

潜在トピックを網羅し差分進化アルゴリズムを用いた複数文書要約

重松遥 小林一郎

お茶の水女子大学大学院人間文化創成科学研究科理学専攻

shigematsu.haruka@is.ocha.ac.jp koba@is.ocha.ac.jp

概要 近年、大量の文書データと接する機会の増加にともない、文書要約技術の必要性が高まっている。文書要約における一般的な手法としては、冗長性を考慮してバランスの良い重要文の組合せを求めることにより要約文を生成する手法がある。最適な文の組合せを求めるには整数計画法などが用いられるが、整数計画法はNP困難に属し、要約対象とする文書集合が大きい場合には厳密解を求めるために膨大な計算時間を要するという問題が存在する。一方、厳密な解を追求せず実用的な時間で近似解を求める最適化手法として、進化的アルゴリズムの有効性が報告されている。そのような背景を踏まえて、本研究では、進化的アルゴリズムの中でも解の精度や計算時間の点で優れているとされている差分進化アルゴリズムを用いて組合せ最適化を行う要約文生成を行う。また、文書中には複数のトピックが含まれているという仮定の下に、文書内の潜在トピックを潜在的ディリクレ配分法を用いて抽出し、各トピックの内容を万遍なく含むような文の組合せを要約文として生成する。

キーワード 複数文書要約, 潜在的ディリクレ配分法, 差分進化

1 はじめに

近年、インターネットの普及により文書の電子化が進み、人々は膨大な量の文書データを簡単に手にすることができるようになった。しかし、これらを一つずつ見ていくには時間も手間もかかるため、情報を有効的に使いきれていないのが現状である。そこで、自動的に情報を取捨選択し、重要な部分だけをまとめて提示することで効率的な内容把握を支援する文書要約技術の必要性が高まっている。

文書要約には、一つの文書に対して要約を生成する単文書要約と、同じテーマに関する複数の文書に対して要約を生成する複数文書要約がある。特に複数文書要約では、文書間で共通する文や、各文書特有の文などを適切にまとめた要約を生成する必要がある。これらの要約生成を実現する一般的なアプローチとしては、文の組合せ最適化問題を解くことによって重要文抽出を行うものがある。

重要文抽出においては、どのような文を重要文とすることが課題となる。近年、文書の表層的な情報に替えて潜在的な話題(トピック)を考慮して重要文を決める手法の有効性が示されている [1, 2]。

一方、組合せ最適化手法としては、多くの研究において整数計画法などの厳密解法が適用されている。しかし厳密解法には、解の探索空間が大きいほど解を求めるコストが大きくなるという問題がある。このことから、厳密解を求めるのではなく、実時間の中で近似解を求める手法として進化的アルゴリズムの有用性が示されている [3, 4]。

これらのことを踏まえて、本研究では文書の潜在ト

ピックを考慮した、組合せ最適化手法として進化的アルゴリズムを適用した要約手法を提案する。

2 関連研究

重要文抽出を基本的な手法とする文書要約において、最適な重要文の組み合わせを求める問題として要約文生成を定式化する研究が多くなされている。これは、与えられた文書から、要約文長などの制約の下で目的関数を最大(または最小)にするような重要文の組合せを見つけることにより要約を生成する手法であり、制約や目的の与え方、最適化の手法は様々である。

文の組合せ最適化に整数計画法を用いている研究は多くある [5, 6, 7, 8]。高村ら [5, 6]、牧野ら [7] は文書の内容をより含意するような文の組み合わせを求めた。平尾ら [8] は、重要文抽出による文書要約だけでなく、文短縮を組込んだ最適化問題として文書要約問題を定式化している。しかし、整数計画法はNP困難と呼ばれるクラスに属する計算困難な厳密解法であり、問題のスケールが大きくなるほど解を求めるために膨大な計算時間を必要とする。そこで、西川ら [9] はNP困難な問題を回避するため、制約に対してラグランジュ緩和を施し、目的関数の中に制約を組み込むことで高速な準最適解の求解を実現した。

一方、厳密解を追求せず、現実的な計算時間で近似解を求める最適化手法として、進化的アルゴリズムも有効であることが報告されている。Wardlaw ら [3]、Azamathulla ら [4] は近似解法である遺伝的アルゴリズム (GA) [10] と、厳密解法である線形計画法・整数計画法を比較したところ、GAの方がコストパフォーマンスに優れているとの結果を出している。また、進化的アルゴリズムにはGAの他にも粒子群最適化 (PSO) [11] や差分進化 (DE) [12]

など多くのアルゴリズムがあり, Chandrasekar ら [13] による進化的アルゴリズムの比較実験では, GA や PSO に比べ, DE は解の精度や計算時間の点で優れた効率的なアルゴリズムであることを示している.

文書要約において, 進化的アルゴリズムを用いる研究は近年徐々に増えている. Liu ら [14] は, 複数文書要約に対して GA を適用し, DUC04 のデータを用いた実験において ROUGE-1 スコアが最高位の手法から 0.55% のみ低だけだったと報告している. Nguyen ら [15] は, 遺伝的プログラミング (GP) を用いて重要文をランキングする手法を提案しており, それにより生成された要約文は人が作成する要約文より良い結果になったことを示している. Nandhini[16] らは, 要約文書の読みやすさなどを考慮して, 要約文生成の目的に沿った文の平均長, 要約のきっかけとなる語の割合や, 名詞, 多音節語の割合などの組合せ最適に GA を用いている. また, Rasim ら [17] は, 文書を構成している文集合の平均を取ることで, 文書の内容を総括するような代表文ベクトルを生成し, このベクトルと類似して, なおかつ冗長が少ない文の組合せを DE を利用して求めた. 彼らの手法は, グラフベースの要約手法である LexRank[18] やクラスタリングベースの要約手法である Centroid[19] などの基本的な要約手法と比べ, 高い精度を実現している.

重要文の組合せ最適化においては, 文の重要度の決め方が大切である. 一般的に, 文の重要度は, その文中に含まれる単語の重要度から計算されることが多い. 単語の重要度の決め方には, 従来からの手法である *tf.idf* の指標を用いた方法に加え, 近年, 文書中に潜むトピックを考慮してトピックの観点から単語の重要度を定める手法の有効性が示されている. 文書内のトピックの抽出には, Blei ら [20] によって提案された潜在的ディリクレ配分法 (LDA: Latent Dirichlet Allocation) が多く用いられ, 文書要約の研究だけでなく, 情報検索や情報推薦など様々な応用に適用されている. 文書要約においては, Murray[21] や Arora ら [1] は, 潜在トピックに基づいて重要文を抽出するのに LDA を用いている. また, 重要文を判定するための様々な方法に LDA で抽出された潜在トピックを利用している研究があり, Gao ら [22] や北島ら [2] は, 文の類似度グラフ作成するために LDA を用いた手法を提案している. 両者とも高い精度を実現しており, 要約生成における LDA の有用性を示している.

上述したことを踏まえて, 本研究では, 重要文の判定に潜在トピックを適用し, 重要文の組合せ最適化手法に, 進化的アルゴリズムの中でも計算効率の良いとされる差分進化アルゴリズムを用いた複数文書要約手法を提案する.

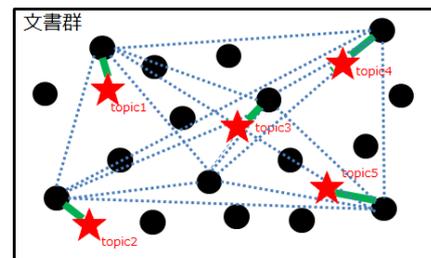
3 提案モデル

まず, 入力として N 個の文書からなる文書セット $D = \{d_1, d_2, \dots, d_N\}$ を与える. そして, 各文書に対して文分割を行い, n 個の文の集合として文書セット $D = \{s_1, s_2, \dots, s_n\}$ を捉え直す. (文書セット D は V 個の単語集合 $T = \{t_1, t_2, \dots, t_V\}$ で構成されているとする.) 最後に, 文書セット D からいくつか文を抽出して, ことで要約文を生成する.

抽出する文の組合せを表現するベクトルとして $X = [x_1, x_2, \dots, x_n]$ を考える. x_1, x_2, \dots, x_n は 1 か 0 のいずれかの二値が入り, 文 i を組合せに含めるときは $x_i = 1$, そうでないときは $x_i = 0$ となる. ここで, 要約文生成における目的と制約が最も達成されたベクトル X を見つけることを目標とする.

要約文生成における目的とは, なるべく文書の核となるような内容を多く含み, なおかつ冗長性が少ない要約を作ることである. 本研究では, 文書は k 個のトピックが集まって構成されているものと考え, トピック t に対して代表単語ベクトル $O_t = [o_{t,1}, o_{t,2}, \dots, o_{t,V}]$, ($t = 1, 2, \dots, k$) を構成し, ベクトル O_t と類似する単語分布を持つ文ほどトピック t における重要な文であると仮定して要約生成の際に考慮する. また, 要約は決められた文字数内で生成する必要があるため, 要約の長さが制約となる.

上述した説明の概要図を図 1 に示す.



★: ある話題における代表文 ●: 文書群を構成する各文

図 1 要約文生成の概要

図中, 四角で囲まれた空間が文書群空間を表し, 黒丸が文書群内の各文を表している. ここでは, 文書群内には 5 つのトピック ($k = 5$) が潜在していると仮定している. 星印で示されたものが文書内に潜在する各トピックの代表ベクトルに相当し, 各トピック代表ベクトルに近い文が, そのトピックにおける重要文と見なされる. トピックの観点から重要とされ, かつ, 選ばれる文の組合せにおいて文間の類似度が小さいものを選ぶことより, 文書群の潜在トピックを捉えた冗長性の少ない要約文を生成する.

上記を踏まえ, 以下のように最適化問題を定式化する.

適合度関数：

$$f(X) = \sum_{i=1}^n \sum_{j=i+1}^{n-1} \left(\frac{\max_{t=1,2,\dots,k} \{sim(s_{ti}, O_t)\}}{sim(s_i, s_j)} + \frac{\max_{t=1,2,\dots,k} \{sim(s_{tj}, O_t)\}}{sim(s_i, s_j)} \right) \cdot x_i \cdot x_j \rightarrow max \quad (1)$$

制約条件：

$$\sum_{i=1}^n l_i \cdot x_i \leq L \quad (2)$$

$$x_i, x_j \in \{0, 1\}; \forall i, j \quad (3)$$

s_i は文 i における単語集合 T 内の単語の有無を 1 または 0 で表現した V 次元の単語ベクトルとして $s_i = [w_{i1}, w_{i2}, \dots, w_{iV}]$ と表す。 $sim(s_i, s_j)$ は文 i と文 j のコサイン類似度を表し、これを最小にすることで、要約文の冗長性を削減する。 s_{ti} はトピック t における文 i の単語ベクトルを表し、 $s_{ti} = \sum_{k=1}^V o_{tk} w_{ik}$ によって求める。 $\max_{t=1,2,\dots,k} \{sim(s_{ti}, o_t)\}$ は文 i に一番近いトピック代表ベクトルとのコサイン類似度を表し、この類似度が大きい文の組合せほど、トピックの内容を捉えた要約となる。すなわち、適合度関数の値を最大にするようなベクトル X とは、なるべくいずれかの代表単語ベクトルと類似しており、なおかつ、選ばれた文同士の類似度は小さくなるような文の組合せを表す。

制約条件 (2) は要約の長さを制限するもので、 L は予め設定された要約長、 l_i は文 i の長さを表す。

本研究では、代表単語ベクトル O_t に対して、LDA で抽出された潜在トピックを割り当てる。そして、DE によってこれらの最適化問題の近似解を得る。

4 潜在的ディリクレ配分法

本研究では、複数文書内の潜在的トピックを確率的に求めるトピックモデルとして潜在的ディリクレ配分法 (LDA)[20] を使用する。LDA は、文書はいくつかの話題 (トピック) が混合されて作られているという仮定の下、そのトピックの確率分布を導き出す手法である。各トピック t は単語分布ベクトル ϕ_t で表され、各文書 d はトピック分布ベクトル θ_d で表される。ベクトル ϕ_t において高い確率が割り振られた単語ほど、そのトピックの特徴を表す単語となり、ベクトル θ_d によって、文書の中にどのような比率でトピックが含まれているのかを推定することができる。本研究では、ベクトル ϕ_t をトピック t の内容を代表する代表単語ベクトル O_t として捉える。

5 差分進化

差分進化 (DE)[12] は進化的アルゴリズムの一種で、個体群を用いて確率的な多点探索を行う最適化アルゴリズムである。決められた世代数の中で、目的関数を最大 (または最小) にするように個体群を進化させていくことで近似解を得ることができ、アルゴリズムの容易さ、計算速度の高速性、計算精度の高さから、最適化問題において有力な手法として注目されている。

5.1 アルゴリズム概要

Step 1. 初期化. 初期個体をランダムに N 個生成し、初期集団 $G(0) = \{P_1(0), P_2(0), \dots, P_N(0)\}$ を構成。

Step 2. 終了判定. 予め設定した最大世代数 g_{max} に達していたら終了。

Step 3. 突然変異. 各個体 P_i に対して、3 個体 P_a, P_b, P_c を、 P_i 及び互いに重複しないように選択する。そして、突然変異ベクトル Q_i を基底ベクトル P_a および差分ベクトル $P_b - P_c$ から以下のように求める。

$$Q_i = P_a + F(P_b - P_c) \quad (4)$$

ここで、 F は調整パラメータである。

Step 4. 交叉. 親ベクトル P_i と突然変異ベクトル Q_i を交叉し、子ベクトル R_i を生成する。

Step 5. 生存者選択. 親ベクトル P_i と子ベクトル R_i を比べ、良い方を次世代に残す。

Step 6. Step 2. に戻る。

step 3~step 5 における処理の概要を図 2 に示す。

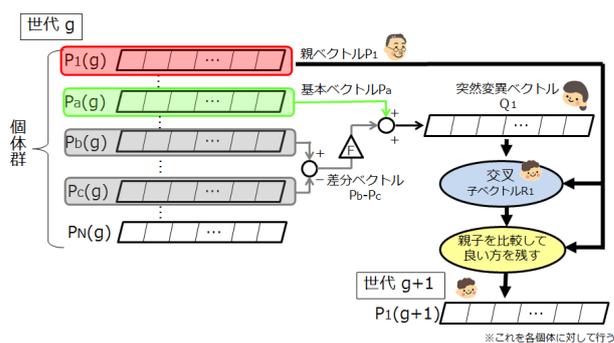


図 2 差分進化アルゴリズムにおける突然変異・交叉・選択

個体群にまとまりがないうちは差分ベクトルが大きく取られ、次世代個体群の解の幅が広がる。反対に、個体群が収束してくると差分ベクトルは小さくなるため、次世代個体群の解も収束していく。このように、突然変異ベクトルの生成に個体群の差分を用いることで、個体分布の情報が次世代個体分布に反映される。従って、DE は広域探索と局所探索のバランスが自動的に取れる収束性の高いアルゴリズムと言える。

文の組合せ最適化においては、各個体が文の組合せを表す。進化の中で組合せを試行錯誤し、式 (1) の値をより大きく取るものを次世代に残すことで、要約に適した文の組合せを求める。

以下、要約生成問題における DE の細かい説明や改良点を示す。

5.2 初期集団生成

DE では世代 $g = 0, 1, \dots, g_{max}$ の中で N 個の個体からなる集団 $G(g)$ を進化させていく。ここで、世代 g の i 番目の個体 $P_i(g)$ は以下のようにおく。

$$P_i(g) = [p_{i,1}(g), p_{i,2}(g), \dots, p_{i,n}(g)]$$

初期集団 ($g = 0$) は、予め与える必要があり、多様性に富んだ個体を用意するために、以下の式でランダムに個体の要素を求める。

$$p_{i,s}(0) = p_s^{min} + (p_s^{max} - p_s^{min}) \cdot rand_{i,s} \quad (5)$$

各要素ごとランダム値 $0 \leq rand_{i,s} \leq 1$ を求め、上限値 p_s^{max} と下限値 p_s^{min} の間の値をランダムに決める。

5.3 突然変異

突然変異ベクトル Q_i を求める一般的な式は、式 (4) であるが、解の精度を高めるため、新たな式を提案している研究が多数ある [23, 24]。本研究では Rasim ら [17] が提案した以下の式を用いる。

$$Q_i(g) = P_a(g) + F \cdot (P_{best}(g) - P_b(g)) + F \cdot (P_{best}(g) - P_c(g)) \quad (6)$$

$P_a(g), P_b(g), P_c(g)$ は、個体 $P_i(g)$ を除いた集団 $G(g)$ の中からランダムに選んだ個体である。また、 P_{best} は、集団 $G(g)$ の中で最も良い個体を表す。調整パラメータ F の値は、Storn ら [25] の研究により $[0.4, 1.0]$ の間の値が効果的だと示されている。

5.4 交叉

親ベクトル $P_i(g)$ と突然変異ベクトル $Q_i(g)$ を交叉率 $CR(g)$ で交叉させ、子ベクトル $R_i(t)$ を生成する。

$$r_{i,s}(g) = \begin{cases} q_{i,s} & (\text{if } rand_{i,s} \leq CR(g) \text{ or } s = s_{rand}) \\ p_{i,s} & (\text{otherwise}) \end{cases}$$

世代が進むにつれ集団は良いものとなってくるため、子ベクトルを生成する際は親ベクトルの要素を多く取り入れた方がよい。そこで、交叉率 $CR(g)$ を次のように求める。

$$CR(g) = CR(0) \cdot \text{sigm}(g_{max}/(2 \cdot (g + 1))) \quad (7)$$

$CR(0)$ は初期世代の交叉率であり、予め与えておく。 $\text{sigm}(z)$ はシグモイド関数である。これを導入することにより、

世代が 0 から g_{max} に近づくにつれ徐々に交叉率を減らし、親ベクトルの要素が強い子ベクトルを生成するようにする。また、 s_{rand} は $1, 2, \dots, n$ のいずれかの値がランダムに選ばれたもので、子ベクトルが親ベクトルと全く同じものにならないように、 s_{rand} 番目の要素は必ず突然変異ベクトルの要素を取るようにする。

5.5 生存者選択

3章で示した適合度関数 (1) と制約条件 (2) で親ベクトル $P_i(g)$ と子ベクトル $R_i(g)$ を評価し、次世代の生存者を選択する。しかし、現在これらのベクトルには実数値が入っているため、このままでは式 (1)(2) の値を求めることができない。そこで、実数値ベクトルを二値化する作業が必要となる。

以下のルールに基づき、実数値ベクトル P を二値ベクトル P' に変える。

$$p'_{i,s}(g) = \begin{cases} 1 & (\text{if } rand_{i,s} < \text{sigm}(p_{i,s}(g))) \\ 0 & (\text{otherwise}) \end{cases}$$

まず、実数値 $p_{i,s}(g)$ をシグモイド関数を利用して 0 ~ 1 の値に変換する。そして、それらの値がランダム値 $0 \leq rand_{i,s} \leq 1$ より大きければ 1、小さければ 0 に二値化する。

実数値ベクトルを二値ベクトルに変換し、式 (1)(2) の値を求めた後、以下の条件に従って親ベクトル $P_i(g)$ と子ベクトル $R_i(g)$ のいずれかを次世代 $P_i(g + 1)$ に選択する。

- どちらも制約条件 (2) を満たしている場合、適合度関数 (1) の値が大きい方を選択
- どちらかが制約条件 (2) を満たしていない場合、いかなる場合も制約条件を満たしている方を選択
- どちらも制約条件 (2) を満たしていない場合、制約条件を大きく違反していない方を選択

6 実験

以上の要約モデルを用いて複数文書要約を行い、提案手法を評価する。本実験では、要約評価ワークショップ DUC04 の Task2 で使用されたデータセットを用いる。データセットには、話題の異なる 50 の文書セットが用意されており、1 文書セットあたり 10 個のニュース記事から成っている。各文書セットに対して、長さ 665 バイト以内の要約を生成し、評価指標 ROUGE[26] を用いて要約の精度を測る。要約文は各文書セットあたり、それぞれ 20 回ずつ生成し、20 個の要約に対する ROUGE 値の平均を最終的な要約精度とする。

6.1 パラメータ設定

LDA では、予め文書に含まれているトピックの数を与える必要がある。本実験においては、トピック数はパー

プレキシティによって調べた。トピックの推定にはギブスサンプリングを用い、反復回数は200回とした。

DEは、最大世代 $g_{max} = 1000$ 、個体数 $N = 50$ として実験を行った。細かなパラメータの設定は、Rasimら[17]を参考にして $CR(0) = 0.7$ 、 $F = 0.45$ 、 $p_s^{min} = -10$ (全ての s に対して)、 $p_s^{max} = 10$ (全ての s に対して)、と設定した。

6.2 結果

1000世代に達した個体群の中で、最も制約と目的が達成されている最良個体を生成要約文として評価することとした。しかし、1文書セットあたり20回試行して求めた20個の要約文のうち、文字数が制約の665バイト以内に収まっているものもあれば、収まっていないものもあり、要約結果が安定していなかったため、今回は提案手法の評価をすることができなかった。

各文書セットに対して、得られた1000世代目の個体群を見てみると、最良個体は665バイト以内の要約となっているが、最悪個体に関しては665バイトを超えるものが多く、1000世代では個体群全体が収束せず、要約文字数が制約の665バイト以内に収まらないという問題が見受けられた。また、各文書あたり生成された20個の要約を比較しても、内容が類似していないものが多々あり、1000世代では近似解を得るのに不十分であった。

7 結論

本論文では、文の組合せ最適化によって内容に冗長性のない、文書の潜在トピックを捉えた要約文を生成した。組合せ最適化手法には、厳密解を求めるのではなく現実的な計算時間で近似解を求める差分進化アルゴリズムを適用した。しかし、1000世代では解が収束しきらないことが判明したことから、今後は世代数をもっと増やすことで、制約を満たし、なおかつ、複数回要約生成を繰り返しても同じような近似解が取り出せるようにする。また、突然変異の式やパラメータ F に対しても工夫を加えることで、世代数を変えることなく、効率のよい進化を実現したいと考えている。具体的には、現在は全ての個体に対して突然変異式は一意であるが、個体の特徴ごとに F の値に変化をもたせるなどの工夫をしていきたいと思う。

参考文献

- [1] Rachit Arora, Balaraman Ravindran: Latent Dirichlet Allocation Based Multi-Document Summarization, Proceedings of the second workshop on Analytics for noisy unstructured text data, pp. 91-97, 2008.
- [2] 北島理沙, 小林一郎: トピックを考慮したグラフによる複数文書要約への一考察, 第19回言語処理学会年次大会, pp. 504-507, 2013.
- [3] Robin Wardlaw, Kampanad Bhaktikul: Comparison of Genetic Algorithm and Linear Programming Approaches for Lateral Canal Scheduling, Journal of Hydro-environment Research, Journal of Irrigation and Drainage Engineering, Vol. 130, No. 4, pp.311-317, 2004.
- [4] H. Md. Azamathulla, Fu-Chun Wu, Aminuddin Ab Ghani, Sandeep M. Narulkar, Nor Azazi Zakaria, Chun Kiat Chang: Comparison between genetic algorithm and linear programming approach for real time operation, Journal of Hydro-environment Research, Vol.2, pp. 172-181, 2008.
- [5] 高村大也, 奥村学: 最大被覆問題とその変種による文書要約モデル, 人工知能学会論文誌, Vol. 23, No. 6, pp. 505-503, 2008.
- [6] 高村大也, 奥村学: 施設配置問題による文書要約のモデル化, 人工知能学会論文誌, Vol. 25, No. 1, pp. 174-182, 2010.
- [7] 牧野拓哉, 高村大也, 奥村学: アスペクト被覆を可能にした最小値最大化問題に基づく文書要約モデル, 情報処理学会研究報告, Vol2011-NL-204 No.9, 2011.
- [8] 平尾努, 鈴木潤, 磯崎秀樹, 最適化問題としての文書要約, 人工知能学会論文誌, Vol. 24, No. 2, pp. 223-231, 2009.
- [9] 西川仁, 平尾努, 牧野俊朗, 松尾義博: ラグランジュ緩和による複数文書要約の高速求解, 言語処理学会論文誌, Vol. 18, pp. 1071-1074, 2012.
- [10] Holland, J.H.: Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence, University of Michigan Press, 1975
- [11] J. Kennedy and R. C. Eberhart: Particle swarm optimization, Proceedings of IEEE International Conference on Neural Networks, Vol. 1498 of Lecture Notes in Computer Science, pp. 1942-1948, 1995.
- [12] Storn, R. and Price, K.: Minimizing the Real Functions of the ICEC96 Contest by Differential Evolution, in Proc. of the International Conference on Evolutionary Computation, pp. 842-844, 1996.
- [13] Chen, N, Vapnik, J. P: Performance Comparison of GA, DE, PSO and SA Approaches in Enhancement of Total Transfer Capability using FACTS Devices, Journal of Electrical Engineering & Technology, Vol. 7, No. 4, pp. 493-500, 2012.
- [14] Dexi Liu, Yanxiang He, Donghong Ji, and Hua Yang, Genetic Algorithm Based Multi-document Summarization, Q. Yang and G. Webb (Eds.): PRICAI 2006, LNAI 4099, pp. 1140 - 1144, 2006.
- [15] Nguyen, Q U. Pham, T A. Truong, C D. Nguyen, X H., A Study on the Use of Genetic Programming for Automatic Text Summarization, 2012 Fourth International Conference on Knowledge and Systems Engineering (KSE), 2012.
- [16] K. Nandhini and S. R. Balasundaram, Use of Genetic Algorithm for Cohesive Summary Extraction to Assist Reading Difficulties, Applied Computational Intelligence and Soft Computing Volume 2013 Article ID 945623, 11 pages, 2013.
- [17] Rasim M. Alguliev, Ramiz M. Aliguliyev, Chingiz A. Mehdiyev: Sentence selection for generic document summarization using an adaptive differential evolution algorithm, Swarm and Evolutionary Computation 1(4), pp. 213-222, 2011.
- [18] G. Erkan, D. Radev: LexRank: graph-based centrality as salience in text summarization, Journal of Artificial Intelligence Research, pp. 457-479, 2004.

- [19] D. Radev, H. Jing, M. Stys, D. Tam: Centroid-based summarization of multiple documents, *Information Processing & Management*, pp. 919-938, 2004.
- [20] David M. Blei, Andrew Y. Ng, Michael I. Jordan: Latent Dirichlet Allocation, *Journal of Machine Learning Research* 3, pp. 993-1022, 2003.
- [21] Kenton W. Murray, Summarization by Latent Dirichlet Allocation: Superior Sentence Extraction through Topic Modeling, A senior thesis for Bachelors degree, Princeton University, 2009.
- [22] Dehong Gao, Wenjie Li, You Ouyang, Renxian Zhang, LDA-Based Topic Formation and Topic-Sentence Reinforcement for Graph-Based Multi-document Summarization, *Lecture Notes in Computer Science Volume 7675*, pp 376-385, 2012.
- [23] A.K. Qin, V.L. Huang, P.N. Suganthan: Differential evolution algorithm with strategy adaptation for global numerical optimization, *IEEE Transactions on Evolutionary Computation*, pp. 398-417, 2009.
- [24] A. Iorio, X. Li: Solving rotated multi-objective optimization problems using differential evolution, *Proceedings of the Australian Conference on Artificial Intelligence*, pp. 861-872, 2004.
- [25] R. Storn, K. Price: Differential evolution - a simple and efficient heuristic for global optimization over continuous space *Journal of Global Optimization*, 11 (4), pp. 341-359, 1997.
- [26] Lin, C.-Y: ROUGE: A Package for Automatic Evaluation of Summaries, in *Proc. of Workshop on Text Summarization Branches Out, Post Conference Workshop of ACL 2004*, pp. 74-81, 2004.