Memory and Language. 27, 124–142 (1988). https://doi.org/10.1016/0749-596X(88)90070-8.
16. Dell, G.S., Chang, F., Griffin, Z.M.: Connectionist Models of Language Production: Lexical Access and Grammatical Encoding. Cognitive Science. 23, 517–542 (1999). https://doi.org/10.1207/s15516709cog2304_6.

17. Clifton, C., Frazier, L.: Should given information come before new? Yes and no. Memory & Cognition. 32, 886–895 (2004). https://doi.org/10.3758/BF03196867.
18. Gundel, J.K.: 'Shared knowledge' and topicality. Journal of Pragmatics. 9, 83–107 (1985). https://doi.org/10.1016/0378-2166(85)90049-9.
19. Gundel, J.K.: Universals of topic-comment structure. In: Hammond, M., Moravcsik, E.A., and Wirth, J. (eds.) Typological Studies in Language. p. 209. John Benjamins Publishing Company, Amsterdam (1988). https://doi.org/10.1075/tsl.17.16gun.
20. Hu, C.: Information Structure in English, Mandarin Chinese and Taiwanese Southern Min: Argument realization of ditransitive objects, (2015).
21. Ibrahim, S.: A corpus-based investigation of the given before new principle in Tanzanian English. Papers from the 9th Lancaster University Postgraduate Conference in Linguistics & Language Teaching 2014. 70–70 (2014).
22. Junge, B., Theakston, A.L., Lieven, E.V.M.: Given–new/new–given? Children's sensitivity to the ordering of information in complex sentences. Applied Psycholinguistics. 36, 589–612 (2015). https://doi.org/10.1017/S0142716413000350.
23. Wan, I.-P.: Mandarin phonology: Evidence from speech errors, (1999). Ph.D. dissertation. State University of New York at Buffalo. N.Y. U.S.A.
24. Wan, I.-P.: On the phonological organization of Mandarin tones. Lingua. 117, 1715–1738 (2007). https://doi.org/10.1016/j.lingua.2006.10.002.
25. CKIP (Chinese Knowledge and Information Processing): Part-of-speech analysis of Academia Sinica Balanced Corpus of Modern Chinese, Technical Report (no.93-05), version 3. Academia Sinica (2004).
26. Ma, W.-Y., Chen, K.-J.: Introduction to CKIP Chinese Word Segmentation System for the First International Chinese Word Segmentation Bakeoff. Proceedings of ACL, Second SIGHAN Workshop on Chinese Language Processing. 168–171 (2003).
27. Huang, C.-R., Lee, L.-H., Hong, J.-F., Yu, S.: Quality assurance of automatic annotation of very large corpora: A study based on heterogeneous tagging systems. LREC 2008. 2725–2729 (2008).
28. Bürkner, P.-C . brms: An R package for Bayesian multilevel models using Stan. Journal of Statistical Software, 80(1) (2017). https://doi.org/10.18637/jss.v080.i01
29. Bürkner, P.-C. Advanced Bayesian multilevel modeling with the R package brms. The R Journal, 10(1), 395 (2018). https://doi.org/10.32614/RJ-2018-017
30. Breiman, L., Friedman, J., Stone, C.J., Olshen, R.: Classification and regression trees. Taylor & Francis, New York (1984).
31. Breiman, L.: Random forests. Machine Learning. 45, 5–32 (2001).
32. Ting, K.M.: Precision and Recall. In: Sammut, C. and Webb, G.I. (eds.) Encyclopedia of Machine Learning. pp. 781–781. Springer US, Boston, MA (2010). https://doi.org/10.1007/978-0-387-30164-8_652.