



ISSN 1823-626X

Malaysian Journal of Fundamental and Applied Sciences

available online at <http://mjfas.ibnusina.utm.my>



Parikh Matrices and Istrail Morphism

K.G. Subramanian*, Pradeep Isawasan And Ibrahim Venkat

School of Computer Sciences, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia

Received 26 November 2012, Revised 10 February 2013, Accepted 15 February 2013, Available online 20 February 2013.

ABSTRACT

A word w is a sequence of symbols. A scattered subword or simply a subword u of the word w is a subsequence of w . Parikh matrix $M(w)$ is an ingenious tool introduced by Mateescu et al (2001) to count certain subwords in a word w . Various properties of Parikh matrices have been established. Two words u and v are said to be M -ambiguous or amiable if their Parikh matrices $M(u)$ and $M(v)$ are the same. On the other hand a morphism f is a mapping on words w whose images $f(w)$ are also words with the property that, $f(uv) = f(u)f(v)$ for given words u and v . Istrailmorphism (Istrail, 1977) is a specific kind of morphism on a set $\{a, b, c\}$ of three symbols. Using this morphism, M -ambiguity or amiability of words based on Parikh matrices is investigated by Atanasiu (2010). Parikh matrices of words that involve certain ratio-property are investigated by Subramanian et al (2009). Here we consider this kind of ratio-property in the context of Istrailmorphism and obtain certain properties of morphic images of words under Istrailmorphism. Using these properties, conditions are obtained for product of Parikh matrices of such morphic images under Istrailmorphism to commute.

| Combinatorics | Word | Subword | Scattered Subword | Parikh Matrix | Morphism |

© 2013 IbnuSina Institute. All rights reserved.
<http://dx.doi.org/10.11113/mjfas.v9n1.72>

1. INTRODUCTION

Combinatorics on words is a growing area of discrete mathematics with applications in many different fields. An excellent source for the study of combinatorial problems on words and their applications is the reference [1]. The topic of combinatorics on words deals with general properties of words. A word is a finite or an infinite sequence of symbols of a finite set called an alphabet. For example, *abbabba* is a word over the alphabet $\{a, b\}$, which has an additional property of being a palindrome. Investigation of properties of words has given rise to challenging problems.

Studies on numerical properties of words have proved to be useful. Parikh mapping or Parikh vector [2] is one such numerical property, which counts the number of occurrences of each symbol in a word over an alphabet. For example, the Parikh vector of the word *abbabba* is (3, 4), since it has three a 's and four b 's. This has been an important notion in formal language theory [3, 4], which is a branch of theoretical computer science.

Recently, Parikh mapping has been extended in [5] by introducing the concept of a Parikh matrix mapping or Parikh matrix, which gives more numerical information than a Parikh vector does. In fact Parikh matrix $M(w)$ counts certain subwords in a word w , also called scattered subwords. Since the introduction of the notion of Parikh matrix, intensive investigations have been done on several problems related to subwords [6-13]. Among various properties of Parikh matrices that have been established, M -ambiguity or amiability of words has been extensively

investigated. Two words u and v are said to be M -ambiguous or amiable if their Parikh matrices are the same, that is $M(u) = M(v)$.

On the other hand a morphism f is a mapping on words w whose images $f(w)$ are also words with the property that $f(uv) = f(u)f(v)$, for given words u and v . Istrailmorphism [14] is a specific kind of morphism ϕ on a set $\{a, b, c\}$ of three symbols. M -ambiguity or amiability of words based on Parikh matrices is recently investigated by Atanasiu [13] using this morphism.

Parikh matrices of words that involve certain ratio-property are investigated by Subramanian et al [9]. This property has been used in [12] to obtain conditions for equality of Parikh matrices of products xy and yx of two words x and y in the case of binary and three-symbol alphabets.

Here we consider morphic images of words under Istrailmorphism and obtain properties of these morphic images in the context of Istrailmorphism, using this kind of ratio-property. In particular certain conditions for commutativity of product of Parikh matrices of such morphic images under Istrailmorphism are obtained.

2. PRELIMINARIES

An alphabet Σ is a finite set of symbols. A finite sequence of symbols from Σ is called a finite word or simply a word over Σ . The set of all words over Σ is denoted by Σ^* . The empty word with no symbols is denoted by λ .

*Corresponding author. E-mail: kgsmani1948@gmail.com
 (K.G. Subramanian) Tel: (60)-4-6534641

The length of a word $w \in \Sigma^*$ is the number of symbols (counting repetitions) in the word w and is denoted by $|w|$.

A word u is called a subword (also called scattered subword) of a word w if there exist words x_1, \dots, x_n and y_0, \dots, y_n , such that $u = x_1 \dots x_n$ and $w = y_0 x_1 y_1 \dots x_n y_n$. For example if $w = babaabbaa$ is a word over the alphabet $\{a, b\}$, the symbols in the positions 2, 5, 6, 7, 9, namely, $aabba$ is a subword of w . The number of occurrences of the word u as a subword of the word w is denoted by $|w|_u$.

An ordered alphabet $\Sigma = \{a_1 < a_2 < \dots < a_n\}$ is an alphabet $\Sigma = \{a_1, a_2, \dots, a_n\}$ with an ordering $a_1 < a_2 < \dots < a_n$.

The Parikh vector [2] is a mapping Ψ from Σ^* to $N \times N$ where $\Sigma = \{a_1, a_2, \dots, a_n\}$ is an alphabet and N is the set of natural numbers including zero, such that for a word w in Σ^* , $\Psi(w) = (|w|_{a_1}, \dots, |w|_{a_n})$. For example, for the word $w = ababab$ over $\{a, b\}$, the Parikh vector is (3,3).

The notion of a Parikh matrix introduced in [5], is an extension of the notion of Parikh vector and gives more numerical information than a Parikh vector does. In fact Parikh matrix counts the number of occurrences of certain subwords in a given word. In the rest of the paper, we are concerned mainly with words over a binary ordered alphabet $\{a < b\}$ or a three-symbol ordered alphabet $\{a < b < c\}$.

Let M_n denote the set of all $n \times n$ upper triangular matrices M such that M has no negative integer entry, has 1's in the main diagonal and has 0's for all entries below the main diagonal.

Let $\Sigma = \{a < b\}$. The Parikh matrix mapping Ψ_2 is a mapping from Σ^* to M_3 , given by

$$\Psi_2(a) = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \Psi_2(b) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

so that

$$\Psi_2(w) = \begin{pmatrix} 1 & |w|_a & |w|_{ab} \\ 0 & 1 & |w|_b \\ 0 & 0 & 1 \end{pmatrix}, w \in \Sigma^*$$

is obtained by substituting for each symbol a or b in w , the corresponding matrix $\Psi_2(a)$ or $\Psi_2(b)$ and by performing matrix multiplication.

The matrix $\Psi_2(w)$ is called the Parikh matrix of the word w . Note that $|w|_a, |w|_b, |w|_{ab}$ give the number of a 's, the number of b 's and the number of subword ab in w .

Note that the notion of Parikh matrix mapping can be defined for words over a general ordered alphabet $\Sigma = \{a_1 < a_2 < \dots < a_n\}$. But we have recalled here the definition in the case of a binary alphabet. For a three-symbol alphabet $\Sigma = \{a < b < c\}$, the Parikh matrix mapping Ψ_3 is a mapping from Σ^* to M_4 , given by

$$\Psi_3(a) = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \Psi_3(b) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$\Psi_3(c) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

As before, for a word $w \in \Sigma^*$, $\Psi_3(w)$ is obtained by substituting for each symbol a, b or c in w , the corresponding matrix $\Psi_3(a), \Psi_3(b)$ or $\Psi_3(c)$ and by performing matrix multiplication.

We illustrate with examples.

Example 1

If $w = ababab$ over $\{a < b\}$, then

$$\begin{aligned} \Psi_2(w) &= \Psi_2(ababab) \\ &= \Psi_2(a)\Psi_2(b)\Psi_2(a)\Psi_2(b)\Psi_2(a)\Psi_2(b) \\ &= \begin{pmatrix} 1 & 3 & 6 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix}. \end{aligned}$$

Note that the word $w = ababab$ has three a 's, three b 's and six subword ab 's. The Parikh vector (3,3) appears at the second diagonal above the main diagonal in the Parikh matrix.

Example 2

If $w = aabcabc$ over $\{a < b < c\}$, then

$$\begin{aligned} \Psi_3(w) &= \Psi_3(aabcabc) \\ &= \Psi_3(a)\Psi_3(a)\Psi_3(b)\Psi_3(c)\Psi_3(a)\Psi_3(b)\Psi_3(c) \\ &= \begin{pmatrix} 1 & 3 & 5 & 7 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \end{aligned}$$

In the word w , there are 3 a 's, 2 b 's, 2 c 's, 5 ab 's, 3 bc 's, 7 abc 's. The entries in the Parikh matrix of w , in the second diagonal (above the main diagonal) give the number of a 's, b 's, c 's, in w ; the entries in the diagonal above it give the number of subwords ab 's, bc 's in w ; and the rightmost entry in the first row gives the number of subword abc 's in w .

The Parikh matrix is known [5, 7] to be not injective which means that two or more words can have the same Parikh matrix. For example the following five words over the alphabet $\Sigma = \{a < b\}$

$$baabaab, baaabba, abbaaab, abababa, aabbbaa$$

have the same Parikh matrix

$$\begin{pmatrix} 1 & 4 & 6 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix}.$$

Words having the same Parikh matrix are called M-equivalent or amiable.

A word $w \in \{a, b\}^*$ with $a < b$ is called M-unambiguous[5, 6] if there is no word $w' \in \{a, b\}^*$ so that $\Psi_2(w) = \Psi_2(w')$. Otherwise w is called M-ambiguous[5, 6]. For example the word $baabaab$ is M-ambiguous. We can analogously define M-ambiguity and M-unambiguity for words over a three-symbol alphabet $\{a < b < c\}$.

M-ambiguity of words has been extensively investigated. Characterizations of M-ambiguous binary words have been established. We state here one such characterization[8].

Two binary words u, v over the alphabet $\{a < b\}$ have the same Parikh matrix and hence M-ambiguous if and only if v can be obtained from u by interchanging an occurrence of ab in u with a distinct occurrence of ba in u . It is known that such a characterization does not immediately carry over to words over a three-symbol alphabet.

We also denote the Parikh matrix of a word w by $M(w)$.

While investigating the notion of M-ambiguity of words over a three-symbol ordered alphabet $\{a < b < c\}$, a property, called weak-ratio property was introduced in[9]. Two words u, v over $\Sigma = \{a < b < c\}$ are said to satisfy the weak-ratio property, written $u \sim_{wr} v$ if

$$\frac{|u|_a}{|v|_a} = \frac{|u|_b}{|v|_b} = \frac{|u|_c}{|v|_c}.$$

Weak-ratio property of binary words u, v over $\{a < b\}$ can be similarly defined by requiring

$$\frac{|u|_a}{|v|_a} = \frac{|u|_b}{|v|_b}.$$

It is shown in [9] that weak-ratio property of words u, v is a sufficient condition for equality of the Parikh matrices of the product binary words uv, vu or in other words, $M(uv) = M(vu)$.

Recently, Atanasiu[13] studied M-ambiguity of words that are morphic images of binary words under a special kind of morphism known as Istrailmorphism[14]. If A and B are two finite nonempty alphabets, a morphism is a mapping $f: A^* \rightarrow B^*$ such that $f(uv) = f(u)f(v)$ for all $u, v \in A^*$.

Let $\Sigma = \{a, b, c\}$. The IstrailMorphism is a mapping $\varphi: \Sigma^* \rightarrow \Sigma^*$ defined by

$$\varphi(a) = abc, \quad \varphi(b) = ac, \quad \varphi(c) = b.$$

3. MORPHIC IMAGES UNDER ISTRAIL MORPHISM

Here we consider the weak-ratio property [9] of words and obtain first some properties of the morphic images under Istrailmorphism φ , restricted to words w over a binary alphabet $\{a < b\}$. Note that the words $\varphi(w)$ are over a three-symbol alphabet $\{a < b < c\}$.

Theorem 1

Let $\Sigma = \{a < b\}$. Then we have the following properties:

i) For any $w \in \Sigma^*$,

$$|\varphi(w)|_a = |w| = |\varphi(w)|_c$$

and

$$|\varphi(w)|_c = |w|_a.$$

ii) If $u, v \in \Sigma^*$ and $u \sim_{wr} v$, then $(u) \sim_{wr} \varphi(v)$.

Proof:

Let $w \in \Sigma^*$. For $w = a_1 \dots a_n$, $a_i \in \Sigma$, for $1 \leq i \leq n$,

$$\varphi(w) = \varphi(a_1)\varphi(a_2) \dots \varphi(a_n)$$

where

$$\varphi(a_i) = \begin{cases} abc, & \text{if } a_i = a \\ ac, & \text{if } a_i = b \end{cases}$$

Hence $\varphi(w)$ has as many a 's as w has a 's and b 's. i.e.

$$|\varphi(w)|_a = |w|_a + |w|_b = |w|$$

Likewise,

$$|\varphi(w)|_c = |w|.$$

But $\varphi(w)$ has as many b 's as w has a 's. i.e.

$$|\varphi(w)|_b = |w|_a.$$

This proves property (i).

Let $u, v \in \Sigma^*$ and $u \sim_{wr} v$, so that

$$\frac{|u|_a}{|v|_a} = \frac{|u|_b}{|v|_b} = \frac{|u|_a + |u|_b}{|v|_a + |v|_b} = \frac{|u|}{|v|} = \alpha$$

where $\alpha \neq 0$ is a constant.

From property (i), we have

$$\frac{|\varphi(u)|_a}{|\varphi(v)|_a} = \frac{|\varphi(u)|_c}{|\varphi(v)|_c} = \frac{|u|}{|v|} = \alpha$$

and,

$$\frac{|\varphi(u)|_b}{|\varphi(v)|_b} = \frac{|u|_a}{|v|_a} = \alpha$$

Hence,

$$\frac{|\varphi(u)|_a}{|\varphi(v)|_a} = \frac{|\varphi(u)|_b}{|\varphi(v)|_b} = \frac{|\varphi(u)|_c}{|\varphi(v)|_c}$$

so that $\varphi(u) \sim_{wr} \varphi(v)$, proving (ii) ■

Theorem 2

Let $\Sigma = \{a < b\}$. For $u \in \Sigma^*$,

$$i) |\varphi(u)|_{ab} = \frac{1}{2}|u|_a(|u|_a + 1) + |u|_{ba}$$

$$ii) |\varphi(u)|_{bc} = \frac{1}{2}|u|_a(|u|_a + 1) + |u|_{ab}$$

Proof:

We have

$$\begin{aligned} |\varphi(u)|_{ab} &= |u|_a + (|u|_a - 1) + \dots + 1 + |u|_{ba} \\ &= \frac{1}{2}|u|_a(|u|_a + 1) + |u|_{ba}. \end{aligned}$$

Note that the application of φ on u yields the factor ab for every a in u and so the contribution coming from these factors to the number of subword ab in $\varphi(u)$ is $|u|_a + (|u|_a - 1) + \dots + 1$ whereas the application of φ on u yields the factor ac for every b in u while removing the b and so the contribution coming from this to the number of subword ab in $\varphi(u)$ is the same as the number of subword ba in u . This proves i).

Likewise we have

$$\begin{aligned} |\varphi(u)|_{bc} &= |u|_a + (|u|_a - 1) + \dots + 1 + |u|_{ab} \\ &= \frac{1}{2}|u|_a(|u|_a + 1) + |u|_{ab}. \end{aligned}$$

This proves (ii). ■

Parikh matrices of products uv and vu of words u, v and their commutativity have been studied in [12]. We state a result in [12] on the commutativity of such products in the following Lemma, which we need to prove a result in the subsequent theorem on Parikh matrices of products of morphic images of two words under the Istrailmorphism.

Lemma 1

Let $\Sigma = \{a < b < c\}$ and let u, v be two words over Σ satisfying

- i) the weak-ratio property, namely, $u \sim_{wr} v$ and
- ii) $|u|_a|u|_{bc} = |u|_{ab}|u|_c$ and $|v|_a|v|_{bc} = |v|_{ab}|v|_c$.

Then the words uv and vu have the same Parikh matrices i.e. $M(uv) = M(vu)$.

Theorem 3

Let $\Sigma = \{a < b\}$ and $x, y \in \Sigma^*$ such that $x \sim_{wr} y$ and

$$|x|_{ab} = \frac{1}{2}|x|_a \cdot |x|_b, \quad |y|_{ab} = \frac{1}{2}|y|_a \cdot |y|_b.$$

Then $\varphi(xy)$ is M-equivalent to $\varphi(yx)$ or

$$M(\varphi(xy)) = M(\varphi(yx)).$$

Proof:

Let $|x|_{ab} = \frac{1}{2}|x|_a|x|_b$. Then we have

$$|x|_{ab} + |x|_{ba} = |x|_a|x|_b,$$

so that

$$|x|_{ba} = \frac{1}{2}|x|_a|x|_b.$$

From Theorem 2, we have

$$\begin{aligned} |\varphi(x)|_{ab} &= \frac{1}{2}|x|_a(|x|_a + 1) + \left(\frac{1}{2}|x|_a|x|_b\right) \\ &= \frac{1}{2}|x|_a(|x| + 1). \end{aligned}$$

Similarly,

$$|\varphi(x)|_{bc} = \frac{1}{2}|x|_a(|x| + 1).$$

Thus

$$|\varphi(x)|_{ab} = |\varphi(x)|_{bc},$$

and likewise

$$|\varphi(y)|_{ab} = |\varphi(y)|_{bc}$$

Hence

$$\begin{aligned} |\varphi(x)|_a \cdot |\varphi(x)|_{bc} &= |x| \cdot \frac{1}{2}|x|_a(|x| + 1) \\ |\varphi(x)|_{ab} \cdot |\varphi(x)|_c &= \frac{1}{2}|x|_a(|x| + 1) \cdot |x| \end{aligned}$$

so that

$$|\varphi(x)|_a \cdot |\varphi(x)|_{bc} = |\varphi(x)|_{ab} \cdot |\varphi(x)|_c.$$

Similarly

$$|\varphi(y)|_a \cdot |\varphi(y)|_{bc} = |\varphi(y)|_{ab} \cdot |\varphi(y)|_c.$$

Also from Theorem 1, $\varphi(x) \sim_{wr} \varphi(y)$ since $x \sim_{wr} y$ and using Lemma 1, we have

$$M(\varphi(xy)) = M(\varphi(yx)).$$

So this proves Theorem 3. ■

We illustrate Theorem 3 with an example.

Example3

Let $\Sigma = \{a < b\}$ and let

$$x = aabbbbaa, y = baaabbbbaaabab$$

Then

$$\frac{|x|_a}{|y|_a} = \frac{|x|_b}{|y|_b} = \frac{1}{2}$$

and

$$|x|_{ab} = 6 = \frac{1}{2}|x|_a \cdot |x|_b, |y|_{ab} = 24 = \frac{1}{2}|y|_a \cdot |y|_b.$$

Thus the words x, y satisfy the hypothesis of Theorem 3. Now

$$\varphi(x) = abcabcacacacabcabc$$

and

$$\varphi(y) = acabcabcabcacacabcabcabcabcacabcac$$

It can be verified that $\varphi(xy) = \varphi(x)\varphi(y)$ and $\varphi(yx) = \varphi(y)\varphi(x)$ have the same Parikh matrix

$$\begin{pmatrix} 1 & 21 & 132 & 1020 \\ 0 & 1 & 12 & 132 \\ 0 & 0 & 1 & 21 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

4. CONCLUSION

In this paper, we have obtained in Theorem 3, a sufficient condition for the Parikh matrices of morphic images under Istrailmorphism of two binary words to commute. The problem of necessity of this sufficient condition remains to be investigated. Other properties of Parikh matrices can also be investigated in the context of Istrailmorphism.

ACKNOWLEDGEMENT

The authors would like to thank the reviewers for their useful comments. The authors gratefully acknowledge support for this research from FRGS Grant No. 203/PKOMP/6711267 and ERGS Grant No.203/PKOMP/6730075 of the Ministry of Higher Education, Malaysia. The second author Pradeep Isawasan would like to thank the Ministry of Higher Education, Malaysia for the award of MyPhD under which this research was jointly carried out by him.

REFERENCES

- [1] M. Lothaire, Combinatorics on words, in: Encyclopedia of Mathematics and its Applications, vol. 17, Addison-Wesley, 1983.
- [2] R. J. Parikh, J. Assoc. Comput. Machinery, 13(4) (1966) 570-581.
- [3] G. Rozenberg and A. Salomaa, Eds., Handbook of Formal Languages, Vols. 1-3, New York: Springer-Verlag New York, Inc., 1997.
- [4] A. Salomaa, Formal Languages, Academic Press, 1973.
- [5] A. Mateescu, A. Salomaa, K. Salomaa and S. Yu, RAIRO-Theor. Inform. and Appl., 35(6) (2001) 551-564.
- [6] A. Atanasiu, Int. J. Found. Comput. Sci., 18(2) (2007) 387-400.
- [7] A. Atanasiu, C. Martin-vide and A. Mateescu, Fundam. Inform., 49(4) (2002) 289-299.
- [8] S. Fosse and G. Richomme, Inf. Process. Lett., 92(2) (2004) 77-82.
- [9] K. G. Subramanian, A. M. Huey and A. K. Nagar, Int. J. Found. Comp. Sci., 20(2) (2009) 211-219.
- [10] A. Salomaa, Theor. Comp. Sci., 411 (2010) 1818-1827.
- [11] A. Atanasiu, R. Atanasiu and I. Petre, Theor. Comput. Sci., 390(1) (2008) 102-109.
- [12] K. Mahalingam and K. G. Subramanian, Int. J. Found. Comput. Sci., 23 (2012) 207-223.
- [13] A. Atanasiu, Parikh Matrices, Int. J. Found. Comput. Sci., 21 (2010) 1021-1033.
- [14] S. Istrail, Bulletin Mathematique, 21(3-4) (1977) 301-308.