applied sciences

MDPI

Formatted: Justified

Article

Source: On the Feature Selection and Classification Based on Information Gain for Document Sentiment Analysis by Asriyanti Indah Pratiwi and *Adiwijaya, used under CC-BY*

~~On the~~ Feature Selection and Classification Based on Information Gain for Document Sentiment Analysis

Firstname Lastname [1], Firstname Lastname [2] and Firstname Lastname [2,*]

[1] Affiliation 1; e-mail@e-mail.com
[2] Affiliation 2; e-mail@e-mail.com
* Correspondence: e-mail@e-mail.com; Tel.: (optional; include country code; if there are multiple corresponding authors, add author initials)

Abstract:

~~Sentiment analysis in a movie review is the needs of today lifestyle~~Sentiment analysis for movie reviews is an increasingly important tool. Unfortunately, the ~~enormous~~massive number of features involved often ~~make~~ causes ~~the~~ sentiment ~~of~~ analysis to become slow and ~~less sensitive~~lose sensitivity. ~~Finding the optimum~~Optimal feature selection and classification is still a significant challenge. ~~In order t~~To handle an ~~enormous~~ large number of features and provide better sentiment classification, an information-based feature selection and classification method ~~are~~ is proposed. The proposed method ~~reduces~~removes more than 90% of unnecessary features, ~~unnecessary features~~ ~~while~~and ~~whereas~~ the proposed classification scheme achieves 96% accuracy ~~of~~ for sentiment classification, whereas previous works have typically reached only 70–88% accuracy. From the experimental results, it can be concluded that the ~~combination of~~ proposed combined feature selection and classification method achieves ~~the best performance so far~~better performance than other methods of this type.

Commented [A1]: I have edited this document for language, grammar, clarity, tone, and logical structure. Please ensure that you read through all of my comments, as some highlight important issues that may still need to be addressed.

I have also ensured that this document matches the template style used by the target journal. Note that this required changing several section headings and their corresponding numbering. Aside from that change, no major structural revisions were required.

Commented [A2]: I have removed "On the" at the beginning of this title, as it made the title unnecessarily longer without adding any information. Alternatively, for a more concise title, you could write "Using Information Gain to Improve Document Sentiment Analysis."

Commented [A3]: Please replace these placeholders with the appropriate author information before submission. Please also do t ...

Commented [A4]: Please consider including the rationale behind the current study in order to strengthen the abstract. I strongly recommend ...

Formatted: MDPI_1.7_abstract, Indent: Left: 0", First line: 0", Line spacing: single

Commented [A5]: I have rewritten this sentence, as the original text was somewhat vague and ungrammatical. ...

Commented [A6]: I have reworded this for clarity.

Commented [A7]: This was slightly redundant wording.

Commented [A8]: It is best to have a brief comparison to previous works here, to put into perspective how much of an improvement 96% ...

Keywords: keyword 1; keyword 2; keyword 3 (List three to ten pertinent keywords specific to the article yet reasonably common within the subject discipline.)

1. Introduction

One ~~of the interesting challenges~~notable challenge in text categorization is sentiment analysis, a ~~study~~ process that analyzes the subjective information of specific ~~object~~ objects [1]. Sentiment analysis can be applied ~~on~~ at various levels, that is~~:~~ the document level, sentence level, and feature level.

Sentiment-based categorization in ~~the movie review~~movie reviews involves ~~is a~~ document-level sentiment analysis. ~~It~~ This method treats ~~the~~ a review as a set of independent words by ignoring the sequence of words ~~on a~~in the text. Every ~~single~~ unique word and phrase can be used as ~~the document features~~a document feature. As a result, ~~it~~ this type of sentiment analysis constructs a massive ~~numbers~~ number of features. ~~In addition, it~~ This abundance of features ~~also~~ slows down the process and ~~makes~~ can introduce bias in the classification task ~~bias~~ [2].

~~Actually~~However, not all features are necessary;~~.~~ ~~Most~~most ~~of the~~ features are irrelevant to the class label. ~~On the other hand,~~Thus, a good feature for classification is ~~the~~ one that has ~~maximum~~ high relevance ~~with~~ to the output class.

As feature selection is a crucial component of ~~in~~ sentiment analysis ~~is a crucial part~~, in this paper, we ~~proposed~~ propose an information ~~gain-based~~gain (IG)-based feature selection method. In addition, we ~~also proposed~~ propose classification schemes based on the dictionary that is constructed by the selected features.

Citation: Lastname, F.; Lastname, F.; Lastname, F. Title. *Appl. Sci.* **2021**, *11*, x. https://doi.org/10.3390/xxxxx

Academic Editor: Firstname Lastname

Received: date

Accepted: date

Published: date

~~1. Previous Work~~

There are two common approaches to sentiment analysis: machine learning methods and knowledge-based methods. Cambria [3] suggested ~~the~~ a combination of both methods,~~:~~ using machine learning to provide the limitations of ~~the~~ sentiment knowledge. ~~On theother hand~~However, ~~it~~this technique cannot be applied ~~in~~ to movie ~~review~~reviews. ~~The sentiment~~Sentiment knowledge, such as that provided by SenticNet, is highly dependent on domain and context. For example, the word "funny" ~~means~~ has a positive connotation for a comedy movie, but a negative connotation for a horror movie [4].

Machine learning-based sentiment analysis ~~on~~ of movie ~~review~~ reviews was ~~initialized~~ first performed by Pang et al. [5]. Their work ~~performed~~ achieved 70~~%~~–80% accuracy, while the human ~~baselines~~ baseline sentiment analysis method only ~~reaches~~ reached 70% accuracy at most. In 2014, Dos Santos and Gatti [6] used a deep learning method for sentence-level sentiment analysis, reaching ~~that reached~~ 70~~%~~–85% accuracy. Words and characters ~~are~~ were used as sentiment features. Unfortunately, the massive number of constructed features resulted in a ~~a long time computation~~long computation time.

T~~In order t~~o provide robust machine learning classification, a feature selection technique is required [7]. Some researchers ~~focus~~ have focused on reducing the number of features [8]. Manurung [9] proposed a feature selection scheme named feature-count (FC). FC selects the *n*-top subfeatures with the highest frequency count~~.~~, an operation which ~~It~~ ~~only costs~~has a time complexity of $O(n)$ ~~to select the subfeatures~~. ~~O then contrary~~However, ~~it~~ this method may select a feature ~~which~~ that has no relevance to the

Commented [A9]: Possible keywords you could use include "Sentiment analysis", "feature selection", and "information gain." Please ensure that this section is filled out appropriately.

Note that the target journal also recommends adding a "Featured Application" section after the Keywords: "Authors are encouraged to provide a concise description of the specific application or a potential application of the work. This section is not mandatory."

Commented [A10]: You have provided the relevant background information, knowledge gap, or limitations of prior studies in the introduction section. You have also focused on the goal of your study.

Commented [A11]: The word "interesting" sounds somewhat subjective or personal for an academic paper.

Commented [A12]: As before, I recommend adding a sentence or two discussing the ...

Commented [A13]: This word is redundant in ...

Commented [A14]: I have revised this to the ...

Commented [A15]: This study is 27 years old. ...

Commented [A16]: This is a somewhat informa ...

Commented [A17]: "Maximum" in this context ...

Commented [A18]: Phrases of the form "[noun] ...

Commented [A19]: This section heading is not ...

Formatted: MDPI_2.1_heading1, Indent: Left: 1.81", No bullets or numbering, Keep with next

Commented [A20]: This is a rather unclear phra ...

Commented [A21]: I recommend adding a ...

Commented [A22]: I recommend using a more ...

Commented [A23]: Please ensure that this ...

Commented [A24]: Small revision to ...

Commented [A25]: Please double-check to ensu ...

Commented [A26]: I suspect that you may have ...

output class, since a high frequency of occurrence does not necessarily indicate high relevance to the output class.

The works of Nicholls and Song [8] ~~research~~ and OKeefe and Koprinska [10] ~~research~~ proposed a similar idea to select features based on the difference between document frequency (DF) in class positive and DF in class negative. ~~It~~ This method was named Document Frequency Difference (DFD). DFD selects the feature that has the highest proportion between the positive ~~DF~~vs. negative DF difference and the total number of documents. ~~Their research may~~This approach may select ~~feature~~ features ~~which~~ that ~~has~~ have high differences in DF but are less relevant to the output class.

Information theory-based feature selection, using factors such as information gain or mutual information, ~~was~~ has also been proposed ~~in~~ for sentiment analysis [11, 12]. ~~In advance,~~ Abbasi et al. proposed a heuristic search procedure, named the entropy weighted genetic algorithm (EWGA), to search ~~optimum subfeature~~for optimal subfeatures based on ~~its~~ their information gain ~~(IG)~~ values ~~named Entropy Weighted Genetic Algorithm (EWGA)~~ [13]. EWGA ~~search~~ searches for optimal subfeatures using a ~~Genetic Algorithm~~genetic algorithm (GA) ~~which~~ with~~its~~ an initial population ~~is~~ selected ~~by~~ using ~~information gain~~IG ~~(IG)~~ thresholding schemes. Compared to ~~the~~ other options in this field, EWGA is the most powerful feature selection method ~~so far~~to date. ~~It~~ This approach selected features ~~that achieved~~with 88% ~~accuracy of classification~~classification accuracy. However, it ~~took high cost computation~~has a high computational cost.

~~This study uses polarity v.2.0 from Cornell review datasets, a benchmark dataset for document level sentiment analysis, that consists of 1000 positive and 1000 negative processed reviews [14]. This dataset split into tenfold crossvalidation.~~

2. ~~Information Gain on Movie Review~~Materials and Methods

2.1. Information Gain in Movie Reviews

Information gain is a quantity that measures how ~~mixed up~~well-organized the features are [15]. In the sentiment analysis domain, ~~information gain~~IG is used to measure the relevance of attribute A ~~in~~ to class C. The higher the value of mutual information between ~~classes~~ class C and attribute A, the higher the relevance between ~~classes~~ them ~~C and attribute A~~.

$$I(C, A) = H(C) - H(C \mid A), \qquad (1)$$

where ~~Where~~ $I(C, A)$ is the information gain, $H(C) = -\sum_{c \in C} p(C) \log p(C)$ is~~,~~ the entropy of the class~~,~~ and $H(C \mid A)$ is the conditional entropy of the class given an attribute, $H(C \mid A) = -\sum_{c \in C} p(C \mid A) \log p(C \mid A)$. Since the Cornell movie review dataset has balanced ~~class~~classes, the probability of class C for both positive and negative results is equal to 0.5. As a result, the entropy of ~~classes~~each class, $H(C)$, is equal to 1. Then, the information gain can be formulated as

$$I(C, A) = 1 - H(C \mid A). \qquad (2)$$

The minimum value of $I(C, A)$ occurs if and only if $H(C \mid A) = 1$, that is, ~~which means~~ attribute A and ~~classes~~ class C are not related at all. ~~On the contrary, we~~We ~~tend~~ attempt to choose an attribute A that mostly appears in one class C as either positive or negative. ~~On~~ For the other words, the best features are the set of attributes that only appear in one class. ~~It~~ This means that the maximum $I(C \mid A)$ is reached when $P(A)$ is equal to $P(A \mid C_1)$, resulting in $P(C_1 \mid A)$ and $H(C_1 \mid A)$ being equal to 0.5. When $P(A) = P(A \mid C_1)$, then the value of $P(A \mid C_2)$ results in $P(C_2 \mid A) = 0$ and $H(C_1 \mid A) = 0$. The value of $I(C, A)$ ~~is varied~~varies from 0 to 0.5.

Commented [A27]: As references have been excluded from this round of editing, I have avoided editing this name in order to avoid conflicts with the references later in the document. However, please double-check the spelling of this name, as it is usually written "O'Keefe" or "O'Keeffe."

Commented [A28]: If it suits your meaning, I recommend revising this to "in the positive class and that in the negative class" for clarity.

Commented [A29]: I recommend adding text clarifying *why* this approach does not necessarily work. For example, you could write "...but are less relevant to the output class, as high DF ...

Commented [A30]: I have deleted this phrase, as the intended meaning was unclear.

Commented [A31]: This term was not capitalized in the original work.

Commented [A32]: You have already defined this abbreviation, so there is no need to define it ...

Formatted: MDPI_3.1_text, Indent: Left: 0", Space After: 0 pt

Commented [A33]: This information is unnecessary in this section, particularly becaus ...

Commented [A34]: Consider adding notes this section about the computer equipment used in ...

Commented [A35]: Please ensure that this revision matches your intent.

Commented [A36]: You have not defined this term. Because it is often used interchangeably ...

Commented [A37]: I assume that you meant to refer to the singular "class" here, judging from ...

Commented [A38]: I have added this variable definition, as this term is not explicitly defined ...

Commented [A39]: Small revision for clarity; please ensure that this is correct.

Commented [A40]: This word choice was somewhat unclear. I have revised this under th ...

2.2. Sentiment Analysis Framework

This study uses the polarity dataset v~~.~~2.0 from Cornell's review datasets. This is~~,~~ a benchmark dataset for document-level sentiment analysis~~,~~ ~~that consists~~consisting of 1000 positive and 1000 negative ~~processed~~ reviews [14]. This dataset was split ~~into~~ for tenfold cross-validation.

Figure 1 shows the ~~process of~~ proposed sentiment analysis process. The process was categorized into a dictionary construction phase and a classification phase. ~~Dictionary~~ The dictionary construction phase constructs a dictionary that can be used to classify the ~~Review~~review as~~:~~ positive or negative. ~~Here are the~~The steps of the dictionary construction phase in this study are as follows: (1) reading the dataset, (2) nonalphabetic removal, (3) tokenization, (4) stopword~~s~~ removal, (5) stemming (optional), (6) initial vocabulary construction, (7) initial feature matrix construction, (8) DF thresholding, (9) information gain and DF thresholding feature selection (IGDFFS), and (10) dictionary construction.

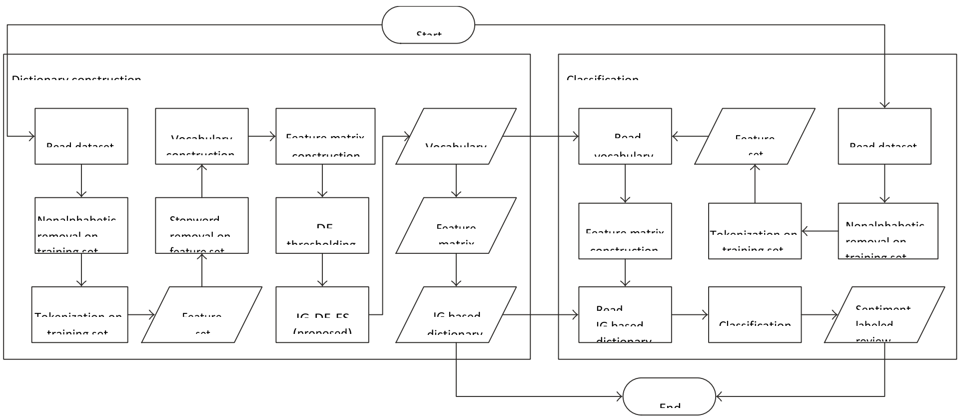


Figure 1: Classification flowchart of the proposed method.

Similar to the dictionary construction phase, the classification phase ~~also~~ consists of preprocessing and feature construction. ~~On the contrary,~~In contrast to the dictionary construction phase, it uses the constructed dictionary instead of selecting ~~feature~~ features and constructs another dictionary. ~~The result of this phase is sentiment labeled~~This phase yields sentiment labeling of movie ~~review~~reviews.

~~*4.1.*~~*2.3. IGDF Feature Selection*

~~.~~Previous work on information gain [16] selected~~s~~ ~~feature~~ features ~~that has~~ having high relevance ~~with~~ to the output class. ~~Those~~ These features commonly appear in positive ~~class~~ classes only or in negative ~~class~~ classes only. Unfortunately, ~~it~~ such features may appear only a few times, ~~since~~ as ~~the~~ a sentiment can be expressed in ~~a various way~~various ways. As a result, overfitting occurs ~~since~~ because those features do not appear frequently.

~~On the other hand~~In contrast, DF thresholding [8, 12] selects ~~feature~~ the features that ~~appears~~ appear most frequently in the training set. ~~It~~ However, this method may select ~~feature~~ features that always ~~appears~~ appear in both classes. ~~Those~~ Such features are unnecessary, as the method ~~since it~~ cannot ~~differentiate~~ determine the ~~class~~ classes to which ~~it~~ these features ~~belongs~~belong.

Commented [A41]: As this word modifies another noun ("removal"), the singular "stopword" is appropriate, even though multiple stopwords are presumably involved.

Commented [A42]: I have consolidated this text into a single paragraph; it was originally separated into two.

Commented [A43]: Consider adding some further details on the steps that you have not explained elsewhere, such as (5). The target journal requests a detailed explanation of all methods used.

Commented [A44]: Please note that the figure labels are not clearly visible here.

Commented [A45]: I have moved this figure (and others) to comply with the journal's guidelines, which state that figures "should be placed in the main text near to the first time they are cited."

Note also that the target journal requires that figures be submitted as separate files in a single .zip (in addition to their inclusion in the manuscript).

Further, please revise "IG-DF-FS (proposed)" to "IGDFFS (proposed)."

Commented [A46]: This is slightly redundant with "similar to".

Commented [A47]: I have revised this for consistency; please ensure that this matches your intent.

Commented [A48]: Minor language detail: "differentiate ... classes" suggests contrasting the *classes* themselves, rather than sorting features into one class or another.

In this study, we propose a combination of information gain and DF thresholding feature selection~~,~~ named IGDFFS. IGDFFS selects ~~a feature~~features that ~~has~~have IG ~~score~~ scores equal to 0.5,~~.~~ ~~It means~~indicating ~~those~~ features highly related to one class only. ~~These schemes~~This scheme ~~succeed~~ succeeds in ~~reducing~~ removing ~~about~~ approximately 90% of unnecessary features (Algorithm 1).

(1) **procedure** IGDF–Feature–Selection(input: {array of attributes A and its class C}, output: {positive and negative feature set})
(2) **for** *each feature~~s~~ in featureset* **do** (3) *calculate* $I(C \mid A)$
(4) **end for**
(5) **for** *each IGscore in* $I(C \mid A)$ **do**
(6) **if** $I(C \mid A) == 0.5$ **then**
(7) ~~*Vocabulary*~~ *Vocabulary* ← ~~*Vocabulary*~~ *Vocabulary* + *A*
(8) **if** $P(A) == P(A \mid C_{positiVe})$ **then**
(9) ~~*featuresetpositiVe*~~ *featuresetpositive* ← ~~*featuresetpositiVe*~~ *featuresetpositive* + *A*
(10) **else**
(11) ~~*featuresetnegatiVe*~~ *featuresetnegative* ← ~~*featuresetnegatiVe*~~ *featuresetnegative* + *A*
(12) **end if**
(13) **end if**
(14) **end for**
(15) **end procedure**

Algorithm 1: Information gain-document frequency (IGDF) feature selection.

~~4.2.~~2.4. Classification

~~. As it is known that entropy~~Entropy and information gain are commonly used in decision ~~tree~~trees. The selected ~~feature~~ features with the highest information gain ~~determines~~ determine the class of the review. Based on this intuition, we categorize our vocabulary into ~~the positive~~positive ~~feature~~ and negative ~~feature~~features. A review ~~will be~~is classified ~~into~~ as a positive review if most ~~of the~~ features are positive and vice versa (Algorithm 2).

(1) **procedure** IG-based–Classifier(input: {Sentiment Feature Vector: Vocabulary × Number of Document}, output: {Sentiment Label: positive or negative})
(2) **for** *each document in featurevector* **do**
(3) **for** *each vocabinVocabulary* **do**
(4) **if** ~~*Vocab*~~ *vocab is positive – features* **then**
~~(4)~~(5) ~~(5)~~ ~~*positiVe*~~ *positive* ← ~~*positiVe*~~ *positive* + 1
(6) **else**
(7) ~~*negatiVe*~~ *negative* ← ~~*negatiVe*~~ *negative* + 1
(8) **end if**
(9) **end for**
(10) **if** ~~*positiVe*~~ *positive* > ~~*negatiVe*~~ *negative* **then**
(11) *classlabel* ← *classlabel* + "*positive*"

Commented [A49]: You have not defined this abbreviation; I recommend doing so here, even though it is relatively clear through context.

Commented [A50]: Assuming that you are referring to IGDFFS specifically, the singular is appropriate here.

Commented [A51]: I have revised some of the formatting in this algorithm for stylistic consistency.

I also recommend changing the line indentations in this algorithm to a more standard style. Typically, the contents of a "for", "if", etc. block are indented, then terms such as "end for", "else", and "end if" are *not* indented, to clarify where each block starts and stops.

Commented [A52]: I have moved this Algorithm as well to comply with the aforementioned journal guidelines regarding figure positioning.

Commented [A53]: Note that the target journal requires abbreviations to be defined "the first time they appear in the abstract, main text, and in figure or table captions." I have revised this caption, as well as that for Algorithm 2, accordingly.

Commented [A54]: Judging from context, I assume that you meant this to be plural; please ensure that this is correct.

Commented [A55]: Please double-check this; I assume you meant "vocab in Vocabulary."

```
else
    class_label ← class_label + "negative"
end if
end for
end procedure
```

Algorithm 2: Information gain (IG)~~IG~~-based classification.

3. Results~~ and Analysis~~

Figure 2 shows the performance of an existing ~~previous~~ feature selection method, that is, the forward feature selection algorithm (FFSA) [16], and that of the proposed feature selection method, ~~(~~IGDFFS~~)~~. The results show that IGDFFS selects better features.

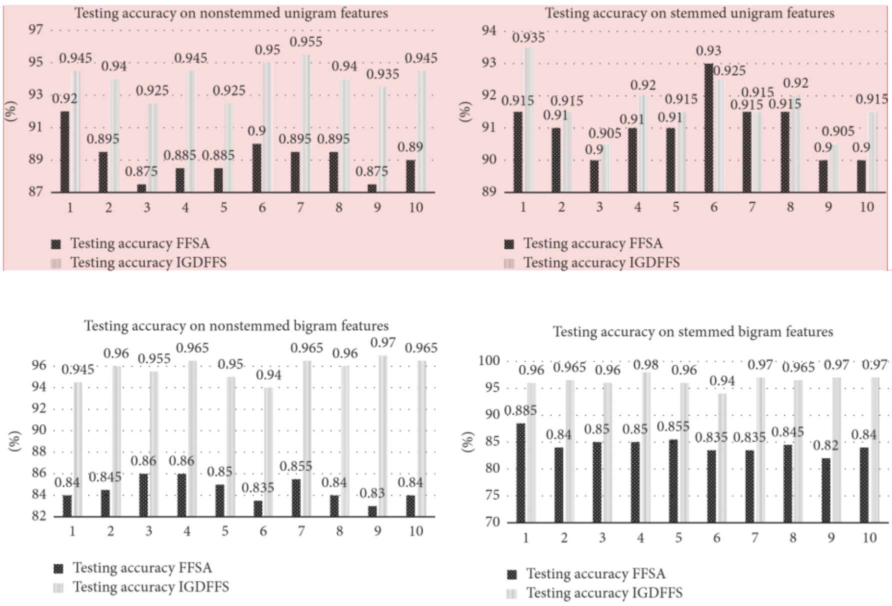


Figure 2: Feature selection performance comparison.

Commented [A56]: I am guessing that the entirety of "label" was meant to be written in subscript, rather than only the first letter. Please revise accordingly if this is correct.

Commented [A57]: I have retyped some of the text in this Algorithm to ensure standard formatting. Please ensure that these revisions match your intent.

Please also see my previous comment on Algorithm 1 regarding indentation.

Commented [A58]: This section clearly presents your results, makes effective use of the data. You have clearly indicated how the results can be interpreted in comparison with previous studies.

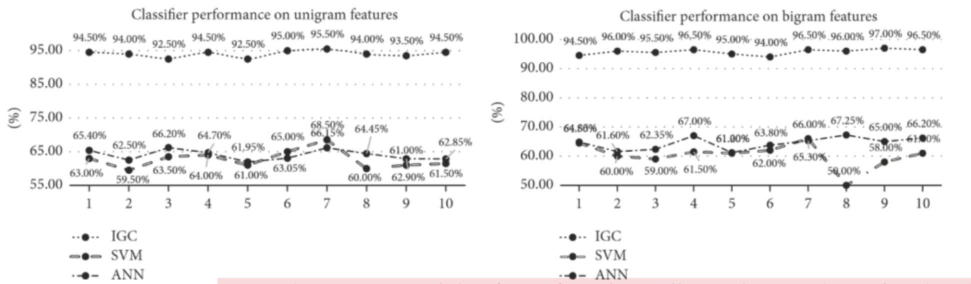
Commented [A59]: Please ensure that this revision and abbreviation definition match your intent.

Commented [A60]: Please revise the labels in these two graphs, as well as those of the two graphs below, by changing "Testing accuracy FFSA" to "Testing accuracy of FFSA" and "Testing accuracy IGDFFS" to "Testing accuracy of IGDFFS."

~~Proposed~~ The proposed method selects ~~feature that has~~features that have both high relevance to the output class and ~~also has the highest occurrence~~high occurrence rates. As a result, the generated feature matrix has less zero value. ~~On the contrary~~In contrast, the previous method may succeed in selecting ~~high~~ highly relevant features, but the selected features are likely to be rare ~~but probably takes rare features~~. ~~The~~ A rare feature does not appear in another ~~movie review~~ document in the training set and may not appear in the testing set. As a result, the generated feature matrix ~~consists of a lot of~~includes many zero ~~value~~values. ~~A lot of~~Many documents ~~which have not any~~without features are ~~hard~~ difficult to ~~be classified~~classify.

One ~~of the~~ feature selection ~~objectives~~ objective is to avoid overfitting, which often results from~~.~~ ~~Actually, in this case,~~ common machine learning techniques ~~may result in overfitting~~. ~~The reason is~~This is because the feature matrix in the testing set ~~consists of a lot of~~has many more zero values ~~more~~ than the feature matrix in the training set does. ~~Since~~ Because ~~the~~ these features affect machine learning ~~model~~models, ~~then~~ it is ~~hard~~ difficult for machine learning to fit the model to the feature matrix in the testing set.

Figure 3 summarizes the performance of the SVM, ANN, and IG ~~classifier~~classifiers. Unfortunately, SVM and ANN suffer from overfitting ~~problems. Their testing accuracy fails~~ and thus fail ~~in achieving~~to achieve 70% accuracy. ~~Different to~~Unlike ANN and SVM, the information gain classifier (IGC) is quite stable in ~~any condition~~all conditions. IGC ~~succeed~~ succeeds in avoiding overfitting ~~problems.~~; ~~It~~ it can be concluded that using the



proposed IGC ~~as proposed classifier performs better~~offers performance better than that of the current classifier.

Figure 3: Sentiment classifier performance comparison.

Information gain ~~value~~ ~~tells how mixed a feature to~~indicates the extent to which a feature is well-organized in a class ~~the class is~~. IG ~~value~~ reaches the highest value (0.5 in this case) when the feature belongs to one class only. ~~It~~ This means that when the feature appears, ~~we make sure that~~ the label must be positive or negative. In this case, the IG ~~value~~ of selected ~~feature~~ features achieves the maximum value (0.5) on average; thus, ~~(0.5) so,~~ it can be used for automatic classification. The ~~specialty~~ uniqueness of the proposed classification scheme ~~is the~~lies in its independence from mathematical ~~model~~models. Since the proposed classification method succeeds in avoiding overfitting, we ~~can say~~conclude that our method is ~~better~~ more effective than ~~the~~ those of previous ~~work~~works.

4. ~~Conclusion and Future Work~~Discussion

~~In order to~~To provide a better sentiment analysis system, ~~an~~ a ~~improvement~~ method of information gain-~~-~~based feature selection and classification was proposed. The proposed ~~feature selection~~method selects ~~feature that has~~features with high information

Commented [A61]: This is somewhat unclear. If you mean that there are not as many zero values in the feature matrix, this should be revised to “fewer zero values.” Otherwise, please clarify.

Commented [A62]: I believe that it would be more logical to revise this to “Rare features do not appear in most documents in the training set” or “A rare feature appears in only one document in the training set”, depending on which better matches your intended meaning.

Commented [A63]: This sentence is somewhat unclear, as it is not clear how a document could lack features entirely. Did you mean something such as “Documents that lack the selected features”? If so, please revise accordingly; if not, please clarify.

Commented [A64]: The phrase “a lot of” is rather informal.

Commented [A65]: It is somewhat unclear which classifier you are referring to here. I recommend clarifying this point, for example, by replacing this with “the SVM and ANN classifiers.”

Commented [A66]: While it is not necessary to restate the contents of the figure below, I recommend including a brief mention of IGC’s numerical accuracy. For example, “IGC achieves an average accuracy of XX% in these conditions ...

Commented [A67]: As before, please ensure that this revision matches your intent.

Commented [A68]: This is slightly unclear and potentially illogical. Writing that “the IG value of selected features [is] the maximum value (0.5) on average” suggests that the IG values of these ...

Commented [A69]: This was slightly unclear; please ensure that this revision matches your intent.

Commented [A70]: The target journal suggests that a Conclusions section should only be added if the Discussion section is particularly long; thu ...

gain and high occurrence. As a result, it succeeded in providing ~~feature that most probably~~features that were most likely to appear ~~appears~~ in testing ~~also~~as well. ~~Proposed~~ The proposed classifier used the positive and negative features obtained from the IG calculation before, performing its task more quickly~~. Then, it takes less time~~ than ~~the~~ previous ~~classifier~~ classifiers can (SVM, ANN, etc.).

~~The~~ A combination of information gain and document frequency ~~in this study~~ was proposed for feature selection in this study.~~;~~ IGDFFS selects subfeatures that satisfy ~~these~~ the following criteria: (1) high relevance to the output class and (2) high occurrence in the dataset. ~~As a result~~Thus, it constructs subfeatures that ~~reach better performance in the classification~~yield better classification performance.

Compared to ~~the current classifier~~current classifiers, the ~~Information Gain Classifier (IGC)~~IGC ~~overcomes the recent high accuracy which belongs to~~ has surpassed the high accuracy of EWGA (only 88.05%). ~~It~~ The IGC succeeded in avoiding overfitting problems in ~~any condition~~diverse conditions~~.~~, yielding ~~The~~ stable performance ~~of IGC is quite stable~~ in both training and testing.

~~We~~ For future work, we are considering ~~to groups~~ grouping ~~the~~ words based on their relevance to positive and negative reviews. Note that there are 171,476 words that are currently used and 47,156 obsolete words in the English domain (~~based~~ according to~~on~~ the Oxford English Dictionary). ~~At least a~~A ~~finite~~ limited number of groups would at least ~~be less~~represent a dataset smaller than the total ~~number~~ set of words.

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "Conceptualization, X.X. and Y.Y.; methodology, X.X.; software, X.X.; validation, X.X., Y.Y. and Z.Z.; formal analysis, X.X.; investigation, X.X.; resources, X.X.; data curation, X.X.; writing—original draft preparation, X.X.; writing—review and editing, X.X.; visualization, X.X.; supervision, X.X.; project administration, X.X.; funding acquisition, Y.Y. All authors have read and agreed to the published version of the manuscript." Please turn to the CRediT taxonomy for the term explanation. Authorship must be limited to those who have contributed substantially to the work reported.

Funding: Please add: "This research received no external funding" or "This research was funded by NAME OF FUNDER, grant number XXX" and "The APC was funded by XXX." Check carefully that the details given are accurate and use the standard spelling of funding agency names at https://search.crossref.org/funding. Any errors may affect your future funding.

Data Availability Statement: In this section, please provide details regarding where data supporting reported results can be found, including links to publicly archived datasets analyzed or generated during the study. Please refer to suggested Data Availability Statements in section "MDPI Research Data Policies" at https://www.mdpi.com/ethics. You might choose to exclude this statement if the study did not report any data.

Acknowledgments: In this section, you can acknowledge any support given which is not covered by the author contribution or funding sections. This may include administrative and technical support, or donations in kind (e.g., materials used for experiments).

Conflicts of Interest: The authors declare no conflict of interest.

~~Conflicts of Interest~~

~~The authors declare that there are no conflicts of interest regarding the publication of this paper.~~

References

[1] B. Agarwal and N. Mittal, Prominent Feature Extraction for Sentiment Analysis, Springer, 2015.

[2] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," IEEE Transactions on Neural Networks and Learning Systems, vol. 5, no. 4, pp. 537–550, 1994.

[3] E. Cambria, "Affective computing and sentiment analysis," IEEE Intelligent Systems, vol. 31, no. 2, pp. 102–107, 2016.

Commented [A71]: "Finite" is not an appropriate word choice here, as this word only means "not infinite", and of course using infinitely many groups is not possible.

Commented [A72]: Please ensure that this revision matches your intent.

Commented [A73]: Please fill out these sections as appropriate, replacing the placeholder text. "Acknowledgments" may be deleted if you have nothing to write in that section.

Commented [A74]: From the template: "Any role of the funders in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results must be declared in this section. If there is no role, please state 'The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results."

Please add text regarding the role of any funders, as detailed above.

Commented [A75]: I have not edited this section, as references were excluded from this round of editing.

[4] P. Chaovalit and L. Zhou, “Movie review mining: a comparison between supervised and unsupervised classification approaches,” in Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS’05), 112c pages, IEEE, 2005.
[5] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up?: sentiment classification using machine learning techniques,” in Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing, pp. 79–86, Association for Computational Linguistics, Stroudsburg, Pa, USA, July 2002.
[6] C. N. Dos Santos and M. Gatti, “Deep convolutional neural networks for sentiment analysis of short texts,” in Proceedings of the 25th International Conference on Computational Linguistics (COLING ’14), pp. 69–78, 2014.
[7] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, FeaturE Extraction: Foundations and Applications, vol. 207, Springer, 2008.
[8] C. Nicholls and F. Song, “Comparison of feature selection methods for sentiment analysis,” in Proceedings of the Canadian Conference on Artificial Intelligence, pp. 286–289, Springer, 2010.
[9] R. Manurung, “Machine learning-based sentiment analysis of automatic indonesian translations of english movie reviews,” in Proceedings of the International Conference on Advanced Computational Intelligence and Its Applications (ICACIA), Depok, Indonesia, 2008.
[10] T. OKeefe and I. Koprinska, “Feature selection and weighting methods in sentiment analysis,” in Proceedings of the 14th Australasian document computing symposium, pp. 67–74, Citeseer, Sydney, Australia, 2009.
[11] B. Agarwal and N. Mittal, “Text classification using machine learning methods-A survey,” in Proceedings of the Second International Conference on Soft Computing for Problem Solving (SocProS 2012), vol. 236 of Advances in Intelligent Systems and Computing, pp. 701–709, Springer, India, December 2012.
[12] M. Ikonomakis, S. Kotsiantis, and V. Tampakas, “Text classification using machine learning techniques,” WSEAS Transactions on Computers, vol. 4, no. 8, pp. 966–974, 2005.
[13] A. Abbasi, H. Chen, and A. Salem, “Sentiment analysis in multiple languages: feature selection for opinion classification in Web forums,” ACM Transactions on Information and System Security, vol. 26, no. 3, article 12, 2008.
[14] B.Pangand L.Lee,“Asentimentaleducation:sentiment analysis using subjectivity summarization based on minimum cuts,” in Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, 271 pages, Association for Computational Linguistics, Barcelona, Spain, July 2004.
[15] R. M. Gray, Entropy and Information Theory, Springer Science and Business Media, 2011.
[16] F. Amiri, M. M. R. Yousefi, and C. Lucas, “Mutual informationbased feature selection for intrusion detection systems,” Journal of Network & Computer Applications, vol. 34, no. 4, pp. 1184–1199, 2011.