DOI: 10.6977/IJoSI.202303_7(5).0003
D. Adrianto, D. Suryadi/Int. J. Systematic Innovation, 7(5), 30-45(2023)

Feature selection using binary particle swarm optimization algorithm to predict repurchase intention from customer reviews

Dimas Adrianto, Dedy Suryadi*
Industrial Engineering Department, Parahyangan Catholic University, Bandung, Indonesia
* Corresponding author E-mail: dedy@unpar.ac.id
(Received 31 March 2022; Final version received 12 August 2022; Accepted 15 January2023)

Abstract

Indonesia has the most prominent digital economy in Southeast Asia and has a promising market for e-commerce companies to compete and dominate the online market share. This also gave rise to an increment in the number of customer reviews of a product or service provided. Online customer reviews can be utilized to analyze the repurchase intention of e-commerce customers. However, many features appearing in customer reviews increased the repurchase intention predictive model complexity. A process to choose a subset of features and reduces the number of features in data is called feature selection. This paper proposed a method of feature selection to pre-process the inputs to the predictive model. The selection is performed using a metaheuristic called Binary Particle Swarm Optimization (BPSO) combined with Sentiment Orientation-Pointwise Mutual Information to sort the features. The sorting corresponds to the particle dimension, which is a part of the particle encodings that affect the metaheuristic's performance in solving the problem. The results show that the proposed method reduces and selects the best features to construct a predictive model of repurchase intention from online customer reviews on two datasets that are written both in Indonesian and English. Compared to the baseline model before performing feature selection, the accuracy of the predictive models evaluated using k-Nearest Neighbors on both datasets increased by 5.40% (75.91% to 81.31%) and 8.50% (71.37% to 79.87%), respectively.

Keywords: Binary Particle Swarm Optimization, Feature Selection, Online Customer Reviews, Repurchase Intention.

1. Introduction

Indonesia is the tenth largest economy in the world and is expected to have the largest digital economy in Southeast Asia and will likely reach $330 billion in value by 2030 (Chandra 2021). An article published by ISEAS in August 2021 (Negara and Soesilowati 2021) stated that Indonesia's digital economy has approximately 40% of the total regional market share. Rising internet penetration and the usage of smartphones in Indonesia are among the main factors contributing to the remarkable growth of e-commerce. The online market or e-commerce in Indonesia has continued its expansion during the COVID-19 pandemic. In addition, We Are Social in Digital 2021: Indonesia (DataReportal 2021) reported that, in January 2021, 87.1% of internet users in Indonesia aged 16 to 64 purchased a product online from any device. Many e-commerce companies have been competing to dominate the online market share in Indonesia. Note that, to dominate the market, companies must develop and improve the quality of products and services they offer and provide.

Repurchase intention is a customer's judgment on repurchasing the same products or services from the same company based on their experience (Hellier et al., 2003). Based on Social Learning Theory (SLT), self-efficacy influences people's decisions about what actions to take. People tend to learn from performance accomplishments (i.e., past experiences) to influence their actions. The role of self-efficacy is vital in the repeat buying behavior of e-commerce customers (Chen 2012).

Quality is an essential factor related to repurchase intention (Suryadi 2020). Customer satisfaction is crucial in mediating the effect of quality variables positively on customers' repurchase intention, such as product quality (Vashti and Antonio 2021). This argument is also supported by the research stating that product quality positively affects repurchase intention and is associated with customer satisfaction. In addition, price perception, through customer satisfaction, affects repurchase intention (Suhaily and Soelasih 2017). The previous research implies that quality and price perception are associated with customer satisfaction. When customers are satisfied, customers tend to make repeat purchases.

Most research and papers on repurchase intention or customer satisfaction rely on common traditional data collection methods, such as interviews (Mendoza 2020), surveys (Nguyen et al., 2021) (Trivedi and Yadav 2018), questionnaires (Tsai et al., 2016). These methods, however, are time-consuming and costly to carry out. This paper offers a relatively new method of predicting repurchase intention using online customer reviews. Customer review contains textual content on customers' opinion of products or services they had experienced that is related to their repurchase intention.

Nowadays, machine learning is an important technique in a broad area, including e-commerce and marketing. In marketing, machine learning offers advantages and is usually compared to traditional methods such as econometric methods. The role of machine learning techniques allows e-commerce to make strategic and crucial decisions on time. Moreover, machine learning provides valuable insights for e-commerce marketers and product manufacturers to improve their products or services (Suryadi 2020). Furthermore, feature selection and optimization may be applied to machine learning to achieve higher efficiency and performance for many problems, including marketing problems (Brei 2020).

Customer reviews are documents that may be represented as a collection of words or bag-of-words (features). These documents are high-dimension vectors where each dimension corresponds to the number of features (Nedjah et al., 2009). To solve this complexity, feature selection may be proposed to reduce the number of irrelevant and redundant features in the feature space while improving the accuracy of the constructed prediction model. In machine learning and statistics, feature selection is a technique to reduce the dimensionality of data by choosing subsets that consist only of relevant features by removing irrelevant, redundant, or noisy features from the original feature set (Miao and Niu 2016). Feature selection has been applied in broad practical applications, such as image processing (Bins and Drapper 2001), bioinformatics (Saeys et al., 2007), text mining (Forman 2003), and (Du et al., 2019).

Text mining has become more popular as it tries to gather valuable information from textual data (Dang and Ahmad 2014). Forman (2003) performed an experimental study on 229 text classification problem instances of twelve feature selection methods. Du et al. also performed feature selection on textual problems, specifically on online customer reviews. However, with the emergence of textual analysis, especially on online customer reviews, the research on analyzing Indonesian textual content, such as customer reviews, is still limited. This paper uses customer reviews (written both in Indonesian and English) collected from the Indonesian cosmetics e-commerce, sociolla.com.

This research aims to apply feature selection to textual features (i.e., customer reviews) using Binary Particle Swarm Optimization (BPSO). Feature selection on textual features is used to help to construct a better and more efficient predictive model. The Binary Particle Swarm Optimization (BPSO) algorithm is a version of Particle Swarm Optimization that has been used in binary problems (Xue et al., 2014). Examples of binary problems are those involving labels such as "yes" or "no", "included" or "not included" (Khanesar et al., 2017). The subsequent section will further detail the arguments for selecting BPSO.

Similar previous works on selecting textual features using BPSO have the purposes of sentiment classification and text summarization (Shang et al., 2016; Suganya & Priya, 2017; Suganya et al., 2019). Those purposes are different from this paper's purpose, i.e., classifying customer repurchase intention. Customers who intend to repurchase may express positive and negative sentiments in their reviews. That is also the case for customers who do not intend to repurchase. Therefore, this paper's purpose is different from sentiment classification in the previous works.

Moreover, considering the essential role of encoding in metaheuristic algorithms, this paper proposes a novel way to sort the features (which correspond to the dimensions of a particle in BPSO) according to a measure called Sentiment Orientation-Pointwise Mutual Information (SO-PMI). Encoding the particle in BPSO is important because encoding influences the algorithm's performance in solving the problems (Zavala et al., 2014). More specifically, the importance is due to the dependence of the movement or variation operators in BPSO on the encoding (Osaba et al., 2022).

The remainder of this paper is divided into four sections as follows. In Section 2, literature reviews are presented. Section 3 presents the methodology used in this research. Section 4 presents the results of a case study and a discussion of the experiments in this research, followed by conclusions and potential future works in Section 5.

2. Literature reviews

2.1 Metaheuristics for feature selection

Feature selection is a technique for choosing a subset of features in data used during the pre-processing step (Cherrington, et al. 2019). Feature selection has two objectives: first, it is to reduce the number of irrelevant and redundant features (Bing et al., 2013) and to improve model significance and performance (Cherrington, et al. 2019). Feature selection is arduous as there can be a complex interaction between features and a large search space (Bing et al., 2013).

Feature selection may be performed by a number of methods. There are three main categories of feature selection algorithms: filter approaches, wrapper approaches, and hybrid approaches. The filter approach uses independent criteria. The wrapper approach, such as the metaheuristics algorithm, is bound to the predetermined classifying algorithms. In other words, this approach considers the interaction between the classifying algorithm and the metaheuristic algorithm (Shroff and Maheta 2015). The hybrid approach combined the filter approach and wrapper approach. Research done by Kohavi and John (Kohavi and John 1997) concludes that wrapper approaches are superior to filter approaches, such as tf-idf, even though filter approaches are arguably less expensive computationally.

Metaheuristic algorithms, such as Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), and Genetic Algorithm (GA), are methods that may obtain a near-optimal solution to a complex problem effectively. The binary version of Particle Swarm Optimization (BPSO) is chosen in this paper, considering that PSO computing time is the fastest among the three algorithms, even though GA generates a better performance (Gunantara and Putra 2019). Also, compared to GA, PSO is easier to implement to its few parameters and can converge faster (Cervante et al., 2012).

Compared to ACO, research that compared ACO, PSO, and a proposed hybrid method of ACO-PSO on a feature selection problem by Menghour and Souci-Meslati in 2016 showed that the simple PSO approach is the second best method in general, after the hybrid ACO-PSO (Menghour and Souici-Meslati 2016). The effectiveness of PSO on feature selection is also supported by survey research on feature selection using five different binary metaheuristic algorithms (Binary Particle Swarm Optimization, Binary Differential Evolution, Binary Antlion Optimizer, Binary Grey Wolf Optimizer, and Binary Gaining Sharing Knowledge Based Algorithm). Even though, in general, the Binary Gaining Sharing Knowledge Based Algorithm performs better than the other four, Binary Particle Swarm Optimization (BPSO) takes relatively less computational time (depending on the size of the datasets) (Agrawal et al., 2021). Due to these results and advantages, PSO is considered a promising method for feature selection.

Other algorithms, such as random forest, AdaBoost, and gradient boosting, also can be utilized for feature selection. However, to the best of our knowledge, AdaBoost and other boosting algorithms select features based on their importance. The top-ranked features (based on their individual feature importance) are selected (Sun et al., 2011). These algorithms above do not generally consider the interaction between features. Wrapper method such as metaheuristic algorithms (Binary PSO in this case) has the ability to consider and capture the interaction between features.

Table 1 summarizes the examples of literature reviews on feature selection. Suryadi (Suryadi 2020) proposed a machine learning-based method to predict repurchase intention using online customer reviews on the cosmetics dataset collected online. The filter approach (tf-IDF and Fisher score) was used to represent a review's textual content and reduce the number of features. Then, three classification models were performed, and the results were significantly higher in the accuracy of three categories of products compared to the baseline model.

Particle Swarm Optimization was used on the feature selection problem. Bing et al. applied PSO to perform feature selection on fourteen datasets collected from the UCI machine learning repository, plus six additional datasets. Kristiyanti and Wahyudi (Kristiyanti and Wahyudi 2017) proposed a method of using the wrapper approach for feature selection on opinion mining cosmetic product reviews. They compared the classification performance of Principal Component Analysis (PCA), Particle Swarm Optimization (PSO), and also Genetic Algorithm (GA). The method proposed in their paper is to use those three algorithms mentioned before so the accuracy of the Support Vector Machine (SVM) as a machine learning classifier algorithm in text classification can be increased. The results showed that the PSO-based SVM algorithm outperformed the other algorithms mentioned above.

Shang et al. (Shang, Zhou and Liu 2016) proposed a method using Binary PSO on feature selection. They modified it to overcome feature selection problems in sentiment classification, including unreasonable velocity update formula and lack of evaluation of a single feature. The modification is called Fitness-sum Proportionate Selection Binary Particle Swarm Optimization (FS-BPSO). The results indicated that FS-BPSO performed better than BPSO.

Suganya et al. (Suganya, Lavanya, & Gowrisankari, 2019) focused on using Fitness Based Particle Swarm Optimization (FBPSO) to select subsets of features for sentiment classification and summarization problems on a hotel review dataset. According to their experiments, FBPSO improved the performance using ROUGE-N metric as a performance evaluation of summary compared to the probabilistic ranking approach. Suganya and Priya (2017) also proposed a similar method on a hotel review dataset using a binary version of PSO (Binary PSO). The results also indicated that ROUGE-N metrics were improved.

Table 1 Summary of the literature reviews on feature selection.

Reference	Dataset	Method
(Bing et al., 2013)	Fourteen datasets from UCI machine learning repository and six additional datasets	Applying Particle Swarm Optimization on feature selection increases classification performance and reduces features and computational time.
(Shang et al., 2016)	Two UCI benchmark datasets	Modify PSO to become Fitness-sum Proportionate Selection Binary Particle Swarm Optimization (FS-BPSO) for feature selection in a sentiment classification problem.
(Kristiyanti and Wahyudi 2017)	Amazon's Cosmetic Product Review	Comparing the performance (accuracy) of PSO, GA, and PCA on feature selection. These algorithms are combined with SVM.
(Suganya & Priya, 2017)	Hotel Review Dataset (collected from TripAdvisor)	Applying PSO on feature selection to the sentiment classification and text summarization problems.
(Suganya et al., 2019)	Hotel Review Dataset of cities such as Beijing and London (collected from TripAdvisor)	Applying a proposed Fitness Based BPSO feature selection to the sentiment classification and text summarization problem.
(Suryadi 2020)	Online Customer Reviews (collected from sociolla.com)	Applying Fischer Score to reduce the dimensionality of textual features

2.2 Binary particle swarm optimization

Particle Swarm Optimization, or PSO, as one of the Evolutionary computation techniques, has been used in broad optimization fields, including feature selection (Bing et al., 2013). In 1995, Particle Swarm Optimization (PSO) was first introduced by Dr. Eberhart and Dr. Kennedy. Particle Swarm Optimization (PSO) is a metaheuristic swarm-intelligence-based algorithm that simulates the social behavior of birds. Each particle is a vector in a multidimensional search space, and each particle within a swarm has its position and velocity. Each particle moves according to the current particle velocity, the best position the particle has explored, and the global best position the swarm has explored (Khanesar et al., 2017). Each particle updated its velocity using formula (1), and the position of each particle is updated using formula (2). r_1 and r_2 are random numbers.

$$v_{pd}^{new} = w \times v_{pd}^{old} + c_1 \times r_1 \times (pbest_{pd} - x_{pd}^{old}) + c_2 \times r_2 \times (gbest_d - x_{pd}^{old}) \quad (1)$$

$$x_{pd}^{new} = x_{pd}^{old} + v_{pd}^{new} \quad (2)$$

As described in the pseudocode below, the algorithm starts with initializing random particles and their position. The process stops when the process has reached the stopping criterion (Bing et al., 2013).

```
Initialize p-particle, particle position;
repeat
evaluate fitness value of each particle f(x);
update pbest of each particle and gbest;
   for (each particle p = 1,2, …, p) do
   update velocity using formula 1;
   update particles position using formula 2;
   if f(x) > pbest, then pbest = f(x);
   if pbest > gbest, then gbest = pbest;
   end for
   until ("stopping criterion is true")
```

Cherrington et al. (2019) summarized several traditional feature selection methods and reviewed the advantage of the Particle Swarm Optimization (PSO) filter based on feature selection. They also highlighted feature selection's limitations, opportunities, and improvement methods. They argued that initialization techniques could affect performance. Particle Swarm Optimization (PSO) must be tuned before performing feature selection to /improve efficiency, performance, and analysis.

The main difference between BPSO and PSO is that the position of each particle is limited to [0,1] using formula (4), and to update the particle position. The velocity is first transformed into a sigmoid value using formula (3).

$$S(v_{pd}) = \frac{1}{(1 + e^{-v_{pd}})} \quad (3)$$

$$x_{pd} = \begin{Bmatrix} 1, if\ r_2 \le S\ (v_{pd}) \\ 0, otherwise \end{Bmatrix} \quad (4)$$

As described in the pseudocode of BPSO below, the main process of the BPSO algorithm is almost identical to PSO. The main difference is that before updating the particle positions, the velocity of each particle is transformed into a sigmoid value using a sigmoid function (Xue et al., 2014).

```
Initialize p-particle, particle position;
repeat
evaluate fitness value of each particle f(x);
update pbest of each particle and gbest;
   for (each particle p = 1,2, …, p) do
   update velocity using formula 1;
   calculate the sigmoid value using formula 3;
   update particles position using formula 4;
   if f(x) > pbest, then pbest = f(x);
   if pbest > gbest, then gbest = pbest;
   end for
```

DOI: 10.6977/IJoSI.202303_7(5).0003
D. Adrianto, D. Suryadi/Int. J. Systematic Innovation, 7(5), 30-45(2023)

until ("stopping criterion is true")

2.3 Sentiment orientation-pointwise mutual information

Pointwise Mutual Information (PMI) is a measure of association between two items based on information theory. In natural language processing, Pointwise Mutual Information compares the probability of two words dependently with the probability of those two words independently. Sentiment Orientation (or Semantic Orientation) is a numerical rating that indicates the direction of the sentiment of a word (i.e., positive or negative). Sentiment Orientation-Pointwise Mutual Information (SO-PMI) is a measure of the relevance of words to a reference sentiment word (i.e., positive or negative) using the information of words' co-occurrence in a corpus (Turney 2002).

The sentiment orientation of a word is calculated by comparing its correlation to a positive reference word with its correlation to a negative reference word (in this case, the words are "yes" or "no" repurchase intention) (Turney 2002). Mathematically, a word is assigned a numerical rating by the correlation to the positive reference word and subtracted to a numerical rating by the correlation to the negative reference word (Turney 2002). If the SO-PMI value of two words' co-occurrence is high, they have a strong correlation (Zhao, Zhang and Chai 2015). SO-PMI is calculated using formulas (6) and (7).

$$PMI(x;y) = \log_2 \frac{p(x,y)}{p(x)p(y)} \tag{6}$$

$$SO\ PMI\ (x) = PMI(x, positive) - PMI(x, negative) \tag{7}$$

$PMI(x;y)$ is a PMI value of x and y. $p(x,y)$ is a value of co-occurrence numbers of x and y in the document. $The\ notation\ p(x)$ is a value of occurrence numbers of x in the document. $The\ notation\ p(y)$ is a value of occurrence numbers of y (repurchase intention "yes" or "no" in this case) in the document (Zhao, Zhang and Chai 2015).

In this paper, SO-PMI provides a meaningful encoding for particle dimensions by sorting features (words) based on their polarity towards repurchase intention reference words. If the SO-PMI value of a word is high (in other words, positive), that word has a strong correlation with "yes" repurchase intention and vice versa.

2.4 k-nearest neighbors

The k-Nearest Neighbors (k-NN) classifier is a supervised machine learning algorithm. The k-NN is proposed to classify labels (repurchase intention "yes" or "no") by ranking the training data based on the Euclidean distance from their neighbors and comparing the label with k-most similar neighbors (Mandong and Munir 2018). Mathematically, formula (5) measures the Euclidean distance within the data.

$$d_i = \sqrt{\sum_{i=1}^{p} (x_{2i} - x_{1i})^2} \tag{5}$$

Govindarajan and Chandrasekaran (Govindarajan and Chandrasekaran 2010) evaluated the k-Nearest Neighbors (k-NN) classifier. They demonstrated their approach on an existing direct marketing dataset that classifies customers based on their characteristics. Their experiments showed that the proposed k-Nearest Neighbors (k-NN) performed better in accuracy. They concluded that k-Nearest Neighbors (k-NN) is not a problem-dependent algorithm and can be used for other problems or datasets.

Generally, k-NN is the simplest machine learning algorithm compared to Naïve Bayes and Support Vector Machine (SVM). Although Naïve Bayes tends to be much faster than k-NN when it applies to big data, Naïve Bayes could suffer from the zero probability problem and would result in a biased prediction. Compared to Support Vector Machine (SVM), k-NN is better if the training data is larger than the number of features. SVM is better with a dataset with a low sample size but a very high number of inputs (features) (Bzdok et al., 2018). Since this paper aims to see how a metaheuristic algorithm (Binary Particle Swarm Optimization) works on a textual feature selection problem, particularly on the problem of predicting repurchase intention, only k-NN is selected to lower the computational time due to its simplicity. Also, k-NN is the most preferred and most used classifier among all (Agrawal et al., 2021).

Therefore, this paper proposed combining a metaheuristic algorithm and a measure of associations between textual features. Binary Particle Swarm Optimization (BPSO) as a metaheuristic algorithm is proposed to select the best features subset according to the highest accuracy of the k-NN prediction model.

3. Research methodology

This paper aims to implement Binary Particle Swarm Optimization (BPSO) combined with Sentiment Orientation-Pointwise Mutual Information (SO-PMI) to select the best features subset to construct a predictive model of repurchase intention from customer reviews. The research methodology in this paper is presented in five main steps, as shown in Fig. 1. This research methodology provides a systematic way of feature selection using a metaheuristic algorithm.

The first step is to gather the data. In this paper, the dataset is a set of customer reviews as well as the repurchase intention of each review. The dataset is gathered from sociolla.com, a beauty and cosmetic product e-commerce website in Indonesia. The product category used in this customer reviews dataset is limited only to the Moisturizer category from March 18, 2019, to August 26, 2019. The dataset is divided into two parts, i.e., Dataset 1: 2,614 reviews and Dataset 2: 6,439 reviews. These datasets are the reduced datasets from more than 120,000 reviews collected.

An imbalanced dataset is a common issue in a machine learning problem, especially in a real-world problem (such as feature selection in this research). Imbalanced datasets would impact the correlation between features, and the accuracy would be biased and inaccurate. The imbalanced issue on both datasets (Dataset 1 and Dataset 2) is addressed by using an undersampling method by randomly deleting data in the dataset from the majority class (in this case, the "no" label is the majority class) to balance it with the minority class. The result of under sampling is displayed in Table 2.

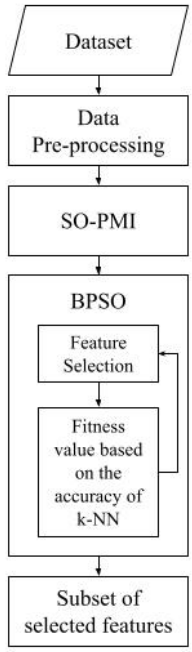

Fig. 1. Research methodology

Table 2 Treating imbalanced dataset by undersampling

Dataset		Label (Repurchase)	
		"Yes"	"No"
1	Before	611	2,003
	After	611	611
	Difference	**0**	**-1,394**
2	Before	2,539	3,900
	After	2,539	2,539
	Difference	**0**	**-1,361**

Both datasets are then randomly split using the sklearn package (Pedregosa, et al. 2011) in Python language programming into two samples, i.e., 80% training sample and 20% test sample. For dataset 1, the training sample has 978 reviews, and the rest, 244 reviews, are in the test sample. For dataset 2, the training sample has 4,062 reviews; the rest, 1,016 reviews, are in the test sample.

The cross-validation method was not used in this research, even though cross-validation would help to reduce the chance of overfitting. Cross-validation would increase training time and is computationally very expensive, as it needs a high and powerful processing system (hardware and software) (Joulani et al., 2015). In this research, there are two relatively large datasets (Dataset 1: 5,698 tokens (or words) with a total frequency of 54,981 and Dataset 2: 15,090 tokens (or words) with a total frequency of 297,422). Therefore, cross-validation was arguably not required

DOI: 10.6977/IJoSI.202303_7(5).0003
D. Adrianto, D. Suryadi/Int. J. Systematic Innovation, 7(5), 30-45(2023)

since these datasets were considered sufficiently large. Also, since the datasets are randomly split into the training set and the test set, even though cross-validation was not used, it is expected that the test set already represents the population of data. Therefore, the performance metrics (accuracy, precision, recall, and F-score) may also represent the expected performance.

Most customer reviews in the datasets are written in Indonesian and English. The examples of customer reviews collected are in Table 3. The first review in the examples written in Indonesian has a repurchase intention of "yes," and the second one is written in English with a repurchase intention of "no." Subsequently, the reviews in both datasets are transformed (during the data pre-processing step) into feature vectors. The contents of the feature vector are the frequency per feature per review. Based on Table 3, the value of $X_{N,n}$ is the frequency of the N-th feature that appears in the n-th review. For example, if the value of $X_{N,n}$ is 3, then the N-th Feature appears three times in the n-th review.

Table 3 Examples of customer reviews collected

Customer Review	Repurchase Intention (Yes/No)
"Pelayanannya bagus, harganya juga lebih murah dibandingkan sama toko sebelah" ("The service was great, also the price was cheaper than other stores")	Yes
"This product is so bad; I don't want to buy it anymore."	No

Next, the SO-PMI values of the features are calculated. Then, the features are sorted from the lowest SO-PMI value (strongly correlated to "no" repurchase intention) to the highest SO-PMI value (strongly correlated to "yes" repurchase intention). For example, "sedih" ("sad") has the lowest SO-PMI value of -5.0348, and "suka" ("like") has the highest SO-PMI value of 4.8457. As an illustration, in Table 4, the first feature is the word with the lowest SO-PMI value, and the last feature is the word with the highest SO-PMI.

Table 4 Illustration of feature vectors as the encodings of particles

Review	Feature			Repurchase Intention
	1^{st} Feature	...	N^{th} Feature	
1^{st} Review	1	...	$X_{N,1}$	Yes
...	...	...	...	...
n^{th} Review	0	...	$X_{N,n}$	No

Subsequently, to classify the repurchase intention of each review, the k-NN algorithm is applied using the selected features. The accuracy is then used on Binary Particle Swarm Optimization (BPSO) as the fitness value of particles. BPSO algorithm guides the search for the best set of features, although not guaranteed to be optimal. The k-NN algorithm was used in this research since k-NN is generally an easy and simple machine learning algorithm.

The accuracy represents the fitness of a particle after implementing the selected features into the k-NN algorithm. The accuracy is the ratio of the number of correct predictions to the total number of predictions.

4. Experiments and results

After the datasets are collected, the data are then split into tokens (tokenization), transformed into standard forms (stemming), and stop word removal is performed. Most of the reviews in the datasets are written in Indonesian and English. Both dictionaries (Indonesian and English) are used in this paper using packages such as NLTK (Bird et al., 2009) and PySastrawi (Robbani 2018). The examples of data pre-processing are shown in Table 5.

Table 5 Data pre-processing

Before	After
"Pelayanannya bagus, harganya juga lebih murah dibandingkan sama toko sebelah" ("The service was great, also the price was cheaper than the other stores")	[*'layan', 'bagus', 'harga', 'lebih', 'murah', 'banding', 'toko', 'sebelah'*]
"This product is so bad; I don't want to buy it anymore"	['product', 'bad', 'want', 'buy']

Tokens are features that will be selected. The total tokens appear in both datasets (5,698 tokens and 15,090 tokens). The frequencies of each token that appears in the documents are calculated and sorted from the highest frequency to the lowest. Features are now reduced to 341 features and 605 features using Pareto Principle (80/20). Pareto Principle (80/20) removes the features (words) that do not belong to the set of the most frequent words that constitute 80% of the total frequency of all words. The remaining features are then sorted using SO-PMI from the lowest SO-PMI value (strongly correlated to "no" repurchase intention) to the highest SO-PMI value (strongly correlated to "yes" repurchase intention). The idea of implementing SO-PMI was to sort the feature index on particles and to capture the effects of sentiment orientation of features towards the repurchase intention ("yes" or "no") on model accuracy.

Some of the results are presented in Table 6. Using the SO-PMI formula, the SO-PMI value of the word "Sedih" (equivalent to "sad" in English) is -5.0348 and is negative, so the word "Sedih" is strongly correlated to "no" repurchase intention. The word "Sedih" is equivalent to "sad" in English, and it is reasonable that the word "Sedih" is correlated with "no" repurchase intention. The word "Suka" (equivalent to "like" in English) has a SO-PMI value of 4.8457 and is positive, so it is strongly correlated with "yes" repurchase intention.

Table 6 Example of SO-PMI values

Features	SO-PMI
Sedih (Sad)	-5.0348
Lacock (Not suitable)	-4.7192
…	…
Wajib (Compulsory)	4.8457
Cantik (Beautiful)	4.8457
Suka (Like)	4.8457

The sorted features are then transformed into a specific model (particle encoding for BPSO example shown in Table 7), and accuracy-based fitness value feature selection is performed. Based on Table 7, if the value of $X_{N,n}$ feature is 1, $X_{N,n}$ feature is selected in the particle. Conversely, $X_{N,n}$ feature is not selected in the particle. The value [0,1] is a particle position for each dimension.

Table 7 Particle encoding in BPSO

Feature No.	1	2	…	N-1	N
Feature	"*Sedih*" (Sad)	"*Gacocok*" (Not suitable)	…	"*Cantik*" (Beautiful)	"*Suka*" (Like)
1st Particle	1	0	…	1	0
…	…	…	…	…	…
nth Particle	0	1	…	1	0

The experiment was performed using PyCharm 2020 with a Python 3.7 64-bit version. The sklearn (Pedregosa, et al. 2011) and PySwarms (Miranda 2018) packages are used in this experiment. The hardware used in this experiment is a laptop with Intel Core i7-7700HQ with 16 GB DDR-4 of RAM. Both datasets are split into a training set (80%) and a test set (20%). Twelve replications of the experiment are performed using Dataset 1 with an average running time of around 6 hours, and five replications using Dataset 2 with an average running time of around 12 hours. Parameters are set by conducting preliminary experiments using Dataset 1 before performing the entire experiment.

The BPSO parameters used in this study are obtained from several experiments, which are conducted using the training dataset. The parameters are selected based on the model's accuracy using the training data considered optimal. The parameter of maximum iteration is set to 55, and the number of particles is 25, which is determined based on the literature (Yassin et al., 2012). First, c_1 and c_2 are determined using five replications on a training dataset in order to see if there is a difference in accuracy between $c_1 > c_2$ and $c_1 < c_2$.

Based on Table 8, the accuracy is higher when $c_1 > c_2$ with the value of $c_1 = 1$ and $c_2 = 0.75$. This result is then tested using a two-sample t-test to see if there is a statistical difference in accuracy between $c_1 > c_2$ and $c_1 < c_2$. Using $\alpha = 0.05$, there is no statistical difference in accuracy. Then, c_1 and c_2 are determined using an identical value of 1. The selection of identical values is also supported by research on the parameter selection of PSO, which stated that if the value of c_1 and c_2 is identical, the fitness value generated by the model would be better (He et al., 2016).

Next, w is determined using the same method. $w = 1.5$ and $w = 0.75$ are tested using the same training dataset used above. The value of $w = 0.75$ does not converge at 55 iterations of BPSO. The value of $w = 1.5$ converged early. Based on the experiment above, $w = 1$ is determined to overcome premature convergence or slow convergence at 55 iterations of BPSO.

Table 8 Results of testing the value of c_1 and c_2

Run	Accuracy	
	$c_1 = 1 > c_2 = 0.75$	$c_1 = 0.75 < c_2 = 1$
1.	0.785851	0.776291
2.	0.804971	0.778203
3.	0.820868	0.791587
4.	0.743786	0.799235
5.	0.797323	0.789675

Parameters on k-Nearest Neighbors, the value of k is determined using the same method above. k = 1 to k = 10 is tested using the training dataset. The results are shown in Fig. 2. The parameter k = 8 generates the highest accuracy of 0.77214 on the training dataset. Then k = 8 is used based on the experiments above.

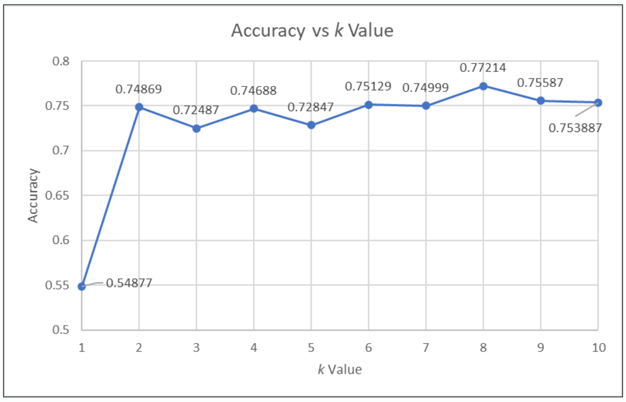

Fig. 2 Accuracy based on the value of k.

Table 9 is a summary of the final BPSO parameters used in this paper. The results are tested using a two-sample t-test to see if there is a statistical difference in accuracy before feature selection and accuracy after feature selection. Using α = 0.05, there is a statistical difference in accuracy for both datasets (Dataset 1: from 341 features reduced to 229 features; Dataset 2: from 605 features reduced to 389 features). Table 8 is a summary of the results.

Table 10 indicates that despite the number of features decreasing, the accuracy of repurchase intention predictive models constructed are increased on both datasets. For each dataset, the first row corresponds to no feature selection. The second row shows the result of selecting features according to the Pareto (80/20) rule. Finally, the last row displays the result of applying BPSO. The respective number of features and the accuracy are shown for each row. The other performance metrics

Table 9 Parameters used in this research

Parameters	Value
c_1	1
c_2	1
Max. Iteration	55
No. of Particles	25
k	8
w	1

Table 10 Model Performance (Accuracy)

Dataset	Pareto (80/20)	SO-PMI	BPSO	Feature	Accuracy
1	-	-	-	618	0.7591
	✓	-	-	341	0.7744
	✓	✓	✓	**229**	**0.8131**
2	-	-	-	2100	0.7137
	✓	-	-	605	0.7788
	✓	✓	✓	**389**	**0.7987**

Table 11 Results Summary

Dataset	Feature	Accuracy	Precision	Recall	F-Score
1	618	0.7591	0.9548	0.7996	0.8701
	229	0.8131	0.9649	0.8249	0.8892
%	-62.9%	**5.40%**	**1.01%**	**2.53%**	**1.91%**
2	2100	0.7137	0.9832	0.7881	0.8748
	389	0.7987	0.9706	0.8067	0.8806
%	-81.5%	**8.50%**	-1.26%	**1.86%**	**0.58%**

As a comparison between no feature selection and applying BPSO, Table 11 shows the performance metrics other than accuracy. Based on Table 11, there are improvements in accuracy and almost all other performance metrics. According to the result, the proposed method of BPSO on feature selection constructs a better-performing predictive model of repurchase intention from customer reviews.

Table 12 shows 20 (out of 150) selected features on both datasets. It is worth noting that in this particular research case, features are tokens, and most of the tokens are "emotional" and "preference" words within a sentence in a review rather than a physical feature of the product itself, such as packaging. These emotional terms could describe customers' personal preferences and mediate customers' repurchase intention.

Since this research is based on sociolla.com, Indonesia-based e-commerce, most of the features (tokens) are in Bahasa Indonesia, and many are in informal or standardized forms. For example, the word "Pudar" (dull) and "Mudarin" (to dull). "Mudarin" is an informal form of the word "Memudarkan" (to dull).

Table 12 Examples of Selected Features

No.	Dataset 1	Dataset 2
1	*efektif* (effective)	*maksimal* (maximum)
2	*kental* (thick)	*ngefek* (effective)
3	*diskon* (discount)	*kental* (thick)
4	*segar* (fresh)	*diskon* (discount)
5	*terjangkau* (affordable)	*segar* (fresh)
6	*halus* (gentle)	*murah* (cheap)
7	*nyaman* (mild)	soothing
8	*dingin* (cool)	*lembut* (gentle)
9	*cerah* (glowing)	*nyaman* (mild)
10	*lembab* (moisturize)	*dingin* (cool)
11	*aman* (safe)	*terang* (glowing)
12	*awet* (long lasting)	*cantik* (beautiful)
13	*wangi* (fragrant)	*lembab* (moisturizing)
14	*wajib* (compulsory)	*aman* (safe)
15	hydrating	*awet* (long lasting)
16	*suka* (like)	*baik* (good)
17	Ngilangin (Removing)	Instant (Instant)
18	Mulus (Smooth)	Bersih (Clean)
19	Pudar (Dull)	Mudarin (Dull)
20	Manfaat (Benefit)	Simple (Simple)

Based on Table 12 above, the 20 (out of 150) selected features show similarities between the two datasets. These selected features (including “emotional” terms) may provide managerial insights in marketing the products, especially regarding the aspects related to customer repurchase intention.

For example, based on dataset 1, customers tend to repurchase a moisturizer product that is affordable, hydrating, long-lasting, gentle, and mild. These emotional terms describe what customers personally want or need in their moisturizing products.

Although this paper only uses moisturizer products as the main case study, this proposed method is not bound only to moisturizer products. Different product categories could also be analyzed using this method. Different product categories might result in different "emotional" or "preference" words as a feature in the model. Nevertheless, in general, using "emotional" and "preference" words as features in the model could help practitioners and companies gather valuable information regarding their products and predict customers' behavior, especially the repurchase intention behavior. Emotional factors could mediate the influence of shopping characteristics and customers' behavior. Keeping the customer happy by satisfying their personal needs increases the customer's intention to repurchase (Pappas et al., 2014). Companies may use these emotional appeals in their marketing campaigns or objectives (Ali et al., 2020).

5. Conclusions

The result in this paper shows that the proposed Binary Particle Swarm Optimization (BPSO) succeeds in selecting features that generate predictive models with high accuracy. This paper proposes the SO-PMI to sort the features and subsequently encode the particle dimensions according to the sorted values.

The accuracy of a repurchase intention prediction model is improved by selecting only the relevant features and also reducing the computational time (k-NN training time).

This research provides a relatively new idea of a feature selection method for the repurchase intention predictive model. To the best of our knowledge, this is the first attempt to utilize BPSO on feature selection to construct a repurchase intention prediction model that uses SO-PMI to create a meaningful particle encoding, as illustrated in Table 4. Futhermore, this is also the first attempt at utilizing data from an Indonesian e-commerce website with reviews that are mainly written in Bahasa Indonesia. In 2022, Bahasa Indonesia had approximately 300 million speakers worldwide (Grehenson 2022).

There are ideas that may be realized for future work due to the limitations of this paper. The ideas that could be considered for future work are designing other encodings of the BPSO particles and comparing the performances of various encodings, applying other types of supervised learning algorithms as the classification method (e.g., Naïve Bayes, Decision Tree, or even Deep Learning method), and exploring approaches to finding the more suitable model hyperparameters to construct a better model in BPSO also for the chosen supervised learning algorithm.

References

Agrawal, Prachi, Hattan F. Abutarboush, Talari Ganesh, & Ali Wagdy Mohamed. (2009-2019)."Metaheuristic Algorithm on Feature Selection: A Survey of One Decade of Research (2009-2019)." IEEE Access, 2021: 26766 - 26791.

Ali, Amjad, Nazia Abdul Rehman, Khurram Shakir, & Ibrahim Noorani.(2020). "Impact of Emotional Advertisement on Consumer Decision Making for Ice Cream Brands in Pakistan." International Journal of Management (IJM), 2020: 320-338.

Bing, Xue, Mengjie Zhang, & Will N. Browne. (2013)."Particle Swarm Optimization for Feature Selection in Classification: A Multi-Objective Approach." IEEE Transactions on Cybernatics 43, no. 6 (2013): 1656-1671.

Bins, Jose, & Bruce A. Drapper. (2001)."Feature selection from huge feature sets." Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001. Vancouver, 2001.

Bird, S., E. Klein, & E. Loper. (2009).Natural Language Processing with Python. O'Reilly Media Inc.

Brei, Vinicius Andrade. (2020)."Machine Learning in Marketing: Overview, Learning Strategies, Applications, and Future Developments." Foundations and Trends in Marketing 14, no. 3 (2020): 173-236.

Bzdok, Danilo, Martin Krzywinski, & Naomi Altman.(2018). "Machine learning: Supervised methods, SVM and kNN." Nature Methods (Nature Publishing Group), 2018: 1-6.

Cervante, Liam, Bing Xue, Mengjie Zhang, & Lin Shang.(2012) "Binary Particle Swarm Optimisation for Feature Selection: A Filter Based Approach." IEEE World Congress on Computational Intelligence. Brisbane: IEEE, 2012.

Chandra, Grace Nadia. (2021). Indonesia to Double the Size of Current SE Asia's Digital Economy 2030. Jakarta: Jakarta Globe, 2021.

Chen, Yue-Yang.(2012). "Why Do Consumers Go Internet Shopping Again? Understanding the Antecedents of Repurchase Intention." Journal of Organizational Computing and Electronic Commerce 22, no. 1 (2012): 38-63.

Cherrington, Marianne, David Airehrour, Joan Lu, Fadi Thabtah, Qiang Xu, and Samaneh Madanian. (2019)."Particle Swarm Optimization for Feature Selection: A Review of Filter-based Classification to Identify Challenges and Opportunities." 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). Vancouver: IEEE, 2019.

Dang, Shilpa, & Peerzada Hamid Ahmad. (2014)."Text Mining: Techniques and its Application." International Journal of Engineering & Technology Innovations 1, no. 4 (2014): 22-25.

DataReportal. Digital(2021): Indonesia. February 11, 2021. https://datareportal.com/reports/digital-2021-indonesia (accessed October 3, 2021).

Du, Jiahua, Jia Rong, Sandra Michalska, Hua Wang, & Yanchun Zhang.(2019). "Feature selection for helpfulness prediction of online product reviews: An empirical study." PLoS ONE 14, no. 12 (2019).

Forman, George.(2003). "An Extensive Empirical Study of Feature Selection Metrics for Text Classification." Journal of Machine Learning Research 3 (2003): 1289-1305.

Govindarajan, M., & R. Chandrasekaran. (2010)."Evaluation of k-Nearest Neighbor classifier performance for direct marketing." Expert Systems with Applications 37, no. 1 (2010): 253-258.

Grehenson, Gusti. (2022).Universitas Gadjah Mada. May 23, 2022. https://ugm.ac.id/id/berita/22527-penutur-bahasa-indonesia-capai-300-juta-jiwa.

Gunantara, Nyoman, & I Dewa Nyoman Nurweda Putra.(2019). "The Characteristics of Metaheuristic Method in Selection of Path Pairs on Multicriteria Ad Hoc Networks." Journal of Computer Networks and Communications, 2019.

He, Y., W. J. Ma, & J. P. Zhang.(2016). "The Parameters Selection of PSO Algorithm Influencing on Performance of Fault Diagnostics." MATEC Web of Conference. Hongkong, 2016.

Hellier, Phillip K., Gus M. Geursen, Rodney A. Carr, & John A. Rickard. (2003)."Customer repurchase intention: A general structural equation model." European Journal of Marketing 37, no. 11/12 (2003): 1762-1800.

Joulani, Pooria, András György, & Csaba Szepesvári.(2015). "Fast Cross-Validation for Incremental Learning." International Joint Conference on Artificial Intelligence (IJCAI-2015). Buenos Aires, 2015.

Khanesar, Mojtaba Ahmadieh, Mohammad Teshnehlab, & Mahdi Aliyari Shoorehdeli. (2017)."A Novel Binary Particle Swarm Optimization." 15th Mediterranean Conference on Control & Automation. Athens, 2017.

Kohavi, Ron, & George H. John. (1997). "Wrapper for feature subset selection." Artificial Intelligence 97, no. 1-2 (1997): 273-324.

Kristiyanti, Dinar Ajeng, & Mochamad Wahyudi.(2017). "Feature selection based on Genetic Algorithm, particle swarm optimization and principal component analysis for opinion mining cosmetic product review." 2017 5th International Conference on Cyber and IT Service Management (CITSM). Denpasar: IEEE, 2017.

Mandong, A., & U. Munir. (2018)."Smartphone Based Activity Recognition using K-Nearest Neighbor Algorithm." International Conference on Engineering Technologies (INCENTE'18). Konya, 2018.

Mendoza, Enrico C. (2020)."A Study of Online Customers Repurchase Intention Using the 4Rs of Marketing Framework." International Review of Management and Marketing 11, no. 2 (2020): 1-10.

Menghour, Kamilia, & Labiba Souici-Meslati. (2016)."Hybrid ACO-PSO Based Approaches for Feature Selection." International Journal of Intelligent Engineering & Systems 9, no. 3 (2016): 65-79.

Miao, Jianyu, & Lingfeng Niu. (2016)."A Survey on Feature Selection." Procedia Computer Science 91 (2016): 919-926.

Miranda, L. J. V.(2018). "PySwarms: a research toolkit for Particle Swarm Optimization in Python." Journal of Open Source Software 3, no. 21 (2018).

Nedjah, Nadia, Luiza de Maced Mourelle, Janusz Kacprzk, Felipe M.G. Franca, & Alberto Ferreira de Souza. (2009). Intelligent Text Categorization and Warsaw: Springer-Verlag Berlin Heidelberg, 2009.

Negara, Siwage Dharma, & Endang Sri Soesilowati. (2021)."ISEAS - Yusof Ishak Institute." August 2, 2021. https://www.iseas.edu.sg/wp-content/uploads/2021/07/ISEAS_Perspective_2021_102.pdf (accessed October 1, 2021).

Nguyen, Lan, Thu Ha Nguyen, & Thi Khanh Phuong Tan. (2021)."An Empirical Study of Customers' Satisfaction and Repurchase Intention on Online Shopping in Vietnam." Journal of Asian Finance, Economics and Business 8, no. 1 (2021): 971-983.

Pappas, Ilias O., Panos E. Kourouthanassis, Michail N. Giannakos, & Vassilios Chrissikopoulos.(2014). "Shiny happy people buying: the role of emotions on personalized e-shopping." Electron Markets, 2014: 193-206.

Pedregosa, F., et al. "Scikit-learn: Machine Learning in Python.(2011)." Journal of Machine Learning Research 12, no. 85 (2011): 2825-2830.

Robbani, H. A. (2018)."PySastrawi 1.2.0." 2018. https://pypi.org/project/PySastrawi/ (accessed 2020).

Saeys, Yvan, Iñaki Inza, & Pedro Larrañaga.(2007). "A review of feature selection techniques in bioinformatics." Bioinformatics 23, no. 19 (2007): 2507-2517.

Shang, Lin, Zhe Zhou, & Xing Liu. (2016). "Particle swarm optimization-based feature selection in sentiment classification." Soft Computing - A Fusion of Foundations, Methodologies, and Applications 20, no. 10 (2016): 3821-3834.

Shroff, Kandarp P., & Hardik H. Maheta.(2015). "A Comparative Study of Various Feature Selection Techniques in High-Dimensional data set to Improve Classification Accuracy." 2015 International Conference on Computer Communication and Informatics (ICCCI -2015). Coimbatore, 2015.

Suhaily, Lily, & Yasintha Soelasih.(2017). "What Effects Repurchase Intention of Online Shopping." International Business Research 10, no. 12 (2017): 113-122.

Sun, Chensheng, Jiwei Hu, & Kin-Man Lam. (2011)."Feature subset selection for efficient AdaBoost training." 2011 IEEE International Conference on Multimedia and Expo. Barcelona: IEEE, 2011.

Suryadi, Dedy.(2020). "Predicting Repurchase Intention Using Textual Features of Online Customer Reviews." 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy. Sakheer: IEEE, 2020.

Trivedi, Shrawan Kumar, & Mohit Yadav. (2018). "Predicting online repurchase intentions with e-Satisfaction as mediator: a study on Gen Y." Journal of Information and Knowledge Management Systems 48, no. 3 (2018): 427-447.

Tsai, Huei-Ting, Hsin-Cheng Chang, & Ming-Tien Tsai.(2016). "Predicting repurchase intention for online clothing brands in Taiwan: quality disconfirmation, satisfaction, and corporate social responsibility." Electron Commer Research 16 (2016): 375-399.

Turney, Peter D.(2002). "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews." Proceeding of the 40th Annual Meeting of the Association for Computational Linguistics (ACL). Philadelphia, 2002.

Vashti, Hannah, & Tony Antonio. (2021)."The Role of Price Promotion and Product Quality in Influencing the Intention to Repurchase Cok-Kis Cookies." International Conference on Entrepreneurship (ICOEN). KnE Social Sciences, 2021. 441-459.

Xue, Bing, Mengjie Zhang, & Will N. Browne. (2014)."Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms." Applied Soft Computing 18, no. 1 (2014): 261-276.

Yassin, I. M., M. N. Taib, R. Adnan, M. K. M. Salleh, & M. K. Hamzah. (2012)."Effect of swarm size parameter on Binary Particle Swarm Optimization-based NARX structure selection." 2012 IEEE Symposium on Industrial Electronics and Applications. Bandung, Indonesia, 2012.

Zhao, M., T. Zhang, & J. Chai. (2015)."Based on SO-PMI Algorithm to Discriminate Sentimental Words 'Polarity in TV Programs' Subjective Evaluation." 8th International Symposium on Computational Intelligence and Design. Hangzhou, 2015.

AUTHOR BIOGRAPHIES

Dimas Adrianto, is a Master's student at the Industrial Engineering Department, Parahyangan Catholic University, Bandung, Indonesia

Dedy Suryadi, is a faculty member at the Industrial Engineering Department, Parahyangan Catholic University, Bandung, Indonesia. His current research interests include machine learning, Natural Language Processing, and metaheuristics.

INSTRUCTIONS TO AUTHORS

Submission of papers

The International Journal of Systematic Innovation is a refereed journal publishing original papers four times a year in all areas of SI. Papers for publication should be submitted online to the IJoSI website (http://www.ijosi.org) In order to preserve the anonymity of authorship, authors shall prepare two files (in MS Word format or PDF) for each submission. The first file is the electronic copy of the paper without author's (authors') name(s) and affiliation(s). The second file contains the author's (authors') name(s), affiliation(s), and email address(es) on a single page. Since the Journal is blind refereed, authors should not include any reference to themselves, their affiliations or their sponsorships in the body of the paper or on figures and computer outputs. Credits and acknowledgement can be given in the final accepted version of the paper.

Editorial policy

Submission of a paper implies that it has neither been published previously nor submitted for publication elsewhere. After the paper has been accepted, the corresponding author will be responsible for page formatting, page proof and signing off for printing on behalf of other co-authors. The corresponding author will receive one hardcopy issue in which the paper is published free of charge.

Manuscript preparation

The following points should be observed when preparing a manuscript besides being consistent in style, spelling, and the use of abbreviations. Authors are encouraged to download manuscript template from the IJoSI website, http://www.ijosi.org.

1. *Language.* Paper should be written in English except in some special issues where Chinese maybe acceptable. Each paper should contain an abstract not exceeding 200 words. In addition, three to five keywords should be provided.
2. *Manuscripts.* Paper should be typed, single-column, double-spaced, on standard white paper margins: top = 25mm, bottom = 30mm, side = 20mm. (The format of the final paper prints will have the similar format except that double-column and single space will be used.)
3. *Title and Author.* The title should be concise, informative, and it should appear on top of the first page of the paper in capital letters. Author information should not appear on the title page; it should be provided on a separate information sheet that contains the title, the author's (authors') name(s), affiliation(s), e-mail address(es).
4. *Headings.* Section headings as well as headings for subsections should start front the left-hand margin.
5. *Mathematical Expressions.* All mathematical expressions should be typed using Equation Editor of MS Word. Numbers in parenthesis shall be provided for equations or other mathematical expressions that are referred to in the paper and be aligned to the right margin of the page.
6. *Tables and Figures.* Once a paper is accepted, the corresponding author should promptly supply original copies of all drawings and/or tables. They must be clear for printing. All should come with proper numbering, titles, and descriptive captions. Figure (or table) numbering and its subsequent caption must be below the figure (or table) itself and as typed as the text.
7. *References.* Display only those references cited in the text. References should be listed and sequenced alphabetically by the surname of the first author at the end of the paper. For example:

Altshuller, G. (1998). *40 Principles: TRIZ Keys to Technical Innovation*, Technical Innovation Center.
Sheu, D. & Lee, H. (2011). A Proposed Process for Systematic Innovation, International Journal of Production Research, Vol. 49, No. 3, 2011, 847-868.

The International Journal of Systematic Innovation
Journal Order Form

Organization Or Individual Name	
Postal address for delivery	
Person to contact	Name: e-mail: Position: School/Company:
Order Information	**I would like to order ___ copy(ies) of the** *International Journal of Systematic Innovation***:** **Period Start: 1st/ 2nd half ____, Year:______(Starting 2010)** **Period End : 1st/ 2nd half ____, Year:** **Price:** **Institutions: US $150 (yearly) / NT 4,500 (In Taiwan only)** **Individuals: US $50 (yearly) / NT 1500 (In Taiwan only)** (Local postage included. International postage extra) **E-mail to**: IJoSI@systematic-innovation.org or fax: +886-3-572-3210 Air mail desired ☐ (If checked, we will quote the additional cost for your consent)
Total amount due	**US$**

Payment Methods:
1. **Credit Card (Fill up the following information and e-mail/ facsimile this form to The Journal office indicated below)**
2. **Bank transfer**
3. **Account:** The Society of Systematic Innovation
4. **Bank Name:** Mega International Commercial BANK
5. **Account No:** 020-53-144-930
6. **SWIFT Code:** ICBCTWTP020
7. **Bank code**：017-0206
8. **Bank Address:** No. 1, Xin'an Rd., East Dist., Hsinchu City 300, Taiwan (R.O.C.)

VISA / Master/ JCB/ AMERICAN Cardholder Authorization for Journal Order

Card Holder Information

Card Holder Name	(as it appears on card)		
Full Name (Last, First Middle)			
Expiration Date	/ (month / year)	Card Type	☐ VISA ☐ MASTER ☐ JCB
Card Number	☐☐☐☐-☐☐☐☐-☐☐☐☐-☐☐☐☐	Security Code	☐☐☐ 
Amount Authorized		Special Messages	
Full Address (Incl. Street, City, State, Country and Postal code)			

Please Sign your name here ________________________ (same as the signature on your card)

The Society of Systematic Innovation
6 F, #352, Sec. 2, Guanfu Rd,
Hsinchu, Taiwan, 30071, R.O.C.