



Text Classification of Conversational Implicatures Based on Lexical Features

Xianbo Li

To cite this article: Xianbo Li (2022) Text Classification of Conversational Implicatures Based on Lexical Features, Applied Artificial Intelligence, 36:1, 2127598, DOI: 10.1080/08839514.2022.2127598

To link to this article: <https://doi.org/10.1080/08839514.2022.2127598>



© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 26 Sep 2022.



Submit your article to this journal [↗](#)



Article views: 753




View related articles [↗](#)



View Crossmark data [↗](#)

Text Classification of Conversational Implicatures Based on Lexical Features

Xianbo Li 

Si-Mian Institute for Advanced Studies in Humanities, East China Normal University, Shanghai, China

ABSTRACT

Following the guiding hypothesis in NLP, similar word frequency vectors may have similar implicatures, but some scholars are more inclined conversational implicatures cannot be obtained only through lexical features. To judge which view is more reasonable and explore the reasons for the divergence between them, whether conversational implicatures can be obtained only through lexical features is verified empirically. Main work of this paper includes: First, based on 600 corpora in the annotated dataset, the values of 20 lexical features of each corpus are obtained by automatic calculation. Second, meta-transformer of logistic regression for selecting features is adopted for feature selection and ranking. Third, after determining the features, the text is classified by the binomial logistic regression with the type of implicatures as labels. Fourth, results are tested for significance to identify relationships between variables. Experiments show that there is a statistical dependence between lexical features and conversational implicatures, and the text classification of implicatures can be performed only based on lexical features. In addition, the results of text classification will not be different due to the difference in context utterance or the type of implicature, and the text classification of implicatures only based on “response utterance” is more efficient.

ARTICLE HISTORY

Received 13 July 2022

Accepted 19 September 2022

Introduction

Conversational implicature is the core topic of pragmatics, and it is also one of the difficult problems to be overcome in natural language processing. However, the research paths of conversational implicatures appear to be contradictory in pragmatics and computer science. Turney and Pantel (2010) have put forward a hypothesis as the cornerstone of natural language processing: “If units of text have similar vectors in a text frequency matrix, then they tend to have similar meanings.” According to this hypothesis, since conversational implicatures are a subset of meaning, similar implicatures should also be reflected by similar lexical frequency metrics; that is, implicatures can be distinguished only by lexical features. However, since conversational implicature is expressed implicitly which is meant without being part of

CONTACT Xianbo Li  Albert_Dillon@sina.com  500 dongchuan road, Minhang, Shanghai, China

© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

what is said (Huang 2017, 156), some linguists are more inclined to believe that a specific context is a necessary condition for understanding conversational implicatures; that is, the conversational implicatures cannot be obtained only according to the features of language forms. For example, Huang (2014, 32) claimed that “a conversational implicature is not part of what a sentence means.” From that viewpoint, it is impossible to deduce conversational implicatures only from literal information of text. Therefore, there is a superficial disagreement between computer science and pragmatics about whether conversational implicatures can be calculated only by vectors generated from lexical features. To judge which view is more reasonable and explore the reasons for the divergence between them carefully, it is necessary to verify whether conversational implicatures can be obtained only through lexical features empirically.

Since computer scholars believe that the similarity of meaning can be judged only by the similarity of frequencies, it seems that some clues can be obtained without contextual information when judging the conversational implicatures. However, a wide range of philosophers including relevance theorists, some neo-Wittgensteinians, and some Sellarsians claimed that every single expression is context sensitive, or “if the only context sensitivity you take into account is that due to the expressions in the basic set, you won’t get a proposition or anything truth evaluable” (Cappelen and Lepore 2005, 7–8). They considered all meaning is determined by context. Similarly, Clark (2013, 15) also believed that the inference of the hearer to the implicature will not use “linguistic meanings of the words,” which is inconsistent with the view of computer scholars. Therefore, even we prove that the conversational implicature is computable, it is also necessary to study whether the computer needs to rely on the context to infer the implicature. If the computer obtains some clues about the implicature without the context information, it can at least show that “all meaning is determined by context” is not completely perfect, which can assist to verify the views “against radical contextualism” held by scholars such as Lassiter (2021).

At present, the research on conversational implicature itself mainly focuses on presupposition, pragmatic function, etc. The research approaches mainly include theoretical analysis and statistical and corpus methods. Sbisà (2021) examined the explicitation practice of implicit meaning through theoretical analysis, and gave a method to distinguish implicature and presupposition. Based on corpus coding, Garassino, Brocca, and Masia (2022) analyzed some implicit strategies of political communication in a corpus of British and Italian tweets by calculating Kappa and AC1 value (Hoek and Scholman 2017). More scholars are willing to pay attention to the cognitive process of implicature. Li et al. (2018) proposed a Bayesian belief network model of indirect speech act theory based on idealized cognitive model and probabilistic pragmatics (Frank and Goodman 2012; Goodman and Frank

2016), and the cognitive process of particularized conversational implicature is explained from a computational perspective. Kecskes (2021) explained “simplicature” from the perspective of social cognition and proposed a model to explain the relationship and interplay between factors that affect implicature processing. Feng, Yu, and Zhou (2021) adopted functional MRI (fMRI) and transcranial direct current stimulation (tDCS) discovering particularized implicature and generalized implicature comprehension shared the multivariate fMRI patterns of language processing, of which particularized implicature could elicit theory-of-mind-related pattern. Wylie et al. (2022) provided the conditions for children to produce implicature and depicted its underlying social-cognitive mechanisms. Based on relevant literature, the current research on conversational implicature is almost all along the path of pragmatics, that is, understanding conversational implicature in a specific context, but ignoring the information about whether conversational implicature can be obtained only through lexical features. If lexical features can indeed provide clues for conversational implicatures, it proves the feasibility of pragmatic computing to a certain extent, which provides a new way for natural language processing.

In order to investigate whether the similarity of lexical features can represent the similarity of conversational implicatures to a certain extent and determine the degree of context dependence when classifying implicatures, three main research questions are proposed here: (1) Can the text classification of conversational implicatures be performed based only on lexical features? How is its accuracy? (2) What features are needed to classify conversational implicatures based on lexical factors? (3) Will the results of this text classification be different because of the context utterance or the type of implicature? To solve the above problems, we first use meta-transformer of logistic regression for selecting features based on importance weights to select the Optimal Feature Set for text classification, and then use the binomial logistic regression model to classify the conversational implicatures to obtain the classification results. Finally, the classification results are statistically tested to draw conclusions.

Materials and Methods

Here, we first introduce the used corpus sources and related concepts, and then the logistic regression and its classifier will be analyzed briefly.

Corpus Source

The corpus required for the experiments is selected from the annotated dataset of conversational implicatures in English dialogue constructed by George and Mamidi (2020), from which 600 items with the particularized implicature as

yes ($N = 300$) and no ($N = 300$) are selected, including context utterance, response utterance, and implicature. Here are two examples:

Example 1:

Context utterance: Are you going for the party?

Response utterance: *Is the pope Catholic?*

The listener asks the speaker if he or she is going to a party, and the speaker asks back if the Pope is Catholic. According to common sense and encyclopedic knowledge, the fact that the Pope is Catholic is beyond doubt, and the answer is “yes.” Meanwhile, the words of interlocutors should be related according to Cooperative Principles. Therefore, the hearer deduces that the speaker means that he or she is going to a party, so here the implicature can be labeled as “yes.”

Example 2:

Context utterance: Did you go to the movies last night?

Response utterance: *I had to study last night.*

The listener asked the speaker if he or she went to the movies last night, and the speaker replied that he or she had to study last night. According to common sense, if he or she was studying last night, he or she would have to take up the time that he or she should go to the movies, so the speaker’s implicature is that he or she didn’t go to the movies last night, and the implicature can be labeled as “no.”

In the experiments, the particularized implicature is regarded as the label of classification, and each context utterance and response utterance are merged and extracted to form a document (the total number of documents is 600), which is used for classification. In addition, information containing only response utterance is extracted as text to judge the relationship between the features of language forms and particularized implicature under non-contextual conditions, which is used to discover the strength of context utterance’s role in text classification.

Binomial Logistic Regression Model

The classifier used here for the text classification of conversational implicatures is a binomial logistic regression model, which is represented by a conditional probability distribution $P(Y|X)$ in the form of a parameterized logistic distribution (Li 2022, 78–79).

Suppose the weight vector is $w = (w^{(1)}, w^{(2)}, \dots, w^{(n)}, b)^T$, and the input vector is $x = (x^{(1)}, x^{(2)}, \dots, x^{(n)}, 1)^T$, then the binomial logistic regression model is as follows:

$$P(Y = yes|x) = \frac{\exp(\hat{w} \cdot x)}{1 + \exp(\hat{w} \cdot x)} \quad (1)$$

$$P(Y = no|x) = \frac{1}{1 + \exp(\hat{w} \cdot x)} \quad (2)$$

The parameters here are estimated by maximum likelihood estimation, let $P(Y = yes|x) = \pi(x)$, $P(Y = no|x) = 1 - \pi(x)$, then the log-likelihood function is

$$\begin{aligned} L(w) &= \sum_{i=1}^N [y_i \log \pi(x_i) + (1 - y_i) \log(1 - \pi(x_i))] \\ &= \sum_{i=1}^N \left[y_i \log \frac{\pi(x_i)}{1 - \pi(x_i)} + \log(1 - \pi(x_i)) \right] \\ &= \sum_{i=1}^N [y_i(w \cdot x_i) - \log(1 - \exp(w \cdot x_i))] \end{aligned}$$

Maximize $L(w)$ to get an estimate of w . Substituting the obtained estimates \hat{w} into formulas (1) and (2), the logistic regression model for the classification of implicatures is obtained.

Experiments and Results

Here the experimental processes are introduced in general, and then the details and results of each experiment are described one by one.

Experiment Procedure

The experiment is mainly divided into four steps: data calculation, feature sorting and selection, automatic text classification, and significance test of classification results.

First, data calculation. Calculate the corresponding data of lexical features according to the indicators. The main lexical indicators include: (1) Descriptive statistics: such as the mean and standard deviation of number of syllables, and the mean and standard deviation of number of letters. (2) Lexical diversity: such as type-token ratio of content word lemmas and all words, MTLTD and VOCD (McCarthy and Jarvis 2010). (3) Other vocabulary information: such as average word frequency for content words and all words; average minimum word frequency in sentences; familiarity, concreteness, imageability, and polysemy for content words; hypernymy for nouns and

verbs, etc. The entire calculation process is automatically completed using Coh-Metrix Web Tool (Graesser et al. 2004, 2014; Graesser, McNamara, and Kulikowich 2011).

Second, feature sorting and selection. The meta-transformer of logistic regression for selecting features based on importance weights is used to calculate the coefficients, and then the coefficients are sorted from large to small in absolute value to determine the importance of features. After sorting, forward selection is used to continuously add new features from the empty set, and then select the feature combination with the highest accuracy as the feature set for text classification.

Third, automatic text classification. After the features are determined, the logistic regression algorithm is used for text classification (Zhou 2021, 62–65). Four-fold cross-validation is used for classification. Each time, 75% of the corpus ($N_{\text{yes_train}} = 225$, $N_{\text{no_train}} = 225$) is selected as the training set, and the remaining 25% of the corpus ($N_{\text{test}} = N_{\text{yes_test}} + N_{\text{no_test}} = 150$) is used as the test set. Because the order of the corpus is shuffled using a random number generator to ensure that different types of corpus are not adjacent to each other, the test sets in each fold cross-validation are directly selected in order, that is, corresponding to “1-75,” “76-150,” “151-225,” and “226-300” respectively. The training set is the data other than the test set.

Fourth, the significance test of the classification results. After obtaining the results of text classification, it is necessary to judge the statistical dependence between lexical features and conversational implicatures by performing goodness-of-fit test with equal expected frequency, and then determine whether the text classification of conversational implicatures can be performed only by lexical features. In addition, to determine whether the inclusion of “context utterance” and different types of conversational implicatures will affect the text classification results, it is necessary to use Independent-Samples *T*-Test, Paired-Samples *T*-Test and Contingency Analysis to determine the difference relationship between variables.

The Text Classification of Conversational Implicatures with Context Utterance

The first experiment conducts text classification based on both context utterance and response utterance. Its main goal is to find a feature set composed of lexical features that is suitable for text classification with implicatures as a label, and then perform text classification on this basis to obtain classification results.

After feature sorting and selection, the Optimal Feature Set F_I for text classification based on both context utterance and response utterance is found. This feature set contains 17 features in total. According to the interpretation in Coh-Metrix and the coefficients obtained in feature selection, the Optimal Feature Set F_I is summarized as shown in Table 1.

Table 1. Features and their coefficients in the Optimal Feature Set F_1 .

Feature	The Explanation of Feature	Coefficient
WRDFRQmc	CELEX Log minimum frequency for content words, mean	0.23078
WRDHYPv	Hypernymy for verbs, mean	0.158253
WRDFRQc	CELEX word frequency for content words, mean	0.1365
DESWLltd	Word length, number of letters, standard deviation	-0.11829
DESWLlt	Word length, number of letters, mean	0.097787
WRDHYPnv	Hypernymy for nouns and verbs, mean	-0.06214
DESWLsyd	Word length, number of syllables, standard deviation	-0.04945
WRDPOLc	Polysemy for content words, mean	0.047379
WRDHYPn	Hypernymy for nouns, mean	-0.03848
DESWLsy	Word length, number of syllables, mean	-0.03355
WRDFRQa	CELEX Log frequency for all words, mean	0.025816
LDTTRc	Lexical diversity, type-token ratio, content word lemmas	0.009295
LDMTLD	Lexical diversity, MTLTD, all words	-0.0049
LDTTRa	Lexical diversity, type-token ratio, all words	-0.00348
WRDFAMc	Familiarity for content words, mean	-0.0023
WRDMEAc	Meaningfulness, Colorado norms, content words, mean	0.002255
WRDIMGc	Imagability for content words, mean	-0.00211

It can be found from Table 1 that when the text used for classification includes both context utterance and response utterance, the required features are not only large in number (17), but also rich in variety, including descriptive statistics, lexical diversity, and those related values used to represent word information. The features haven't included in F_1 include the mean of concreteness for content words and the mean of age of acquisition for content words. Although these two features have a large variance, their contribution to the discriminant category is not enough. The variance of the value of lexical diversity of all words from VOCD is 0, which cannot be used as a decision feature for text classification obviously.

Next, data analysis is performed on the results of text classification. The experiment records the classification results of each cross-validation. First, the basic classifier evaluation measures are shown here, including the number of true positive (TP) and true negative (TN). Based on this basis, the true positive rate (sensitivity), true negative rate (specificity), and recognition rate (accuracy) are reported. Next, the Independent-Samples T -Test is also performed on the texts labeled "yes" and "no" to examine whether the accuracy rates of different categories are significantly different to judge whether the reliability of the classification is related to the type of implicature. Finally, a goodness-of-fit test with equal expected frequency is performed on the accuracy, and then it is determined whether the logistic regression based on the Optimal Feature Set F_1 is an effective method for the classification of implicatures to verify whether automatic classification of conversational implicatures can be performed based on lexical features. The experimental results are shown in Table 2.

According to Table 2, the accuracy of each fold cross-validation is concentrated around 60%, and the overall accuracy is 59%. This accuracy cannot fully explain that the feature set F_1 can be used as the basis for the text classification of conversational implicatures, because random factors may also cause the

Table 2. Results of the performance measures, *T*-test and Goodness-of-fit test for the text classification of conversational implicatures with context utterance.

Folds	Performance measures					T-test Sig.	Goodness of fit test	
	TP	sensitivity	TN	specificity	accuracy		χ^2	Sig.
fold 1	50	0.667	46	0.613	0.640	0.500	11.760	0.001
fold 2	45	0.600	42	0.560	0.580	0.622	3.840	0.050
fold 3	42	0.560	48	0.640	0.600	0.321	6.000	0.014
fold 4	41	0.547	40	0.533	0.540	0.871	0.960	0.327
total	178	0.593	176	0.587	0.590	0.868	19.440	0.000

accuracy to be higher than the general value. According to the p -value in the goodness-of-fit test, two of the four tests are significant ($p_1 = 0.001$, $p_3 = 0.014$), one is marginally significant ($p_2 = 0.050$), and one is not significant ($p_4 = 0.327$). This shows that under the current sample size corpus, it can basically be determined that the feature set F_1 is the influencing factor of implicature. From the overall result of goodness-of-fit test, the significance is obvious ($\chi^2 = 19.440$, $p = 0.000$). The results of this test are not surprising, because as the sample size increases, when the accuracy is fixed at 59%, the interference of random factors will become smaller and smaller. Therefore, the Optimal Feature Set F_1 can be used as the basis for the text classification of conversational implicatures.

Next, consider whether the above classification is balanced in the positive and negative categories. According to the sensitivity and specificity in each fold of cross-validation, there is no obvious rule for the accuracy of positive and negative classes. Overall, the value of sensitivity (59.3%) and specificity (58.7%) are not much different, and after Independent-Samples *T*-Test, whether it is cross-validation for each fold or the whole, the significance of all tests is greater than 0.05. That is, when text classification is performed based on context utterance and response utterance at the same time, the classification accuracy has nothing to do with the type of implicature. This shows that with context utterance, the text classification of conversational implicatures is balanced in positive and negative categories.

The Text Classification of Conversational Implicatures Without Context Utterance

The previous experiments show that lexical features can significantly contribute to the text classification of conversational implicatures when context utterance is included. However, since the text used in that experiment includes both context utterance and response utterance, it is difficult to determine that the contributor to the correct classification is the response utterance spoken by the speaker; that is, the true source of the successful classification is uncertain. To determine the contribution of response utterance to text classification, all context utterances in the corpus can be eliminated, and the text classification

can be performed again. Then, by observing whether there is any change in the experimental results, it can be judged to what extent the influence on the text classification of conversational implicatures comes from “response utterance.” Like the previous experiment, the main goal of this experiment is to find a feature set composed of lexical features that is suitable for text classification with implicatures as labels. Based on this, the text classification is carried out, and the classification results are obtained in the similar way.

After feature sorting and selection, this paper finds the Optimal Feature Set F_2 for text classification only based on response utterance, which contains five features in total. According to the interpretation in Coh-Metrix and the coefficients obtained in feature selection, the Optimal Feature Set F_2 is summarized as shown in Table 3.

From Table 3, when the text used for classification only contains response utterance, the required number of features is only 5, and the types are more single, including only three descriptive statistics and two related values used to represent word information. It can be found that when the text classification of conversational implicatures is performed only according to the response utterance, the statistics of the number of letters and syllables and the frequency for content words may be able to be used as effective features for classification. However, the specific effectiveness needs to be judged according to the experimental results.

Next, data analysis is performed on the results of text classification. Like the first experiment, the basic classifier evaluation measures are shown here, and then Independent-Samples T -Test is performed on the text labeled “yes” and “no” to check whether the accuracy of different categories is significantly different to judge whether the reliability of the classification is related to the type of implicature. Finally, a goodness-of-fit test with equal expected frequency is performed on the accuracy, and then it is determined whether the logistic regression based on the Optimal Feature Set F_2 is an effective method for the classification of implicatures to verify whether automatic classification of conversational implicatures can be performed based on lexical features. The experimental results are shown in Table 4.

According to Table 4, the accuracy of each fold cross-validation is concentrated around 60%, and the overall accuracy is 60.7%. This accuracy cannot fully explain that the feature set F_2 can be used as the basis for the text classification of conversational implicatures, because random factors may

Table 3. Features and their coefficients in the Optimal Feature Set F_2 .

Feature	The Explanation of Feature	Coefficient
WRDFRQmc	CELEX Log minimum frequency for content words, mean	0.290257
DESWLltd	Word length, number of letters, standard deviation	-0.255635
DESWLlt	Word length, number of letters, mean	0.182417
WRDFRQc	CELEX word frequency for content words, mean	0.136391
DESWLsyd	Word length, number of syllables, standard deviation	-0.108100

Table 4. Results of the performance measures, *T*-test and Goodness-of-fit test for the text classification of conversational implicatures without context utterance.

Folds	Performance measures					T-test Sig.	Goodness of fit test	
	TP	sensitivity	TN	specificity	accuracy		χ^2	Sig.
fold 1	42	0.560	44	0.587	0.573	0.743	3.227	0.072
fold 2	46	0.613	53	0.707	0.660	0.230	15.360	0.000
fold 3	36	0.480	51	0.680	0.580	0.013	3.840	0.050
fold 4	49	0.653	43	0.573	0.613	0.318	7.707	0.006
total	173	0.577	191	0.637	0.607	0.133	27.307	0.000

also cause the accuracy to be higher than the general value. According to the *p*-value in the goodness-of-fit test, two of the four tests are very significant ($p_2 = 0.000$, $p_4 = 0.006$), and twice are marginally significant ($p_1 = 0.072$, $p_3 = 0.050$). This shows that under the current sample size corpus, it can basically be determined that the feature set F_2 is the influencing factor of implicature. From the overall results of goodness-of-fit test, the significance is still obvious ($\chi^2 = 27.307$, $p = 0.000$). According to the above results, it can be determined that only based on the response utterance, the Optimal Feature Set F_2 can be used as the basis for the text classification of conversational implicatures.

Likewise, consider whether the above classification is balanced in the positive and negative categories. According to the sensitivity and specificity in the cross-validation of each fold, the accuracy of the positive class is lower than that of the negative class in most cases, but fold 4 is an exception. Overall, the value of sensitivity (57.7%) is lower than that of specificity (63.7%). However, after Independent-Samples *T*-Test, the significance of most cross-validation and overall *T*-test is greater than 0.05. It shows that when text classification is performed only based on response utterance, the classification accuracy has nothing to do with the category of implicature; that is, in the absence of context utterance, the text classification of conversational implicatures is also balanced in positive and negative categories.

Comparison Between with and without Context Utterance

The influence of context utterance and response utterance on the classification results can be determined by comparing the first two experiments horizontally. In addition to determining the size relationship between the two by comparing the relevant indicators in the accuracy and goodness-of-fit test, and since classification with context utterance and classification without context utterance are in one-to-one correspondence, Paired-Samples *T*-Test can be performed on them to determine whether the results of the two experiments are significantly different. The results of the comparison are organized as shown in Table 5.

Table 5. Comparison between with and without context utterance.

Folds	Performance measures		Goodness of fit test		Effect comparison	Matched
	accuracy		χ^2			T-test
	with C.	without C.	with C.	without C.		Sig.
fold 1	0.640	0.573	11.760	3.227	1 > 2	0.105
fold 2	0.580	0.660	3.840	15.360	1 < 2	0.064
fold 3	0.600	0.580	6.000	3.840	1 > 2	0.687
fold 4	0.540	0.613	0.960	7.707	1 < 2	0.153
total	0.590	0.607	19.440	27.307	1 < 2	0.473

In Table 5, the accuracy and goodness-of-fit test statistics of the two experiments are first compared. It can be found that in the four-fold cross-validation, the situations of “with context utterance” (with C.) greater than and less than “without context utterance” (without C.) appear at the same time. The results of the Paired-Samples *T*-Test show that there is no significant difference in each fold of cross-validation (*p*-values are all greater than 0.05), and the overall significance is $0.473 > 0.05$. It shows that overall, the impact of lexical features on the text classification of conversational implicatures is not significantly different when including “context utterance” or not.

The above results show that when classifying texts based on lexical features and using conversational implicatures as labels, context is not involved as a factor, and the classification is only done based on the literal features of response utterance. This shows that computers take a completely different path from humans when using lexical features to judge conversational implicatures. The core contextual information in pragmatics for judging conversational implicatures is not used in the text classification based on logistic regression. However, according to the conclusion of the text classification of conversational implicatures without context utterance, that is, lexical features have a significant impact on the binary classification of conversational implicatures, it is proved that literal features based on the language itself also play an important role in judging particularized implicatures. This computational approach, rarely covered in traditional pragmatics, can expand methods of reasoning about implicatures. To a certain extent, it proves the rationality of computational pragmatics as an effective supplement to classical theories of conversational implicatures.

The Contingency Relationship Between Context and Implicature in the Case of Inconsistent Classification Results

The previous experiments give the overall test results of the text classification of conversational implicatures in the situations of with and without context utterance. Besides, there is another important issue worth considering, that is, from the perspective of the accuracy of classification results, whether the inclusion of context utterance and the type of conversational implicatures

Table 6. Results of frequency statistics and contingency analysis for context and implicature.

Folds	1			-1			Contingency analysis	
	yes	no	total	yes	no	total	value	sig.
fold 1	14	10	24	6	8	14	0.849	0.357
fold 2	11	4	15	12	15	27	3.249	0.071
fold 3	17	12	29	11	15	26	1.460	0.227
fold 4	12	12	24	20	15	35	0.293	0.589
total	54	38	92	49	53	102	2.206	0.138

are related. If they are related, different corpora should be chosen when targeting texts of different types of implicatures; if they are irrelevant, the selected method and corpus are universal, and a unified approach can be adopted to classify the text of various implicatures. To study this problem, we need to pay special attention to those cases where the judgment is different between “with context” and “without context.” To do this, a new variable needs to be defined, that is, “score difference.”

The “score difference” is defined as follows: if the classification is correct in “with context” but incorrect in “without context,” then score difference = 1; if the classification is incorrect in “with context” but correct in “without context,” then score difference = -1; if the classification is correct or incorrect in both “with context” and “without context,” then score difference = 0. According to the definition, when the score difference = 0, it means that the classification results of the two groups of experiments are consistent, and there is no need to care too much. When score difference = ± 1 , it means that the classification results of the two groups of experiments are different, and the circumstances under which this discrepancy arises need to be analyzed in detail. Therefore, this experiment only considers the case where the classification results of the two groups of experiments are different, that is, only statistical analysis is performed for the data with score difference = ± 1 .

By performing contingency analysis on the score difference and the type of conversational implicatures, the corresponding frequencies and their significance test results can be obtained, as shown in Table 6.

According to Table 6, the p -values of all the significance tests of contingency analysis are greater than 0.05, which suggests there is no correlation between the inclusion of context utterance and the type of conversational implicatures. So, the unified and universal methods and corpus can be used to classify the text of various implicatures, and there is no need to consider the type of implicature.

Discussion

This section discusses two issues: first, answer the three questions posed in “introduction;” second, discuss which type of text is more effective in the practice of natural language processing.

Answers to Research Questions

Now, the three research questions raised in the “introduction” are answered based on the experimental results. The first question is whether the text classification of conversational implicatures can be performed based only on lexical features. The answer to this question is yes. According to the experimental results of “with context utterance,” the text classification of conversational implicatures can be performed based on the feature set F_1 , and the accuracy rate is 59% ($\chi^2 = 19.440$, Sig. = 0.000). According to the experimental results of “without context utterance,” the text classification of conversational implicatures can be performed based on the feature set F_2 , with an accuracy rate of 60.7% ($\chi^2 = 27.307$, Sig. = 0.000). This suggests that different types of conversational implicatures can be distinguished only by lexical features, regardless of whether “context utterance” is included. Thus, although conversational implicatures are highly context dependent (Potts 2005, 25), through representations of sentences and words may not only be able to calculate semantic similarity (Ahmad and Faisal 2022), but also characterize the similarity of conversational implicatures.

The second question is what features are needed to classify conversational implicatures based on lexical factors. Comparing the two experiments, it can be found that whether “context utterance” is included or not affects the selection of optimal classification features. If “context utterance” is included, the Optimal Feature Set F_1 for classification contains 17 features; If “context utterance” is not included, the Optimal Feature Set F_2 contains only 5 features. Comparing Tables 1 and 3, it can be found that the feature set F_2 is a subset of F_1 ; that is, the features of F_2 have all appeared in F_1 . Therefore, if conversational implicatures are classified based on lexical factors, the most needed features are the statistics of the number of letters and syllables and the frequency for content words, as presented by the feature set F_2 .

The third question is whether the results of text classification will be different because of the difference in including “context utterance” or the type of implicature. The answer to this question is no. According to the results of the Independent-Samples *T*-Test in Tables 2 and 4, there is no significant difference in the classification results between different types of implicatures. According to the results of the Paired-Samples *T*-Test in Table 5, the classification results are also not significantly different between texts with or without context utterance. According to the results of the contingency analysis in Table 6, even if the two variables of context utterance and the type of implicature are considered at the same time, there is no significant difference in the local results between them. In summary, the results of text classification will not be different due to the difference in context utterance or the type of implicature.

What Kind of Corpus Should Be Selected in the Classification of Implicatures?

The next question that needs to be discussed is which text should be selected in the specific practice of classification. First, according to the answer to the third research question, the results of text classification will not be different due to the difference in context utterance or the type of implicature, that is, a unified method and corpus can be used to classify the text of various implicatures. That is, this kind of classification has universality.

Under the guarantee of this universality, we further examine the feature numbers of “with context utterance” and “without context utterance.” Obviously, the number of features contained in the Optimal Feature Set F_1 is 17, which is greater than the number of features contained in the Optimal Feature Set F_2 , which is 5. Generally, if the classification performance is similar, the smaller the number of features, the simpler the model, and the more effective each feature is. Therefore, it is more effective to select the text of “without context utterance” and adopt the Optimal Feature Set F_2 . And according to the accuracy of the two methods and the χ^2 statistic of the goodness-of-fit test, the result of “without context utterance” is slightly better than the result of “with context utterance.” This further suggests that using only “response utterance” for the classification of implicatures is a better choice. Therefore, when performing the text classification of conversational implicatures based on lexical features, there is no need to add “context utterance,” and the text can be limited to “response utterance.” This can explain, to a certain extent, the difference between the way computers deal with conversational implicatures and the way linguists do. From the perspective of “meaning,” pragmatic scholar (Kecskes 2008) argued that meaning is mostly dependent on context, and there is no doubt that conversational implicature is no exception. But the computer’s processing of meaning is almost entirely based on linguistic forms, and what is most sought after in AI are ways of representing resource extraction in symbols (Kavanagh 2022; Monte-Serrat and Cattani 2021, 177). So for now, the computational processing of conversational implicatures is still mainly based on the lexical features and various statistics in response utterances. This difference in research perspectives, to a certain extent, has led to the divergence of human and computer research paths of “meaning.” From the experimental results of this study, the computer’s processing path for conversational implicatures based on the language form also passed the chi-square test, which verifies the validity of this computational model and method to some extent.

Conclusions

The conclusions drawn from this study are summarized as follows: First, after a goodness-of-fit test with equal expected frequency, this paper proves that there is a statistical dependency between lexical features and conversational

implicatures. So, the text classification of conversational implicatures can be done only with lexical features. Second, in the classification process, the most effective lexical features contain both the statistics of the number of letters and syllables and two frequencies for content words. They have a significant effect on the classification of implicatures. Finally, the results of text classification do not differ due to the context utterance or the type of implicature. Therefore, common methods and corpora can be used to classify implicatures automatically.

Future research directions can be considered from the following perspectives: First, adopt a more suitable classification algorithm or carry out algorithm improvement aimed at implicature. Second, expand the amount of data. When the amount of corpus is larger, its laws and knowledge will be more obvious. Third, replace the feature. This paper uses lexical features. In the future, other features such as in syntax or discourse analysis can be further considered, and for pragmatic researchers, an important task is to propose new features according to theories of pragmatics, and then the classification of conversational implicatures based on new features can be carried out, which has the potential to improve the classification performance greatly.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Xianbo Li  <http://orcid.org/0000-0002-2949-2753>

Data Availability Statement

The authors confirm that the data supporting the findings of this study are available within the article and its supplementary materials. The data named “LexicalData.rar” is also available at <https://github.com/Albert-Dillon/CompImplicature>.

References

- Ahmad, F., and M. Faisal. 2022. A novel hybrid methodology for computing semantic similarity between sentences through various word senses. *International Journal of Cognitive Computing in Engineering* 3:58–77. doi:10.1016/j.ijcce.2022.02.001.
- Cappelen, H., and E. Lepore. 2005. *Insensitive semantics: A defense of semantic minimalism and speech act pluralism*. Oxford: Blackwell Publishing Ltd.
- Clark, B. 2013. *Relevance theory*. Cambridge: Cambridge University Press.
- Feng, W.-S., H.-B. Yu, and X.-L. Zhou. 2021. Understanding particularized and generalized conversational implicatures: Is theory-of-mind necessary? *Brain and Language* 212:104878. doi:10.1016/j.bandl.2020.104878.

- Frank, C., and N. Goodman. 2012. Predicting pragmatic reasoning in language games. *Science* 336 (6084):998. doi:10.1126/science.1218633.
- Garassino, D., N. Brocca, and V. Masia. 2022. Is implicit communication quantifiable? A corpus-based analysis of British and Italian political tweets. *Journal of Pragmatics* 194:9–22. doi:10.1016/j.pragma.2022.03.024.
- George, J., and R. Mamidi. 2020. Conversational implicatures in English dialogue: Annotated dataset. *Procedia Computer Science* 171:2316–23. doi:10.1016/j.procs.2020.04.251.
- Goodman, D., and M. Frank. 2016. Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences* 20 (11):818–29. doi:10.1016/j.tics.2016.08.005.
- Graesser, C., D. McNamara, Z.-Q. Cai, M. Conley, H.-Y. Li, and J. Pennebaker. 2014. COH-METRIX MEASURES TEXT CHARACTERISTICS at MULTIPLE LEVELS of LANGUAGE and DISCOURSE. *The Elementary School Journal* 115 (2):210–29. doi:10.1086/678293.
- Graesser, C., D. McNamara, and J. Kulikowich. 2011. Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher* 40 (5):223–34. doi:10.3102/0013189x11413260.
- Graesser, C., D. McNamara, M. Louwerse, and Z.-Q. Cai. 2004. Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods Instruments & Computers* 36 (2):193–202. doi:10.3758/bf03195564.
- Hoek, J., and M. Scholman. 2017. Evaluating discourse annotation: Some recent insights and new approaches.
- Huang, Y. 2014. *Pragmatics*. 2 ed. Oxford: Oxford University Press.
- Huang, Y. 2017. Implicature. In *The Oxford Handbook of Pragmatics*, ed. Y. Huang, pp. 155–79. Oxford: Oxford University Press.
- Kavanagh, R. 2022. Fairness and communication-based semantics for session-typed languages. *Information and Computation* 285: 104892. doi: <https://doi.org/10.1016/j.ic.2022.104892>
- Kecskes, I. 2008. Dueling contexts: A dynamic model of meaning. *Journal of Pragmatics* 40 (3):385–406. doi:10.1016/j.pragma.2007.12.004.
- Kecskes, I. 2021. Processing implicatures in English as a Lingua Franca communication. *Lingua* 256:103067. doi:10.1016/j.lingua.2021.103067.
- Lassiter, C. 2021. Particularized conversational implicatures and miraculous communication. *Language Sciences* 86:101401. doi:10.1016/j.langsci.2021.101401.
- Li, H. 2022. *Methods of machine learning*. Beijing: Tsinghua University Press.
- Li, X.-B., H.-R. Qiao, Z.-C. Ma, D.-D. Yang, Y.-T. Pan, Z.-X. Ma, and Machinery Assoc Comp. 2018. “Bayesian belief network model of indirect speech act theory.” International Conference on Artificial Intelligence and Virtual Reality (AIVR)/International Conference on Big Data and Machine Learning (BDML), Nagoya, JAPAN, Nov 23-25.
- McCarthy, M., and S. Jarvis. 2010. MTL, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods* 42 (2):381–92. doi:10.3758/brm.42.2.381.
- Monte-Serrat, M., and C. Cattani. 2021. *The natural language for artificial intelligence*. Cambridge: Academic Press.
- Potts, C. 2005. *The logic of conventional implicatures*. Oxford: Oxford University Press.
- Sbisà, M. 2021. Presupposition and implicature: Varieties of implicit meaning in explicitation practices. *Journal of Pragmatics* 182:176–88. doi:10.1016/j.pragma.2021.05.027.
- Turney, D., and P. Pantel. 2010. From frequency to meaning: Vector space models of semantics. *The Journal of Artificial Intelligence Research* 37:141–88. doi:10.1613/jair.2934.
- Wylie, E., S. George, K. McWilliams, A. Evans, and S. Stolzenberg. 2022. Children’s acquiescence to polysemous implicature questions about coaching: The role of parental support. *Journal of Applied Developmental Psychology* 78:101370. doi:10.1016/j.appdev.2021.101370.
- Zhou, Z.-H. 2021. *Machine learning*. Singapore: Springer Nature Singapore Pte Ltd.