



Ramkhamhaeng Research Journal of Sciences and Technology

ปีที่ 24 ฉบับที่ 1 มกราคม - มิถุนายน 2564 Vol. 24 No. 1 January - June 2021

•เครื่องจักรเชิงนำจะเป็นสำหรับเรียนรู้ส่วนเพิ่มเพื่อการจำแนกลำดับสัญลักษณ์

จิตรกร พลุโพธิ์ทอง 1

•Study on Carcinogenicity Conditions of Benzene in 25 % Pasteurized Orange Juice Contain Benzoic and Ascorbic Acid as an Ingredient

Chulaporn Lertborwornwong 14

•Investigation for the Antibacterial Activity of Five Citrus Fruits' Peel Extracts

Tuangrat Tunvongvinis and Wimon Chanchaem 27

Ramkhamhaeng Research Journal of Sciences and Technology

Ramkhamhaeng University

Vol. 24 No. 1 January - June 2021

สถาบันวิจัยและพัฒนา
อาคารสุโขทัย ชั้น 12 มหาวิทยาลัยรามคำแหง
หัวหมาก บางกะปิ กรุงเทพมหานคร 10240
โทรศัพท์ 0-2310-8119 โทรสาร 0-2310-8696

วารสารวิจัยรามคำแหง
วิทยาศาสตร์และเทคโนโลยี

สพ.21XXX

<http://www.rd.ru.ac.th>

วารสารวิจัยรามคำแหง

ฉบับวิทยาศาสตร์และเทคโนโลยี



ผู้จัดพิมพ์

มหาวิทยาลัยรามคำแหง

ที่ปรึกษาบรรณาธิการ ผู้ช่วยศาสตราจารย์ วุฒิศักดิ์ ลากเจริญทรัพย์
ผู้ช่วยศาสตราจารย์ ดร.บุญชาล ทองประยูร

อธิการบดีมหาวิทยาลัยรามคำแหง
รองอธิการบดีฝ่ายวิชาการและวิจัย

บรรณาธิการ รองศาสตราจารย์ ดร.มณี อัครวานนท์

ผู้อำนวยการสถาบันวิจัยและพัฒนา

กองบรรณาธิการ
สาขาวิทยาศาสตร์
และเทคโนโลยี

ศาสตราจารย์ ดร.สนิท อักษรแก้ว
ศาสตราจารย์ ดร.ประสาธ สืบคำ
ศาสตราจารย์ ดร.กฤษณะ สาคริก
ศาสตราจารย์ ดร.อภิชาติ สุขสำราญ
ศาสตราจารย์ ดร.งามผ่อง คงคาทิพย์
รองศาสตราจารย์ ดร.สุทัศน์ สุบินประเสริฐ
รองศาสตราจารย์ ดร.อำนาจ มีเวที
รองศาสตราจารย์ ดร.ชัยณรงค์ คันธพนิต
รองศาสตราจารย์ ดร.มาลี ณ นคร
รองศาสตราจารย์ ดร.เจริญ นิตธรรมมยง
รองศาสตราจารย์ ดร.วรรณมา ตั้งเจริญชัย
รองศาสตราจารย์ ดร.สมนึก บุญเกิด
รองศาสตราจารย์ ดร.ไทยถาวร เลิศวิทยาประสิทธิ์
ผู้ช่วยศาสตราจารย์ ดร.ธรรมศักดิ์ ยี่มิน
ผู้ช่วยศาสตราจารย์ ดร.วิมล จันทรแจ่ม
ผู้ช่วยศาสตราจารย์ ดร.สุนันท์ ไรจนกิจ
ผู้ช่วยศาสตราจารย์ ดร.ปิติกานต์ กร้ามาตร
ผู้ช่วยศาสตราจารย์ ดร.วรรณท กงสง
ศาสตราจารย์ ดร.บุญทัน ดอกไธสง
ศาสตราจารย์ ดร.สุภาวค์ จันทวานิช
รองศาสตราจารย์ ดร.พงษ์ศักดิ์ พันธธาก
รองศาสตราจารย์ ดร.เสาวนีย์ ลิขณวัฒน์
รองศาสตราจารย์ ดร.ชูชีพ พิพัฒน์คีติ
รองศาสตราจารย์ ดร.ดุขฎิ โยเหลา
รองศาสตราจารย์ ดร.พรสุข หุ่นรินทร์
รองศาสตราจารย์ บุญผา บุญทิพย์
รองศาสตราจารย์ อนุกุล พลศิริ
รองศาสตราจารย์ เอมอร ดิสปัญญา
ผู้ช่วยศาสตราจารย์ ดร.บงกชรัตน์ เตชะไตรศักดิ์
ผู้ช่วยศาสตราจารย์ ดร.รวิจวน เบญจศิริ
ผู้ช่วยศาสตราจารย์ ดร.ทัศนัยวรรณ ภู่อารีย์
อาจารย์ ดร.กรวิภา สรรพกิจจำนง
อาจารย์ ดร.สยามพร โยธาสุมุท

สาขามนุษยศาสตร์
และสังคมศาสตร์

สถาบันสิ่งแวดล้อมไทย
มหาวิทยาลัยเทคโนโลยีสุรนารี
มหาวิทยาลัยเทคโนโลยีสุรนารี
มหาวิทยาลัยรามคำแหง
มหาวิทยาลัยเกษตรศาสตร์
มหาวิทยาลัยรามคำแหง
มหาวิทยาลัยเชียงใหม่
สถาบันคลังสมองแห่งชาติ
มหาวิทยาลัยเกษตรศาสตร์
จุฬาลงกรณ์มหาวิทยาลัย
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
มหาวิทยาลัยรามคำแหง
จุฬาลงกรณ์มหาวิทยาลัย
มหาวิทยาลัยรามคำแหง
มหาวิทยาลัยรามคำแหง
มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
มหาวิทยาลัยรามคำแหง
สถาบันบัณฑิตพัฒนบริหารศาสตร์
จุฬาลงกรณ์มหาวิทยาลัย
มหาวิทยาลัยรามคำแหง
มหาวิทยาลัยธรรมศาสตร์
มหาวิทยาลัยเกษตรศาสตร์
มหาวิทยาลัยศรีนครินทรวิโรฒ
มหาวิทยาลัยศรีนครินทรวิโรฒ
มหาวิทยาลัยรามคำแหง
มหาวิทยาลัยรามคำแหง
มหาวิทยาลัยรามคำแหง
มหาวิทยาลัยรามคำแหง
มหาวิทยาลัยรามคำแหง
มหาวิทยาลัยศรีนครินทรวิโรฒ
มหาวิทยาลัยรามคำแหง
มหาวิทยาลัยเกษตรศาสตร์

วารสารวิจัยรามคำแหง

ฉบับวิทยาศาสตร์และเทคโนโลยี



ฝ่ายจัดทำ	ผู้ช่วยศาสตราจารย์ พรชัย วงศ์วาสนา อาจารย์ ยิ่งยง เมฆลอย นางสาวสายพิน ริตมัต นางลักขิกา ขจัดภัย นางสาวฐนิตฐา ยิบพิกุล นางญนันทพร กลั่นเจริญ นางบุปผา เอี่ยมจรรยา นายอำนาจ สิงห์ตาลง นางสาวแพรวไพไล เจริญสิทธิ์ก้องคำ นางสาวจุริภรณ์ ปานแก้ว	ประธานฝ่ายจัดทำ กรรมการ กรรมการ กรรมการ กรรมการ กรรมการ กรรมการ กรรมการและเหรัญญิก กรรมการและเลขานุการ กรรมการและผู้ช่วยเลขานุการ
-----------	--	--

วารสารวิจัยรามคำแหง ฉบับวิทยาศาสตร์และเทคโนโลยี รับพิจารณาและตีพิมพ์บทความวิจัยสาขาวิชาด้านวิทยาศาสตร์และเทคโนโลยี

กำหนดเผยแพร่ ออก 2 ฉบับต่อปี ตั้งแต่เดือนมกราคม – เดือนมิถุนายน, เดือนกรกฎาคม – เดือนธันวาคม

การเผยแพร่ มอบให้ห้องสมุดหน่วยงานรัฐบาล สถาบันการศึกษาในประเทศ

ข้อมูลการติดต่อ ฝ่ายจัดทำ
กองบรรณาธิการวารสารวิจัยรามคำแหง
สถาบันวิจัยและพัฒนา มหาวิทยาลัยรามคำแหง
อาคารสุโขทัย ชั้น 12 หั้วหมาก บางกะปิ กรุงเทพฯ 10240
โทร. 02-310-8696
Email: ruresearch@ru.ac.th

พิมพ์ที่ สำนักพิมพ์มหาวิทยาลัยรามคำแหง

-
- บทความทุกบทความได้รับการตรวจสอบถูกต้องทางวิชาการโดยผู้ทรงคุณวุฒิ (peer review)
 - มีผู้ทรงคุณวุฒิจากภายนอกประจำกองบรรณาธิการมากกว่าร้อยละ 50
 - มีบทความจากนักวิชาการภายนอกลงตีพิมพ์ไม่น้อยกว่า ร้อยละ 50 ทุกฉบับ
 - ข้อความและบทความในวารสารวิจัยรามคำแหงเป็นแนวคิดของผู้เขียน การตีพิมพ์บทความขึ้นเป็นความรับผิดชอบของผู้เขียน ไม่ใช่ความคิดเห็นและความรับผิดชอบของคณะผู้จัดทำ บรรณาธิการ กองบรรณาธิการและมหาวิทยาลัยรามคำแหง
 - กองบรรณาธิการไม่สงวนสิทธิ์การคัดลอก แต่ควรอ้างอิงแสดงที่มา

วารสารวิจัยรามคำแหง

ฉบับวิทยาศาสตร์และเทคโนโลยี



บทบรรณาธิการ

วารสารวิจัยรามคำแหง ฉบับนี้ เป็นวารสารวิจัยของมหาวิทยาลัยรามคำแหง โดยสถาบันวิจัยและพัฒนา ซึ่งได้ก้าวเข้าสู่ปีที่ 24 ฉบับที่ 1 มกราคม – มิถุนายน 2564 ซึ่งผ่านการประเมินคุณภาพวารสารวิชาการที่อยู่ในฐานข้อมูล TCI

การเปลี่ยนผ่านเข้าสู่ยุคดิจิทัล ที่เทคโนโลยีเข้ามามีผลกระทบต่อสังคมมนุษย์อย่างมาก และรวดเร็ว โดยเฉพาะสังคมไทยที่มีความซับซ้อนทั้งเรื่องอายุ ทัศนคติ ช่องว่างของระดับรายได้และความเหลื่อมล้ำในสังคม การนำเสนอบทความทางวิชาการ อาจเป็นช่องทางหนึ่งในการสื่อสารอย่างพินิจพิเคราะห์ ให้เห็นถึงปัญหาต่างๆ เพื่อนำไปสู่วิถีทางในแก้ไข คลื่นคล้ายปมปัญหาต่างๆ ของสังคมไทยได้อย่างค่อยเป็นค่อยไป ในช่วงเวลาแห่งการเปลี่ยนผ่านนี้

กองบรรณาธิการและฝ่ายจัดทำขอขอบพระคุณ ศูนย์ดัชนีการอ้างอิงวารสารไทย (TCI) กับความพยายามในการรักษามาตรฐานวารสารในประเทศไทยให้คงอยู่ในระดับนานาชาติ อันเป็นความก้าวหน้าทางวิชาการที่สำคัญของประเทศไทย และขอขอบคุณนักวิชาการ ผู้ทรงคุณวุฒิ นักวิจัย ที่ให้เกียรติ ให้ความสนใจนำบทความอันทรงคุณค่าของท่านตีพิมพ์ในวารสารวิจัยรามคำแหง และหวังเป็นอย่างยิ่งว่าจะได้รับความไว้วางใจเช่นนี้ตลอดไป

ขอขอบพระคุณ
บรรณาธิการ

เครื่องจักรเชิงน่าจะเป็นสำหรับเรียนรู้ส่วนเพิ่มเพื่อการจำแนกลำดับสัญลักษณ์

Incremental Learning Probabilistic State Machine for Symbolic Sequences

Classification

จิตรกร พูลโพธิ์ทอง¹



บทคัดย่อ

การจำแนกลำดับสัญลักษณ์สามารถนำไปใช้ประโยชน์ได้หลากหลาย เช่น การวิเคราะห์ดีเอ็นเอ การตรวจจับการบุกรุก การวิเคราะห์คลื่นไฟฟ้าหัวใจ ปัจจุบันมีวิธีการมาตรฐานที่ประยุกต์ใช้ได้กับเรื่องนี้ ได้แก่ แบบจำลองภาษาเชิงน่าจะเป็น แบบจำลองโครงข่ายประสาทเทียม เครื่องจักรเวกเตอร์สนับสนุน เป็นต้น อย่างไรก็ตาม ในกรณีที่ข้อมูลอินพุตสำหรับเรียนรู้เป็นสายอักขระที่มีความยาวมากและไม่มีแบ่งส่วน เช่น สายอักขระดีเอ็นเอ เป็นต้น การนำข้อมูลเหล่านี้ไปเรียนรู้ จะต้องแบ่งคำสมมติที่มีความยาวจำกัด ซึ่งทำให้ความน่าจะเป็นและตำแหน่งของคำที่ถูกต้องบิดเบือนไปจากข้อมูลต้นฉบับ ผลที่ตามมาคือ การนำไปประยุกต์ใช้จะได้ผลลัพธ์ไม่ตรงเท่าที่ควร อีกทั้งจะทำให้การเรียนรู้ส่วนเพิ่มด้วยข้อมูลใหม่เพิ่มเติมจะได้ผลลัพธ์ที่คลาดเคลื่อนเช่นกัน งานวิจัยนี้จึงเสนอแบบจำลองการเรียนรู้สายอักขระแบบใหม่ ที่สามารถเรียนรู้ส่วนเพิ่มได้ไม่จำกัด และไม่ต้องแบ่งคำสมมติ แต่ยังคงข้อมูลสถิติของสายอักขระย่อยที่ถูกต้องให้ได้มากที่สุด ในขณะที่มีข้อมูลอยู่ในขอบเขตที่สามารถจัดการได้ โดยจะใช้วิธีการแบ่งคำอัตโนมัติด้วยสายอักขระเอกลักษณ์และสายอักขระเกิดซ้ำ เพื่อทดสอบประสิทธิภาพของแบบจำลองในงานวิจัยเราได้ทดลองจำแนกสายอักขระดีเอ็นเอของแบคทีเรีย อี.โคไล (E. Coli) 2 กลุ่ม คือ กลุ่มที่เป็นตัวสนับสนุน และกลุ่มที่ไม่เป็น ผลการทดลองพบว่ามีความแม่นยำในการจำแนกกลุ่มถูกต้องร้อยละ 96.23 ซึ่งถือว่ามีความแม่นยำสูงเมื่อเทียบกับวิธีการมาตรฐานอื่น

คำสำคัญ: การจำแนกลำดับสัญลักษณ์ ภาษาเชิงน่าจะเป็น สายอักขระไม่แบ่งส่วน ออโตมาตาเชิงน่าจะเป็น เอ็นแกรม

ABSTRACT

Symbolic sequence classification can be used in a variety of applications such as DNA sequences analysis, intrusion detection, electrocardiography (ECG) analysis. The convention methods which can be applied to this field are for example, probabilistic language model, support vector machine, and artificial neural network. However, sometime the list of words in very long sequences such as DNA sequences are unknown. The fix size of imitation words are used for this case. Consequently, the probability and position of original unknown words are distorted which can lead to incorrect results in long term. Moreover, the update learning may cause difficulty. This research proposes a novel probabilistic language model that can learn infinitely increments and do not have to fix the size of imitation words, but still retain the statistical information of substring accurately as possible while the size of model is tractable. The experiment of classification is

¹ ผู้ช่วยศาสตราจารย์ สาขาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยรามคำแหง

applied to DNA sequences of promoter and non-promoter bacteria *E. Coli*. The accuracy of classification is 96.23 % which is highly accurate compared to the other standard methods.

Keyword: symbolic sequence classification, probabilistic language, unsegmented string, probabilistic finite automata, n-gram

บทนำ

การจำแนกลำดับสัญลักษณ์ (symbolic sequence classification) เป็นวิธีการจำแนกกลุ่มของข้อมูลประเภทสายอักขระหรือลำดับสัญลักษณ์ ซึ่งสามารถนำไปใช้ประโยชน์ได้หลากหลาย เช่น การวิเคราะห์ดีเอ็นเอ การตรวจจับ การบุกรุก การวิเคราะห์คลื่นไฟฟ้าหัวใจ (Xing et al., 2010) วิธีการมาตรฐานที่สามารถประยุกต์ใช้ได้กับเรื่องนี้ ได้แก่ แบบจำลองภาษาเชิงน่าจะเป็น (Vidal et al., 2005) เช่น เอ็นแกรม (n-gram) (Cavnavar and Trenkle 1994) แบบจำลองมาร์คอฟแบบซ่อน (hidden Markov model) (Rabiner, 1989) (Khreich et al., 2012) และเครื่องจักรเวกเตอร์สนับสนุน (support vector machine) (Lodhi et al., 2002) (Leslie et al., 2002) เป็นต้น แต่เนื่องจากข้อมูลสายอักขระในปัจจุบัน มีแนวโน้มเพิ่มขึ้นอย่างมาก ข้อมูลบางประเภทเป็นสายอักขระที่มีความยาวมาก และไม่มีการแบ่งคำใน สายอักขระ เช่น สายอักขระดีเอ็นเอ เป็นต้น ทำให้เกิดปัญหาในการวิเคราะห์ เช่น คำที่มีผลต่อการวิเคราะห์ข้อมูล มีอะไรบ้าง และควรมีความยาวเท่าไรจึงจะเหมาะสม ซึ่งหากวิเคราะห์สถิติการปรากฏของสายอักขระย่อยทั้งหมด เพื่อหาตัวแปรสำคัญ จะทำให้มิติของข้อมูลเพิ่มขึ้นจนไม่สามารถจัดการได้ โดยปัจจุบันการวิเคราะห์สายอักขระ มักจะใช้วิธีแบบจำลองภาษา เช่น เคเมอร์ (k-mer) โดยทั่วไปเรียกว่าเอ็นแกรม (n-gram) ซึ่งเป็นวิธีการเลือก สายอักขระย่อยความยาวจำกัด แทนที่จะวิเคราะห์สถิติของสายอักขระย่อยทั้งหมด ถึงกระนั้นยังพบปัญหาคือ สายอักขระดีเอ็นเอหรือข้อมูลลักษณะเดียวกันนี้ อาจจะมีสายอักขระสำคัญ ที่มีความยาวหลากหลาย การใช้สายอักขระความยาวจำกัดบางส่วน จะทำให้บิดเบือนสถิติของการเกิดคำ และตำแหน่งของคำที่ถูกต้อง ซึ่งส่งผลให้ การนำไปใช้งานมีความคลาดเคลื่อน การแก้ปัญหานี้จึงใช้คำที่มีความยาวเหมาะสมต่อข้อมูล ซึ่งงานวิจัยที่ผ่านมาจะ ใช้วิธีการหาค่า k (ค่า n สำหรับ n -

gram) ที่ดีที่สุด เช่น ค่า k ที่ทำให้เกิดความแตกต่างของค่ามากที่สุด (Rayan and Paul, 2014) หรือ หาค่า k ที่แปรผันได้ โดยหาจากสถิติของค่าความยาว k ที่มีนัยสำคัญ (Ron et al., 1996), (Galata et al., 2001) อย่างไรก็ตามความเห็นของผู้วิจัย การประมาณค่า k ที่เหมาะสมตามงานวิจัยก่อนหน้านี้จะใช้ ความยาวคงที่ ซึ่งยังคงให้ผลสถิติของข้อมูลที่มีความยาวนอกเหนือจากที่กำหนดคลาดเคลื่อนไปได้ นอกจากนี้ การนำแบบจำลองไปเรียนรู้ข้อมูลส่วนเพิ่มก็อาจจะเกิดความคลาดเคลื่อน เพราะความยาวเหมาะสมที่เลือกใช้ อาจจะ เปลี่ยนไปตามสภาพของข้อมูล ดูรายละเอียดเรื่องการเรียนรู้ส่วนเพิ่มจาก (Ade and Deshmukh, 2013), (Geppert and Hammer, 2016)

งานวิจัยนี้จึงสนใจการสร้างแบบจำลองภาษาเชิงน่าจะเป็นแบบใหม่ที่สามารถเก็บสถิติของสายอักขระย่อยให้ คงความถูกต้องให้ได้มากที่สุด โดยไม่จำเป็นต้องแบ่งคำหรือจำกัดความยาวของคำที่ใช้เรียนรู้ และสามารถเรียนรู้ ส่วนเพิ่มได้อย่างมีประสิทธิภาพ ในขณะที่มิติของข้อมูลเพิ่มขึ้นในขณะที่เรียนรู้จะยังคงสามารถจัดการได้ โดยมี แนวคิดว่าการวิเคราะห์สถิติของสายอักขระย่อย ควรพิจารณาจากสายเอกลักษณ์และอักขระเกิดซ้ำทุกความยาวที่มีเพื่อเก็บสถิติให้ตรงมากที่สุด และเพื่อรองรับสถิติของข้อมูลใหม่ที่จะเกิดขึ้นได้โดยไม่ต้องทำการเรียนรู้ใหม่ทั้งหมด โดยตัดสายอักขระย่อยที่ไม่เกิดประโยชน์เพื่อลดทอนขนาดของข้อมูลที่ได้จากการเรียนรู้ จากงานวิจัย (Ilie, 2011) พบว่าสายอักขระทุกสายจะมีคุณสมบัติสายอักขระย่อยเอกลักษณ์สั้นสุด (minimum unique substring) และจะเกิดขึ้น คู่กันกับสายอักขระย่อยเกิดซ้ำยาวสุด (maximum repeated substring) เสมอ ซึ่งหลักการนี้จะเป็นแนวทางที่ช่วยในการตัดสายอักขระเอกลักษณ์ให้สั้นลง ในขณะที่ยังคงความน่าจะเป็นที่ถูกต้องได้ งานวิจัยนี้จึงได้ค้นคว้าวิจัยเพื่อสร้างแบบจำลองแบบใหม่โดยใช้ชื่อว่า ออโตมาตาเชิงน่าจะเป็นของ

สายอักขระเอกลักษณ์ที่เหมาะสม (proper unique substring finite automata: PUSFA) ซึ่งเป็นแบบจำลองที่ใช้เก็บสายอักขระย่อยเอกลักษณ์ที่เหมาะสมและสายอักขระย่อยเกิดซ้ำ รวมถึงความถี่ของสายอักขระเหล่านั้น โดยสามารถประยุกต์ใช้คำนวณหาความน่าจะเป็นของสายอักขระใดๆ เพื่อจำแนกข้อมูล หรือเพื่อทำนายข้อมูล รวมถึงเพื่อวิเคราะห์ สายอักขระที่เกิดขึ้นเชิงสถิติ โดยสามารถเรียนรู้จากสายอักขระไม่จำกัดความยาวได้โดยไม่ต้องทำการแบ่งคำ เราได้เสนอขั้นตอนวิธีในการสร้างแบบจำลอง และทำการทดลองนำแบบจำลองไปประยุกต์ใช้จำแนกกลุ่มข้อมูล เพื่อทดสอบว่ามีความแม่นยำมากน้อยเพียงใดเมื่อเทียบกับวิธีการมาตรฐานอื่น และทำการทดสอบว่าขนาดของแบบจำลองจะเติบโตอย่างไร และเวลาที่ใช้ประมวลผลนั้นเติบโตอย่างไร เมื่อเทียบกับความยาวของสายอักขระ ที่นำมาเรียนรู้ เพื่อที่จะรับรองได้ว่า เมื่อนำแบบจำลองไปใช้งานแล้ว แบบจำลองจะยังคงสามารถจัดการได้จริง ในทางปฏิบัติ เนื้อหาในบทความนี้จะประกอบไปด้วย วิธีการดำเนินการวิจัย ซึ่งจะกล่าวถึงทฤษฎีและนิยามที่เกี่ยวข้อง รายละเอียดแบบจำลองที่นำเสนอ วิธีการสร้างแบบจำลอง วิธีการนำแบบจำลองไปใช้งาน การทดลองที่เกี่ยวข้อง ตามด้วยผลการวิจัยและสรุปและวิจารณ์ผล

วิธีดำเนินการวิจัย

การวิเคราะห์สายอักขระ S ใดๆ เพื่อหาสถิติของคำที่เกิดขึ้นให้ได้ความถูกต้องสูงสุดนั้น จะต้องพิจารณาจาก สายอักขระย่อยที่เป็นไปได้ทั้งหมดของ S หากสมมติให้ $S = x_1 x_2 x_3 \dots x_l$ สายอักขระย่อยของ S ที่เป็นไปได้ทั้งหมดนั้นได้แก่ สายอักขระย่อยความยาว 1, 2, 3... ไปจนถึงความยาว l ซึ่งจะต้องพิจารณาคำ

ทั้งหมด $1+2+3+\dots+l$ หรือ ประมาณ $(l^2 + l)/2$ แต่เนื่องจากในกรณีที่ความยาวของ S ยาวมาก จำนวนของคำที่จะต้องพิจารณา จะมีจำนวนมาก เกินกว่าจะจัดการได้ การตัดทอนสายอักขระย่อยบางส่วนออกไปจะทำให้ลดมิติของข้อมูลไปได้ ขณะเดียวกันจะต้องยังคง ความน่าจะเป็นที่ถูกต้องของสายอักขระย่อยเอาไว้ให้ได้ด้วย ซึ่งจากการสังเกตพบว่าสายอักขระย่อยที่ยาวมากขึ้นโอกาสจะเป็นสายอักขระเกิดครั้งเดียวก็มากขึ้น ซึ่งมีจำนวนมากที่สุด การลดทอนในงานวิจัยนี้จึงตัดเฉพาะสายอักขระเอกลักษณ์ ให้เหลือเพียงสายอักขระเอกลักษณ์ที่เพียงพอต่อการคำนวณความน่าจะเป็นเท่ากับ 1 ยกตัวอย่าง ให้ $S = CACGTGCGA$ เห็นได้ว่า GC เกิดขึ้นครั้งเดียว ดังนั้นความน่าจะเป็น $P(G|GC)$ คือ 1 และ $P(G|TGC) = P(G|GTGC) = P(G|CGTGC) = 1$ ดังนั้นคำที่เพียงพอต่อการหาความน่าจะเป็น จะใช้เพียง GC ก็พอ คำที่ยาวกว่านั้นจะตัดทิ้งไป แต่ในกรณี CG เป็น สายอักขระที่ปรากฏซ้ำ 2 ครั้ง ดังนั้น $P(T|CG) = 1/2$ แต่ $P(T|ACG) = 1$ และ $P(A|CG) = 1/2$ แต่ $P(A|GCG) = 1$ การใช้ CG เป็นเงื่อนไขคำนวณความน่าจะเป็นจะไม่เพียงพอ ดังนั้นการหาความน่าจะเป็นตามเงื่อนไขที่ครบจึงต้องมีสถิติของ คำที่มี CG เป็นคำเติมท้ายที่ยาวกว่าเดิม ซึ่งก็คือ ACG และ GCG ด้วย คำเหล่านี้เราได้นิยามไว้ในด้านล่าง และแบบจำลองเพื่อหาความน่าจะเป็น จะพิจารณาจากคำเหล่านี้ ดูตัวอย่างการสร้างแบบจำลองขึ้นด้วย $S = CACGTGCGA$ ตามรูปที่ 1 ซึ่งเราได้เสนอขั้นตอนวิธีการสร้างแบบจำลองนี้ด้วยขั้นตอนวิธีออนไลน์ เพื่อที่จะสามารถรับข้อมูล มาทีละตัวแล้วสร้างเพื่อรองรับการเรียนรู้ส่วนเพิ่มเพื่อรองรับการเรียนรู้แบบไม่จำกัด รายละเอียดอยู่ในเนื้อหา ส่วนถัดไป

สถานะ (Q)	สถานะเอกลักษณ์ยาวสุดก่อนหน้า (ρ)	สถานะปลายทางและความถี่ของการเกิดตัวอักษรต่อจากสถานะ (δ, τ)				ประเภทสายอักขระ (U, R)
		อักขระ A	อักขระ C	อักขระ G	อักขระ T	
ε	-	(A, 2)	(C, 3)	(G, 3)	(T, 1)	เกิดซ้ำ
A	-	(—, 0)	(AC, 1)	(—, 0)	(—, 0)	เกิดซ้ำ
C	-	(CA, 1)	(—, 0)	(CG, 2)	(—, 0)	เกิดซ้ำ
G	-	(GA, 1)	(GC, 1)	(—, 0)	(GT, 1)	เกิดซ้ำ
T	-	(—, 0)	(—, 0)	(TG, 1)	(—, 0)	เกิดซ้ำ
AC	CA	(—, 0)	(—, 0)	(ACG, 1)	(—, 0)	เอกลักษณ์สิ้นสุด
CA	-	(—, 0)	(AC, 1)	(—, 0)	(—, 0)	เอกลักษณ์สิ้นสุด
CG	-	(GA, 1)	(—, 0)	(—, 0)	(GT, 1)	เกิดซ้ำ
GA	GCG	(—, 0)	(—, 0)	(—, 0)	(—, 0)	เอกลักษณ์สิ้นสุด
GC	TG	(—, 0)	(—, 0)	(GCG, 1)	(—, 0)	เอกลักษณ์สิ้นสุด
GT	ACG	(—, 0)	(—, 0)	(TG, 1)	(—, 0)	เอกลักษณ์สิ้นสุด
TG	GT	(—, 0)	(GC, 1)	(—, 0)	(—, 0)	เอกลักษณ์สิ้นสุด
ACG	-	(—, 0)	(—, 0)	(—, 0)	(GT, 1)	เอกลักษณ์เหมาะสม
GCG	-	(GA, 1)	(—, 0)	(—, 0)	(—, 0)	เอกลักษณ์เหมาะสม

รูปที่ 1 ตัวอย่างแบบจำลองเก็บสายอักขระเกิดซ้ำและสายอักขระเอกลักษณ์

นิยามที่เกี่ยวข้อง

สายอักขระ S สามารถแทนได้ด้วย $S = x_1 x_2 x_3 \dots x_j$ ความยาวของ S เขียนแทนได้ด้วย $|S|$ มีค่าเท่ากับ l สายอักขระย่อยของ S แทนด้วย $S_{i,j} = x_i x_{i+1} x_{i+2} \dots x_j$ เมื่อ $i, j \in I^+$ ความยาวของสายอักขระย่อยแทนด้วย $|S_{i,j}|$ สามารถคำนวณได้จาก $(j - i) + 1$ เช่น $|S_{i,j}|$ มีค่าเท่ากับ 3 คำเติมท้าย (suffix) ของ S แทนได้ด้วย $S_{i,l}$ เป็นสายอักขระย่อยของ S ที่เริ่มจากตำแหน่ง i ไปจนถึง l เช่น ถ้า $S = TGCGA$ แล้ว คำเติมท้ายของ S ที่เป็นไปได้ทั้งหมดได้แก่ $\{\varepsilon, A, GA, CGA, GCGA, TGCGA\}$ และ เซตของคำเติมท้ายแท้ (proper suffix) ของ S หมายถึงคำเติมท้ายทุกตัวยกเว้น S ได้แก่ $\{\varepsilon, A, GA, CGA, GCGA\}$ คำนำหน้า (prefix) ของ S แทนได้ด้วย $S_{1,j}$ เป็นสายอักขระย่อยของ S ที่เริ่มจากตำแหน่ง 1 ไปจนถึง j เช่น ถ้า $S = TGCGA$ แล้ว คำนำหน้าของ S ที่เป็นไปได้ทั้งหมดได้แก่ $\{\varepsilon, T, TG, TGC, TGCG, TGCGA\}$ และเซตของคำนำหน้าแท้ (proper prefix) ของ S หมายถึงคำนำหน้าทุกตัวยกเว้น S

ได้แก่ $\{\varepsilon, T, TG, TGC, TGCG\}$ สายอักขระเกิดซ้ำ (repeating substring) หมายถึง สายอักขระย่อย $S_{i,j}$ ใดๆ ที่ปรากฏขึ้นมากกว่า 1 ครั้งบนสายอักขระ S ตัวอย่างเช่น $S = CACGTGCGA$ แล้วสายอักขระเกิดซ้ำได้แก่ $\{\varepsilon, A, C, G, T, CG\}$ เพื่อความสะดวกในการพิจารณา กำหนดให้สายอักขระที่มีความยาวน้อยกว่า 2 ตัวอักษรเป็นสายอักขระเกิดซ้ำทั้งหมด ไม่ว่าจะปรากฏกี่ครั้งก็ตามสายอักขระเอกลักษณ์ (unique substring) หมายถึง สายอักขระย่อย $S_{i,j}$ ใดๆ ที่ปรากฏขึ้นเพียง 1 ครั้งบนสายอักขระ S ตัวอย่างเช่น $S = CACGTGCGA$ แล้วสายอักขระเอกลักษณ์ได้แก่ AC, CA, GA, GC, GT, ACG, GCG, CGT, GTG, TGC,... เป็นต้น สายอักขระเอกลักษณ์สั้นสุด (minimum unique substring) หมายถึง สายอักขระเอกลักษณ์ที่สั้นที่สุดเท่าที่เป็นไปได้ ซึ่งหากพิจารณาสายอักขระย่อยที่สั้นกว่านั้นจะกลายเป็นสายอักขระเกิดซ้ำ กล่าวได้ว่าหากมีสายอักขระ เอกลักษณ์สั้นสุด $S_{i,j}$ แล้วจะมี $S_{i+1,j}$ และ $S_{i,j-1}$ เป็นสายอักขระเกิดซ้ำ ยกตัวอย่างเช่น $S = CACGTGCGA$

สายอักขระเอกลักษณ์สั้นสุดที่เป็นไปได้ทั้งหมด คือ AC, CA, GA, GC, GT, TG สังเกตได้ว่า สายอักขระย่อย ของคำเหล่านี้ที่สั้นกว่า 1 ตัวอักษรจะเป็นสายอักขระเกิดซ้ำสายอักขระเอกลักษณ์ที่เหมาะสม (proper unique substring) หมายถึง สายอักขระเอกลักษณ์ที่มีคำ นำหน้าเป็นสายอักขระเอกลักษณ์สั้นสุดและไม่มีคำเติมท้ายแท้เป็นสายอักขระเอกลักษณ์สั้นสุดตัวอื่นอีก ยกตัวอย่าง $S = CACGTGCGA$ สายอักขระเอกลักษณ์ที่เหมาะสมที่เป็นไปได้ทั้งหมดได้แก่ AC, CA, GA, GC, GT, TG, ACG, GCG แต่ GTG ไม่เป็นเพราะ มี TG เป็นคำเติมท้าย ข้อสังเกตคือ สายอักขระสั้นสุดทุกตัวเป็นสายอักขระเอกลักษณ์ ที่เหมาะสม

แบบจำลองที่นำเสนอ

แบบจำลองที่นำเสนอในงานวิจัยนี้ เรียกว่า ออโตมาตาจำกัดเชิงนำจะเป็นของสายอักขระเอกลักษณ์ที่เหมาะสม (proper unique substrings finite automata: PUSFA) ซึ่งเรียกโดยย่อว่า พัสฟา มีองค์ประกอบ 7 องค์ประกอบ ได้แก่ $M = (Q, R, U, \Sigma, \delta, \tau, \rho)$ โดยมีรายละเอียดดังนี้ Q คือ เซตของสถานะ (state) โดยแต่ละสถานะจะมีฉลากชื่อกำกับไว้ ซึ่งจะชี้แทนสายอักขระข้อมูลที่ได้ จากการเรียนรู้ ประกอบไปด้วย สายอักขระเกิดซ้ำ สายอักขระเอกลักษณ์เหมาะสม R คือ เซตของสถานะที่เป็นสายอักขระเกิดซ้ำ (repeated substrings) และ $R \subset Q \cup U$ คือ เซตของสถานะที่เป็นสายอักขระเอกลักษณ์ที่เหมาะสม (proper unique substrings) และ $U \subset Q \cup \Sigma$ คือ เซตจำกัดของสัญลักษณ์อินพุต (input alphabet) δ คือ ฟังก์ชันเปลี่ยนสถานะ (transition function) ซึ่งโดยเปลี่ยนสถานะและอินพุตไปเป็นสถานะปลายทาง แสดงได้ด้วยความสัมพันธ์ $\delta: (Q, \Sigma) \Rightarrow Q$ τ คือ ฟังก์ชันความถี่ (frequency function) จะเป็นฟังก์ชันที่เปลี่ยนจากสถานะและอินพุตไปเป็นจำนวนเต็ม แสดงได้ด้วยความสัมพันธ์ $\tau: (Q, \Sigma) \Rightarrow \mathbb{N}$ ρ คือ ฟังก์ชันสายอักขระเอกลักษณ์เหมาะสมก่อนหน้า (previous proper unique substring function) ทำหน้าที่ คำนวณสายอักขระเอกลักษณ์เหมาะสมที่มีความยาวมากที่สุด ที่มีเส้นทางมาถึงสถานะเอกลักษณ์สั้นสุดที่สนใจ แสดงด้วยความสัมพันธ์ $\rho(u) \Rightarrow u \cup \{e\}$ จากรูปที่ 1 สถานะ

GT จะมีสถานะเอกลักษณ์เหมาะสมก่อนหน้า ที่ยาวที่สุดที่มีเส้นทางมาถึง คือ ACG ตัวอย่างอื่นเช่น $\rho(GC) \Rightarrow TG$ แบบจำลองนี้สามารถแทนได้ด้วยตารางตามตัวอย่างในรูปที่ 1

ขั้นตอนวิธีสร้างแบบจำลอง

กำหนดให้ S คือข้อมูลที่นำมาเรียนรู้เพื่อสร้างแบบจำลอง และสมมติให้ $S = x_1 x_2 x_3 \dots x_l$ แล้วหลักการสร้างแบบจำลองอธิบายโดยคร่าว คือ อ่านสายอักขระอินพุตแต่ละตัวจาก x_1 จนถึง x_l ในขณะที่อ่าน x_i แต่ละตัวจะมีสถานะ Q_i หลายตัวที่ต้องพิจารณาอยู่ในคิวซึ่งเป็นเซตของคำเติมท้ายของตัวที่ยาวที่สุดเรียงคิวตามลำดับน้อยไปมาก เริ่มต้นให้ Q_i มีตัวเดียวคือ ϵ หากมีเส้นทางจาก Q_i ไปด้วย x_i อยู่แล้วให้เพิ่มความถี่ แต่หากยังไม่มีเส้นทาง ให้สร้างสถานะใหม่ โดยใช้ชื่อใหม่ด้วย Q_i ต่อกันกับ x_i แล้วตั้งเป็นสถานะปลายทางร่วมเพื่อสร้างเส้นทางของ Q_i ตัวที่ยาวกว่าที่เหลือในคิวชี้เส้นทางมายังสถานะปลายทางร่วมนี้ ซึ่งทุกครั้งที่สร้างสถานะใหม่ นั้นหมายความว่าสายอักขระเอกลักษณ์ปรากฏขึ้น เนื่องจากยังไม่มีเส้นทางมาก่อน สายอักขระเอกลักษณ์นั้นจะเป็นสายอักขระเอกลักษณ์สั้นสุด สายอักขระ เอกลักษณ์สั้นสุดจะมีข้อมูลสายอักขระเอกลักษณ์ก่อนหน้าเก็บไว้ด้วยฟังก์ชัน ρ เพื่อใช้ปรับเปลี่ยน ในกรณีที่ สายอักขระเอกลักษณ์สั้นสุดนั้นเปลี่ยนเป็นสายอักขระเกิดซ้ำ ดูรายละเอียดในขั้นตอนวิธีที่ 1 ในกรณีที่ไม่มีเส้นทางปรากฏอยู่แล้วและปลายทางเป็นสถานะเอกลักษณ์มาก่อน หากเพิ่มความถี่แล้วจะต้อง พิจารณาปรับเปลี่ยนปลายทางให้เป็นสถานะเกิดซ้ำด้วย และเมื่อปรับสถานะบางตัวในคิวให้เป็นสถานะเกิดซ้ำแล้ว เส้นทางของสถานะอื่นที่เคยมีเส้นทางมาที่สถานะตัวนี้รวมกันจะใช้เป็นปลายทางร่วมไม่ได้อีกต่อไป เพราะถูกเปลี่ยน เป็นสายอักขระเกิดซ้ำแล้ว จึงต้องขยายสถานะอื่นๆเหล่านั้นยาวออกไป โดยจะต้องเปลี่ยนเส้นทางใหม่ด้วยการสร้าง สถานะขึ้นมาใหม่ แล้วให้ตัวใหม่นี้จะถูกพิจารณาให้เป็นสถานะปลายทางร่วมสำหรับตัวอื่นๆที่ยาวกว่าต่อไป ซึ่งมีรายละเอียดค่อนข้างมากและซับซ้อน ดูรายละเอียดขั้นตอนวิธีในขั้นตอนวิธีที่ 1 และฟังก์ชันแตกสถานะเพิ่มใน ขั้นตอนวิธีที่ 2

ขั้นตอนวิธี 1 การสร้าง PUSFA

Input: input string S **Output:** Probabilistic Unique Substring Finite Automata M

for x_i **in** S

{ **for** Q_j **in** $activeQueue$

{ **if** $(\tau(Q_j, x_i) == 0)$

{ **if** $target == null$

{ $newQ = Q_j + x_i, \tau(Q_j, x_i) = 1, \delta(Q_j, x_i) = newQ$

if $(|newQ| \geq 2)$

$target = newQ$, **add** $newQ$ **to** U

add $newQ$ **to** $nextActiveQueue$ }

else

{ **if** $\rho(target)$ **is a suffix of** Q_j **then**

{ $\tau(Q_j, x_i) = 1, \delta(Q_j, x_i) = target, \rho(target) = Q_j$

else

{ $newQ = Q_j + x_i$, **add** $newQ$ **to** U , $\tau(Q_j, x_i) = 1, \delta(Q_j, x_i) = newQ$

splitNode $(target, Q_j, x_i)$

$target = newQ$ }

add $target$ **to** $nextActiveQueue$ } }

else

{ $nextState = \delta(Q_j, x_i)$, $\tau(nextState, x_i) += 1$

if $nextState \in U$ **and** $\rho(nextState) \neq NULL$

$target = nextState$

if $\tau(nextState, x_i) > 1$

$nextState \in R$

add $nextState$ **to** $nextActiveQueue$ }}

copy $nextActiveQueue$ **to** $activeQueue$ }}

ขั้นตอนวิธี 2 ฟังก์ชันแตกสถานะเพิ่ม (Split Node)

```

function splitNode(splitState, activeState,  $x_i$ )
{
    newTarget = NULL, splitState  $\in R$ , add splitState to nextActiveQueue
    previousState =  $\rho$ (splitState)
    for each  $Q_k$  where  $Q_k$  is a suffix of previousState
    {
        if  $|Q_k| \geq |splitState|$ 
        {
            if newTarget == NULL
            {
                newNode =  $Q_k + x_i$ ,  $\delta(Q_k, x_i) = newNode$ , newNode  $\in U$ 
                if  $Q_k$  is a suffix of activeState
                {
                    add newNode to nextActiveQueue, newNode  $\in R$ 
                }
                if  $|Q_k| == |activeState|$ 
                {
                    newTarget = newNode
                }
                for c in  $\Sigma$ 
                {
                    if  $\tau(splitState, x_i) > 0$ 
                    {
                        nextOfSplitState =  $\delta(splitState, c)$ ,  $\tau(newNode, c) = \tau(splitState, c)$ ,
                         $\rho(nextSplitState) = newNode$ 
                         $\delta(newNode, c) = nextOfSplitState$ 
                    }
                }
            }
            else
            {
                 $\delta(Q_k, x_i) = newTarget$ , newTarget  $\in U$ 
                 $\rho(newTarget) = Q_k$ 
            }
        }
    }
}

```

การประยุกต์ใช้จำแนกข้อมูล

การจำแนกข้อมูลในที่นี้หมายถึง ปัญหาที่ว่า มีกลุ่มข้อมูล 2 กลุ่ม สมมติเป็น กลุ่ม A และ กลุ่ม B แล้วสายอักขระอินพุต S ที่สนใจนั้นจะเป็นสมาชิกของกลุ่มข้อมูลใดมากกว่ากัน เราเลือกใช้วิธีตัวจำแนกประเภท แบบเบย์อย่างง่าย (Naive Bayes Classifier) ซึ่งใช้กันอย่างแพร่หลายและมีประสิทธิภาพดีในการจำแนกภาษา (Rish, 2001) (Ting et al., 2011) ตัวจำแนกจะตัดสินใจว่าเมื่อใช้แบบจำลอง M เรียนรู้ข้อมูล A และ B แล้ว S จะอยู่ในกลุ่มใด โดยคำนวณ S จากแบบจำลอง เราประยุกต์วิธีการดังกล่าวกับแบบจำลองจากงานวิจัยนี้ด้วยการสร้างแบบจำลองพัลฟายซ์ขึ้นมา 2 ตัว ตัวแรกสร้างขึ้นจากข้อมูลกลุ่ม A สมมติแทนได้ด้วย M_A และอีกตัวหนึ่งสร้างขึ้นจากข้อมูลกลุ่ม B แทนด้วย M_B จากนั้นจะใช้ M_A อ่านอินพุต S เพื่อคำนวณหา

ความน่าจะเป็นของการเกิดสายอักขระ S แทนด้วย $P(S | M_A)$ และใช้ M_B อ่านอินพุต S เพื่อคำนวณความน่าจะเป็นของการเกิดสายอักขระ S แทนด้วย $P(S | M_B)$ ตามหลักการในการ จำแนกสายอักขระใดๆเราจะพิจารณาว่า S อยู่ในกลุ่ม A หรือ B โดยพิจารณาจากสมมติฐานที่มีความน่าจะเป็น มากที่สุด (Maximum a posteriori hypothesis) ซึ่งได้สมการในการจำแนกดังนี้

$$l_{max} = \underset{i \in \{A, B\}}{\operatorname{argmax}} P(M_i) \cdot P(S | M_i) \quad (1)$$

จากสมการที่ (1) l_{max} หมายถึงกลุ่มข้อมูลที่เป็นคำตอบ อาจจะเป็น A หรือ B ซึ่งจะได้มาจากการเลือกค่า M_A หรือ M_B ที่ทำให้ได้ค่าความน่าจะเป็นตามสมการที่มีค่ามากที่สุด ก็จะเป็นคำตอบว่า S ควรอยู่ในกลุ่มใด ปัญหาต่อมา คือ จะมีวิธีการคำนวณ $P(S | M)$ ได้อย่างไร

วิธีการคำนวณความน่าจะเป็น

กำหนดให้สายอักขระย่อยของ S ที่ต้องการหาความน่าจะเป็นคือ $x_{1...i} = x_1 x_2 x_3 \dots x_i$ แล้ว อักขระตำแหน่งที่ i แทนด้วย x_i และความน่าจะเป็นของการเกิด x_i คือ $P(x_i | x_1, x_2, x_3, \dots, x_{i-2}, x_{i-1})$ สามารถคำนวณได้ตามหลักการของ ลูกโซ่มาร์คอฟ แต่เนื่องจากการพิจารณาความน่าจะเป็น ด้วยเงื่อนไขจาก x_{i-1} ย้อนกลับไปถึง x_1 สำหรับข้อมูลที่มี ขนาดยาวมากนั้นจะทำให้ไม่สามารถจัดการได้ การคำนวณในงานวิจัยนี้จึงพิจารณาด้วยเงื่อนไขเพียง x_{i-1} ไปถึง x_{i-k} เมื่อ $x_{i-1...i-k}$ เป็นสายอักขระเอกลักษณ์ที่ปรากฏบนสถานะของแบบจำลองซึ่งเพียงพอต่อการคำนวณความน่าจะเป็น ที่ถูกต้องดังที่ได้อธิบายเหตุผลก่อนหน้านี้ ดังนั้นความน่าจะเป็นในการเกิด x_i คือ $P(x_i | Y)$ เมื่อ Y คือป้ายชื่อของ สถานะปลายทางที่ได้จากฟังก์ชันการใช้ δ เพื่อหาสถานะปลายทางจากอินพุต $x_1, x_2, x_3, \dots, x_{i-2}, x_{i-1}$ ไปตามลำดับ และ $P(x_i | Y)$ คำนวณได้จาก

$$P(x_i | Y) = \frac{\tau(Y, x_i)}{\sum_{c \in \Sigma} \tau(Y, c)} \quad (2)$$

จากสมการที่ (2) $\tau(Y, c)$ คือ ฟังก์ชันความถี่ของการเปลี่ยนสถานะด้วยอักขระ c ที่สถานะ Y เช่น จากรูปที่ 1 $\tau(CG, T)$ จะมีค่าเท่ากับ 1 เป็นต้น และผลรวมของความน่าจะเป็นทุกตัวอักษรคือ $\tau(CG, T) + \tau(CG, A) = 2$ ดังนั้น $P(A|CG) = 1/2$. และความน่าจะเป็นของการเกิดสายอักขระ $x_{1...i} = x_1 x_2 x_3 \dots x_i$ สามารถหาได้จากกฎลูกโซ่

$$P(S|M) = P(x_1|\epsilon) \cdot P(x_2|x_1) \cdot \dots \cdot P(x_{i-1}|x_1 x_2 \dots x_{i-2}) \cdot P(x_i|Y) \quad (3)$$

ตัวอย่างการคำนวณความน่าจะเป็นของ $S = CGTGC$ จากเครื่องจักรในรูปที่ 1 จะได้ว่า $P(CGTGC | M) = P(C | \epsilon) \cdot P(G | C) \cdot P(T | CG) \cdot P(G | GT) \cdot P(C | TG) = 3/9 \cdot 2/3 \cdot 1/2 \cdot 1 \cdot 1 = 1/9$ อย่างไรก็ตาม ในการคำนวณความน่าจะเป็นบางครั้ง $\tau(Y, x_i) = 0$ ยกตัวอย่าง $\tau(TG, T)$ ในรูปที่ 1 ซึ่งจะทำให้การคำนวณ $P(S | M)$ มีค่าเป็น 0 วิธีการแก้ปัญหานี้วิธีหนึ่ง ที่ง่าย เหมาะสำหรับแบบจำลองภาษา คือ การปรับเรียบแบบแบคออฟ (back off smoothing) (Chen, 1999) โดยมีหลักการว่า ถ้ามองหดย้อนกลับไปถึงไกลเกินไป ก็มองกลับไปให้สั้นลง ดังนั้นเมื่อมองเงื่อนไขสั้นลงในเส้นทางเดิมก็จะทำให้ได้ความน่าจะเป็นที่ไม่

เป็น 0 เราประยุกต์ใช้ เทคนิคเดียวกับการแบคออฟ (back off) โดยใช้ค่าเดิมท้ายที่สั้นลงของค่าที่มีปัญหา ซึ่งสามารถให้ความน่าจะเป็นที่ดีที่สุด เป็นเงื่อนไขแทนค่าเดิม ดังนั้นในกรณีที่เป็น $\tau(Y, x_i) = 0$ จะสามารถคำนวณความน่าจะเป็นใหม่ได้จาก สถานะ Z ซึ่งเป็นค่าเดิมท้ายที่ยาวที่สุดของ Y ที่ทำให้ $\tau(Y, x_i) \neq 0$ ซึ่งจะใช้สมการที่ (4) แทนสมการที่ (2) ได้ดังนี้

$$P(x_i | Y) = \underset{Z \in \text{suffix}(Y)}{\operatorname{argmax}} \frac{\tau(Z, x_i)}{\sum_{c \in \Sigma} \tau(Z, c)} \quad (4)$$

ตัวอย่างการคำนวณความน่าจะเป็นของสายอักขระ S ด้วยสมการที่ 4 เช่น ให้ $S = ACGCTG$ แล้วจะได้ $P(ACGCTG | M) = P(A | \epsilon) \cdot P(C | A) \cdot P(G | AC) \cdot P(C | G) \cdot P(T | \epsilon) \cdot P(G | T) = 2/9 \cdot 1 \cdot 1 \cdot 1/3 \cdot 1/9 \cdot 1 = 2/243$ โดยพจน์ที่ขีดเส้นใต้คือ การคำนวณโดยใช้ค่าเดิมท้ายด้วยสมการที่ (4)

การทดลอง

ในงานวิจัยนี้สนใจว่าแบบจำลองจะสามารถนำเอาไปใช้งานเพื่อจำแนกข้อมูลได้แม่นยำขึ้นเมื่อเปรียบเทียบกับวิธีการมาตรฐานที่ตัดคำสมมติด้วยความยาวคงที่หรือไม่ และในการนำแบบจำลองไปเรียนรู้ข้อมูลนั้นแบบจำลอง จะเติบโตอย่างไรและใช้เวลาในการเรียนรู้เป็นอย่างไรเมื่อเทียบกับความยาวข้อมูล เพื่อให้เชื่อมั่นได้ว่าจะสามารถนำแบบจำลองนี้ไปเรียนรู้กับข้อมูลที่มีขนาดใหญ่ได้ เราจึงได้ทำการทดลอง 2 ส่วนคือ 1. ทดสอบความแม่นยำในการจำแนกข้อมูล และ 2. ทดสอบการเติบโตและเวลาที่ใช้ในการเรียนรู้ของแบบจำลองเมื่อเทียบกับความยาวข้อมูล การทดลองเพื่อทดสอบความแม่นยำในการจำแนกข้อมูลนั้น จำเป็นจะต้องเปรียบเทียบประสิทธิภาพกับงานวิจัยอื่นๆ ภายใต้สถานการณ์และข้อมูลชุดเดียวกัน โดยงานวิจัยนี้ได้เลือกใช้ข้อมูลดีเอ็นเอของ แบคทีเรีย อี.โคไล ซึ่งเป็นข้อมูลมาตรฐานในการทดสอบประสิทธิภาพแบบจำลอง (Dua and Graff, 2019) โดยแบ่งกลุ่มข้อมูลดีเอ็นเอจากชุดข้อมูลตัวอย่าง 2 กลุ่ม กลุ่มแรกเป็นข้อมูลดีเอ็นเอที่เป็นตัวสนับสนุน (promoters) และอีกกลุ่มหนึ่ง เป็นดีเอ็นเอที่ไม่เป็นตัวสนับสนุน (non-promoters) โดยตัวสนับสนุนจะเป็นส่วนหนึ่งของดีเอ็นเอ ที่อยู่ส่วนต้นของยีน มีรูปแบบการจัดเรียง นิวคลีโอไทด์เฉพาะตัว ซึ่งถูกนำไปใช้ประโยชน์ทางการวิเคราะห์ดีเอ็นเอ การทดลอง

นี้จะนำข้อมูลดีเอ็นเอทั้งสองกลุ่ม มาสร้างแบบจำลอง เพื่อรู้จำทั้งส่วนที่เป็นตัวสนับสนุนและส่วนที่ไม่เป็น แล้วนำแบบจำลองที่ได้ไปทดสอบ จำแนกส่วนของดีเอ็นเอที่ไม่เคยพบมาก่อนว่าจะเป็นตัวสนับสนุนหรือไม่ ข้อมูลดีเอ็นเอของแบคทีเรียจะมี 2 กลุ่มคือกลุ่มตัว 53 สาย มีความยาวสายละ 57 ตัวอักษร รวม 2 กลุ่มมีจำนวน 106 สาย การแบ่งข้อมูลเพื่อใช้สำหรับทดสอบนั้น จะใช้วิธีการแบ่งข้อมูลเช่นเดียวกับวิธีการมาตรฐานอื่นที่ได้ใช้ข้อมูลเดียวกันนี้ในการทดสอบ เพื่อเปรียบเทียบกัน (Towell et al., 1990) (Xing et al., 2008) โดยใช้หลักการเอาออกหนึ่งสาย (leave one out) เพื่อนำข้อมูลออกจากกลุ่ม ที่ใช้สร้างแบบจำลอง ให้เป็นตัวทดสอบที่แบบจำลองไม่เคยพบมาก่อน เช่น ข้อมูลตัวสนับสนุน 53 สาย เอาออก 1 สายจะเหลือข้อมูลที่น่าไปสร้างแบบจำลอง 52 สาย โดยทำซ้ำจำนวน 53 ครั้ง และอีกกลุ่มหนึ่งคือชุดที่ไม่เป็น ตัวสนับสนุนก็ทำการสร้างแบบจำลองด้วยข้อมูล 53 สาย และทำในวิธีการเช่นเดียวกันนี้ แต่เปลี่ยนชุดที่ดึงออกมาทำเช่นนั้นครบ 53 สาย ข้อมูลแต่ละสายที่ดึงออกมาเพียงสายเดียวในแต่ละครั้งนั้น จะถูกนำมาทดสอบด้วย แบบจำลองว่า มีความน่าจะเป็นเท่าไร และมีความน่าจะเป็น ในกลุ่มใดมากกว่ากัน โดยได้ผลการทดลองดังนี้

ผลการวิจัย

การทดลองได้ผลลัพธ์จากการดึงข้อมูลออกจำนวน 106 ตัวแบ่งเป็นกลุ่มตัวสนับสนุน 53 สาย และกลุ่มที่ไม่เป็นตัวสนับสนุน 53 สาย พบว่าแบบจำลองในงานวิจัยนี้ทำนายกลุ่มตัวสนับสนุนผิดพลาดเพียง 4 สายจาก 53 สาย โดยแบบจำลองได้ทำนายสายดีเอ็นเอที่เป็นตัวสนับสนุนว่าไม่เป็น ซึ่งถือเป็นผลลบลง (false negative) และกลุ่มที่ไม่เป็นตัวสนับสนุนนั้นแบบจำลองสามารถจำแนกกลุ่มได้ถูกต้องครบทุกตัว คิดเป็นความถูกต้องได้ถึง 102 สายจาก 106 สาย ผิดพลาดเพียง 4 สายจาก 106 สาย ซึ่งถูกต้องแม่นยำถึงร้อยละ 96.23 หรือผิดพลาดเพียงร้อยละ 3.77

นอกจากนี้งานวิจัยนี้ได้นำไปเปรียบเทียบกับผลลัพธ์จากวิจัยอื่นๆ ที่เป็นมาตรฐานซึ่งใช้ข้อมูลชุดเดียวกันและ วิธีการทดสอบแบบเดียวกันได้แก่ 1. วิธีการโครงข่ายประสาทเทียมแบบฐานความรู้ (knowledge based neural network) หรือ KBANN (Towell et al., 1990) 2. วิธีการโครงข่ายประสาทเทียมมาตรฐาน (standard neural network) แบบส่งค่าย้อนกลับ (back propagation) หรือ BP (Rumelhart et al., 1986) 3. วิธีการกฎการจำแนกลำดับ (sequential classification rule:SCR) และต้นไม้ตัดสินใจสำหรับลำดับแบบทั่วไป (generalized sequential decision tree: GSDT) (Xing et al., 2008) 4. วิธีการเฉพาะทางในทางชีวสารสนเทศของโอเนล (O'neil, 1989) 5. วิธีการต้นไม้ตัดสินใจ (decision tree) เรียกโดยย่อ ID3 (Mingers, 1989) และวิธีการเอ็นแกรมขนาด 2, 3, 4, 8 แกรม ซึ่งได้ผลดังตารางที่ 1 แสดงให้เห็นว่า เมื่อใช้การแบ่งค่าโดยอัตโนมัติเพื่อคำนวณความน่าจะเป็นตามงานวิจัยนี้จะมีค่าแม่นยำสูงกว่าวิธีการที่ใช้การแบ่งค่าสมมติด้วยความยาวคงที่

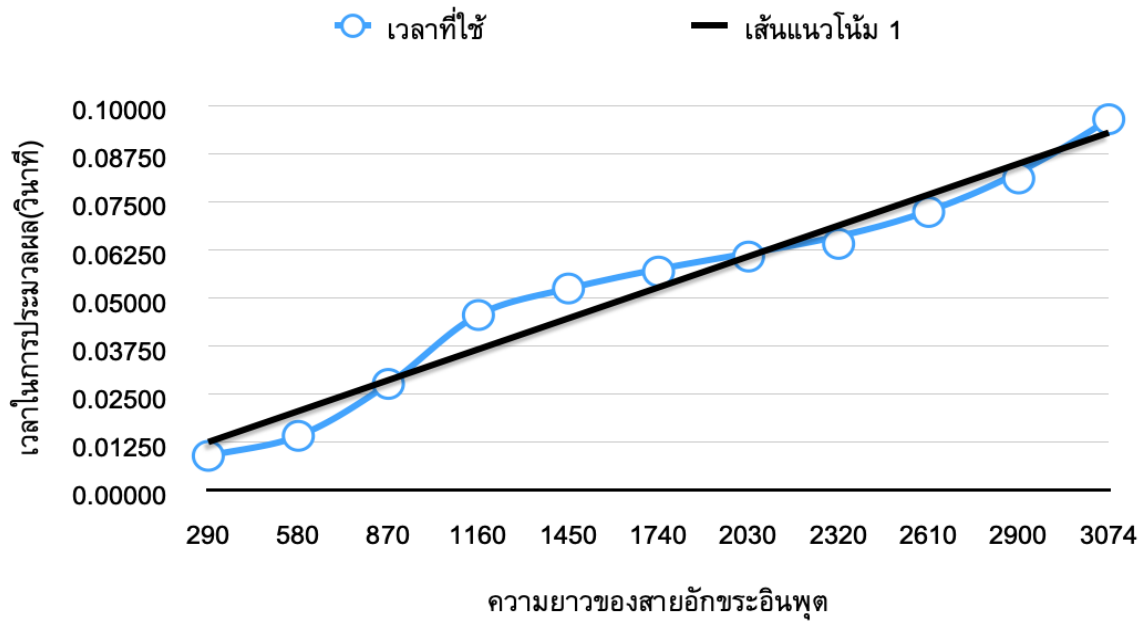
ตารางที่ 1 จำนวนครั้งของความผิดพลาดในการจำแนกข้อมูล

วิธีการที่ใช้จำแนกกลุ่มข้อมูล	จำนวนข้อมูลที่จำแนกผิดพลาด $n/106$
KBANN	4
BP	8
O'neil	12
ID3	21
SCR	7
GSDT	10
2-gram	18
3-gram	13
4-gram	19
8-gram	8
PUSFA	4

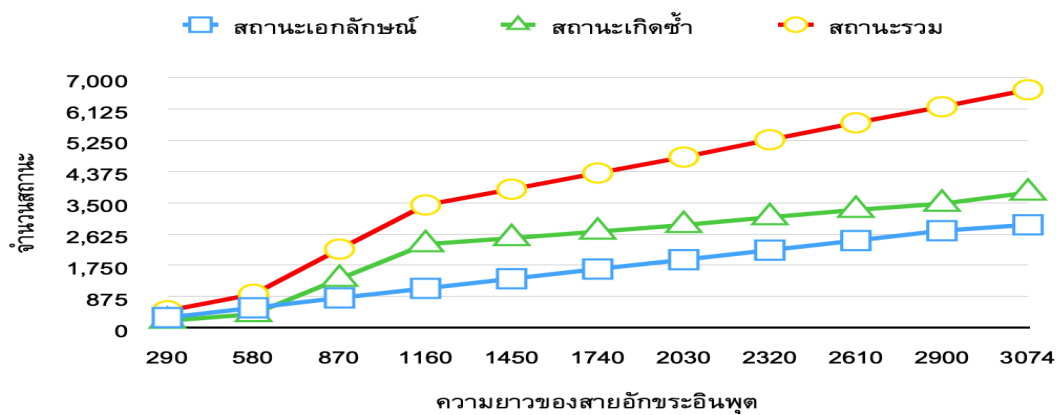
ความซับซ้อนเชิงเวลาและหน่วยความจำในการทำงาน

การทดลองในส่วนที่ 2 เพื่อดูว่า ขนาดของแบบจำลองและเวลาที่ใช้ในการสร้างแบบจำลองนั้นจะเติบโตอย่างไร เพื่อยืนยันว่าเมื่อนำวิธีการสร้างแบบจำลองนี้ไปใช้งานกับข้อมูลขนาดใหญ่จะยังคงจัดการได้ในทางปฏิบัติ โดยเราทำการวิเคราะห์ในทางทฤษฎีได้ว่า การพัฒนาโปรแกรมเพื่อสร้างแบบจำลองตามขั้นตอนวิธี ในงานวิจัยนี้ มีการวนซ้ำเพื่ออ่านสายอักขระเรียนรู้สมมติให้มีความยาว n และในขณะที่อ่านอักขระแต่ละตัว จะต้องพิจารณา แถวค้อยขนาดเท่ากับสายเอกลักษณ์เหมาะสมแต่ละคำสมมติให้เป็น m ซึ่งมีจำนวนตัวแปรผันได้ขึ้นอยู่กับสภาพข้อมูล อย่างไรก็ตามค่า m จะมีค่าไม่เกิน n ดังนั้นความซับซ้อนของขั้นตอนวิธีสามารถคำนวณได้ด้วย $m \cdot n$ ซึ่งจะได้ $O(n^2)$ อย่างไรก็ตามในกรณีที่แย่ที่สุดในกรณี m มีค่าเท่ากับ 2 และหากค่า m มีค่าไม่เกิน k แล้วจะทำให้กรณีที่แย่ที่สุดเป็น $O(n)$ ในทางปฏิบัติเนื่องจาก ข้อมูลที่นำมาเรียนรู้นั้นจะมีค่า m ไม่เกินความยาว 57 ตัวอักษรตามรูปแบบข้อมูล ดังนั้นในทางปฏิบัติแล้ว การเติบโตของเวลาก็ควรเป็น $O(n)$ สำหรับขนาดของแบบจำลองก็

สามารถประมาณได้โดย หากว่าสายอักขระเอกลักษณ์มีความยาวรวม n จะมีสายอักขระย่อยที่เป็นไปได้ทั้งหมด เท่ากับ $\sum_{i=1}^n i$ ซึ่งประมาณได้เป็น $O(n^2)$ แต่หากข้อมูลสายอักขระเกิดซ้ำ มีความยาวไม่เกิน k จะมีสายอักขระย่อยประมาณ k^2 ตัว โดยจะมีจำนวนทั้งหมดแบ่งเป็น n/k จำนวน ดังนั้นจะมีจำนวนสายอักขระย่อยทั้งหมด $k^2 \cdot n/k = kn$ ซึ่งมีขอบเขตเป็น $O(n)$ ซึ่งค่า k ในทางปฏิบัติสำหรับการทดลองนี้ จะมีค่าเท่ากับ 57 ตัวอักษร งานวิจัยนี้ได้ทำการทดลอง เพื่อดูว่าในทางปฏิบัติแล้วจะใช้เวลา และ หน่วยความจำที่ใช้สอดคล้องกับทางทฤษฎีหรือไม่ เราใช้เครื่องคอมพิวเตอร์ระบบปฏิบัติการ macOS v.10.x หน่วยประมวลผล 1.3 GHz Dual-Core Intel Core i5 ทำการสร้างแบบจำลองด้วยข้อมูลเดียวกันกับการทดลองแรก แล้ววัดเวลาและจำนวนสถานะที่ใช้ในแต่ละช่วงที่อ่านข้อมูลแต่ละตัว พบว่ามีการใช้เวลาในการทำงานตามรูปที่ 2 ซึ่งมีแนวโน้มการเติบโตของฟังก์ชันเวลาเป็นเชิงเส้น และ หน่วยความจำที่ใช้วัดจากจำนวนสถานะที่สร้างขึ้นตามรูปที่ 3 ซึ่งมีแนวโน้มการเติบโตของฟังก์ชันพื้นที่เป็นเชิงเส้นเช่นกันซึ่งตรงกับที่วิเคราะห์ทางทฤษฎี ที่มีการเติบโตของเวลาและขนาดเป็น $O(n)$



รูปที่ 2 กราฟแสดงเวลาที่ใช้ในการดำเนินงานเมื่อเทียบกับความยาวของอินพุต



รูปที่ 3 กราฟแสดงจำนวนสถานะที่ใช้ในการดำเนินงานเมื่อเทียบกับความยาวของอินพุต

สรุปและวิจารณ์ผล

งานวิจัยนี้ได้นำเสนอแบบจำลองภาษาเพื่อเรียนรู้สายอักขระหรือลำดับสัญลักษณ์ โดยไม่ต้องแบ่งคำสมมติ ด้วยความยาวคงที่ในกรณีที่สายอักขระมีความยาวมาก โดยสามารถเรียนรู้ส่วนเพิ่มเติมได้ไม่จำกัดโดยไม่ต้องปรับขนาดคำสมมติ โดยในงานวิจัยนี้จะใช้การตัดคำด้วยสายอักขระเอกลักษณ์ที่เหมาะสม และเก็บสายอักขระเกิดซ้ำ ทุกตัวที่เป็นไปได้ เพื่อให้ได้ค่าทางสถิติที่ใช้คำนวณความน่าจะเป็นได้แม่นยำยิ่งขึ้น งานวิจัยนี้ได้เสนอวิธีการนำแบบจำลองนี้ ไปประยุกต์ใช้ในการจำแนกข้อมูลที่มีลักษณะความยาวมากและไม่มีให้เห็นว่าแบบจำลองนี้สามารถประยุกต์ใช้กับข้อมูลขนาดใหญ่ได้ แบบจำลองในงานวิจัยนี้จึงเหมาะสมกับข้อมูลที่มีลักษณะดังกล่าวมา อย่างไรก็ดี แบบจำลองนี้ยัง

การแบ่งคำ เช่น ข้อมูลสายอักขระดีเอ็นเอ เป็นต้น ผลการทดลองแสดงให้เห็นว่า การจำแนกกลุ่มข้อมูลแบบคี่เรียก อี.โคไล ด้วยแบบจำลองของงานวิจัยนี้ ทำได้แม่นยำกว่า แบบจำลองที่แบ่งคำด้วยความยาวคงที่อื่นๆ แต่เท่ากันกับวิธีการ KBANN ซึ่งมีการใช้ความรู้เกี่ยวกับกฎเกณฑ์ ความเป็นโปรโมเตอร์มาเสริมทำให้มีความแม่นยำมากกว่าวิธีการอื่นๆ ในขณะที่ยานวิจัยนี้สามารถทำได้ โดยอัตโนมัติ ปราศจากการใช้ความรู้อื่นมาเสริม การทดลองเพื่อดูอัตราการเติบโตของเวลาและขนาดของแบบจำลองในขณะเรียนรู้เมื่อเทียบกับความยาวข้อมูลที่ใช้เรียนรู้ พบว่าเติบโตแบบเชิงเส้น แสดงให้เห็นว่าการทฤษฎีเกี่ยวกับสายอักขระเอกลักษณ์เหมาะสมมาอธิบายสนับสนุน เช่น ดีกว่าแบบจำลอง เอ็นแกรมทางทฤษฎีในขอบเขตใดบ้าง อีกทั้งยังต้องการขั้นตอน

วิธีสร้างแบบจำลองที่กระชับและง่ายกว่าเดิม เพื่อให้สามารถนำไปใช้งาน ได้ง่ายขึ้น รวมถึงการนำแบบจำลองไปประยุกต์ใช้กับการทำนายข้อมูลและจำแนกข้อมูลอื่นๆ ให้มีความแม่นยำยิ่งขึ้นต่อไป

กิตติกรรมประกาศ

งานวิจัยนี้ได้รับทุนวิจัยจากมหาวิทยาลัยรามคำแหง ประจำปีงบประมาณ 2559 และได้ใช้ข้อมูลชุดสำหรับ ทดลองจากคลังข้อมูลการเรียนรู้เครื่องจักรของมหาวิทยาลัยแคลิฟอร์เนียเออร์ไวน์ (UCI machine learning repository) (Dua and Graff, 2019)

เอกสารอ้างอิง

Ade, R. R. and Deshmukh, P.R. 2013. Methods for Incremental Learning: A Survey. *International Journal of Data Mining and Knowledge Management Process*, 119-125.

Cavnar, W. B. and Trenkle, J. M. 1994. N-Gram-Based Text Categorization. In *Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval*, 161-175.

Chen, S.F. and Goodman, J., 1999. An empirical study of smoothing techniques for language modeling. *Computer Speech and Language*, 13(4), 359-394.

Dua, D. and Graff, C. 2019 UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science.

Galata, A. Johnson, N. and Hogg, D. 2001. Learning Variable Length Markov Models of Behavior. *Computer Vision and Image Understanding*. 81(3): 398-413

Geppert, E. and Hammer, B. 2016. Incremental learning algorithms and applications. *ESANN*, 357-368

Ilie, L., Smyth, W. F. 2011. Minimum Unique Substrings and Maximum Repeats. *Fundam. Inf*, 110 (1-4), 183 - 195.

Leslie, C., Eskin, E., and Noble, W. S. 2002. The spectrum kernel: a string kernel for SVM protein classification. *PAC Symp Biocomput*, 564-575.

Lodhi, H., Saunders, C., Shawe-Taylor, J., Cristianini, N., and Watkins, C. 2002. Text classification using string kernels. *Journal of Machine Learning Research*, 2, 419-444.

Mingers, J. 1989. An Empirical Comparison of Pruning Methods for Decision Tree Induction, *Machine Learning* 4 227-243.

O'Neill, Michael. 1989. Escherichia coli promoters. I. Consensus as it relates to spacing class, specificity, repeat substructure, and three-dimensional organization. *The Journal of biological chemistry*. 264. 5522-30.

Rabiner, L. R. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77 (2), 257-286.

Rayan Chikhi, and Paul Medvedev. 2014. Informed and automated k-mer size selection for genome assembly, *Bioinformatics*, 30(1), 31-37.

Rish, I. 2001. An empirical study of the naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence* 3(22) 41-46.

Ron, D., Singer, Y., and Tishby, N. 1996. The Power of Amnesia: Learning Probabilistic Automata with Variable Memory Length. *Machine learning*, 117-149.

Rumelhart, D. E., G. Hinton, G. E., and Williams, R. J. 1986 Learning Internal Representations by Error Propagation, in: *Parallel Distributed Processing: Explorations in the microstructure of cognition*. Volume 1: Foundations, D. E. Rumelhart and J. L. McClelland (Eds.), 318-363.

- Ting, S. L., Ip, W. H., and Tsang, A. H. 2011. Is Naive Bayes a good classifier for document classification. *International Journal of Software Engineering and Its Applications*, 5(3), 37-46.
- Towell, G., Shavlik, J., and Noordewier, M. 1990. Refinement of Approximate Domain Theories by Knowledge-Based Neural Networks. In *Proceedings of the Eighth National Conference on Artificial Intelligence*, 861-866.
- Vidal, E., Thollard, F., de la Higuera, C., Casacuberta, F., and Carrasco, R. C. 2005. Probabilistic Finite-State Machines-Part I. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27 (7), 1013--1025.
- Xing, Z., Pei, J., Dong, G., and Yu, P. S. 2008. Mining Sequence Classifiers for Early Prediction. *Proceedings of the SIAM International Conference on Data Mining*, 644--655.

การศึกษาสภาวะการเกิดสารก่อมะเร็งเบนซีนในน้ำส้มพาสเจอไรส์ 25 % ที่มีกรดเบนโซอิกและกรดแอสคอร์บิกเป็นส่วนผสม

Study on Carcinogenicity Conditions of Benzene in 25 % Pasteurized Orange Juice Containing Benzoic and Ascorbic Acid as an Ingredient

Chulaporn Lertborwornwong¹



บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาความเป็นไปได้ในการเกิดเบนซีนที่สภาวะการผลิตและการเก็บรักษาแตกต่างกันของน้ำส้มพาสเจอไรส์ 25% ซึ่งได้แก่ ปริมาณโซเดียมเบนโซเอต และกรดแอสคอร์บิกทั้งชนิดเดี่ยวและแบบผสม ค่าพีเอช อุณหภูมิและระยะเวลาการให้ความร้อน สภาวะการเก็บรักษา (อุณหภูมิ เวลา และแสงสว่าง) โลหะหนัก และ สารจับโลหะหนัก ผลการวิจัยพบว่า การเติมโซเดียมเบนโซเอต และ กรดเบนโซอิก 200 ppm และ 120 ppm ตามลำดับ ในผลิตภัณฑ์ที่มี pH 3.7 ฆ่าเชื้อที่อุณหภูมิ 100 °ซ นาน 10 นาที เก็บรักษาที่อุณหภูมิ 30 °ซ นาน 24 ชั่วโมง ตรวจไม่พบเบนซีน แต่พบเบนซีนปริมาณสูงสุดในผลิตภัณฑ์ที่มี pH เท่ากับ 2 ปริมาณการพบเบนซีนในผลิตภัณฑ์ เพิ่มขึ้นจาก 0 เป็น 5.6 ppb ตามค่า pH ที่ศึกษาระหว่าง 3.7 - 2 ตามลำดับ การเพิ่มอุณหภูมิพาสเจอไรส์ จาก 90 - 100 °ซ นาน 20 นาที พบเบนซีน เพิ่มขึ้นจาก 0 - 8.82 ppb ตามลำดับ การเพิ่มระยะเวลาการพาสเจอไรส์ (45 °ซ) เร่งการเกิดเบนซีน มากกว่าระยะเวลาเก็บ แสงสว่างไม่มีผลต่อปริมาณเบนซีน สารโซเดียมเบนโซเอตร่วมกับ กรดแอสคอร์บิกเร่งการเกิดเบนซีนมากกว่าสารเพียงชนิดเดียว (ที่อุณหภูมิ 45 °ซ) การเติมสารจับไอออนโลหะ (Disodium EDTA และ Sodium hexametaphosphate, SHMP) สามารถช่วยลดการเกิดเบนซีนได้

คำสำคัญ: กรดแอสคอร์บิก เบนซีน กรดเบนโซอิก น้ำส้ม

ABSTRACT

The objective of this experiment was to study the possibility of benzene formation at different conditions used in 25% pasteurized orange juice production and storage : benzoate and ascorbic acid alone or together, pH, heating temperature and time, storage conditions (temperature, time and light exposure), metal ions, and chelating agents. An addition of 200 ppm sodium benzoate and 120 ppm ascorbic acid to 25% orange juice, pH 3.7, pasteurized at 100 °C 10 min, stored at 30 °C 24 hrs., did not induce benzene formation. The amount of benzene increased from 0 - 5.6 ppb at pH 3.7 - 2. Increased pasteurization temperature from 90 °C to 100 °C for 20 minutes, found benzene increased from 0 to 8.82 ppb and processing time from 10 to 20 minutes at 100 °C, found benzene risen from 5.32 to 8.82 ppb. Higher storage temperature (45 °C) caused more benzene formation than storage time. Light exposure did not increase

¹ ผู้ช่วยศาสตราจารย์ ภาควิชาเทคโนโลยีอาหาร คณะวิทยาศาสตร์ มหาวิทยาลัยรามคำแหง

benzene formation. Benzene formation in product added only sodium benzoate or ascorbic acid was less than sodium benzoate and ascorbic acid together (stored at 45 °C). The addition of chelating agent (disodium EDTA and sodium hexametaphosphate, SHMP) can decrease benzene formation.

Keywords: ascorbic acid, benzene, benzoic acid, orange juice

Introduction

Benzene (C₆H₆) is recognized by the IARC (International Agency for Research on Cancer) (2018) as carcinogenic to humans and cause damage to germ cells (cell mutagen). Foods or Beverages containing high levels of benzene can cause vomiting, irritation of the stomach, dizziness, sleepiness, convulsions, rapid or irregular heartbeat, death. Long-term exposure to benzene cause harmful effects on the bone marrow, resulting in anemia and excessive bleeding. It can also affect the immune system, increasing the chance of infection. Some women who breathed high levels of benzene for many months had irregular menstrual periods and a decrease in the size of their ovaries (ATSDR, 2018). Its toxicological aspects, occurrence, formation mechanisms, mitigation measures and analyzes data reporting benzene levels in foods has been review by Dos Santos et al. (2015). No limits for benzene levels in beverages and food in general studies have adopted references for drinking water in a range from 1-10 ppb. The presence of benzene has been reported in various food/beverage substances with soft drinks often reported in the literature. Although the analyses reported low levels of benzene in most of the samples studied, and some exceeded permissible limit. The available data on dietary exposure to benzene is minimal from the viewpoint of public health. Often benzene levels were low as to be considered negligible and not a health risk to the consumer, but there is still a need for more studies for a better understanding of their effects on

human health through the ingestion of contaminated food. Factors that influence the formation of benzene in foods have been investigated by many scientists. McNeal et al. (1993) used aqueous models containing 0.025 % ascorbic acid and 0.04 % sodium benzoate the same concentrations used in processed beverages to study the effect of temperature and U.V. on benzene formation. A total of 300 ng / g of benzene formed in models solution stored at 45 °C or under intense U.V. light (wavelength and light intensity not reported) for 20 hours. On the other hand, only 4 ng / g of benzene formed in the dark and at room temperature. However, after 8 days, the benzene concentration had risen to 266 ng / g. Chang and Ku (1993) conducted a study similar to that of McNeal et al. (1993) using the same ascorbic acid and sodium benzoate concentration but did not add copper and iron ions to the samples. After eight days stored in the dark, the solutions had a benzene concentration of 176 ng / g. The addition of chelators, such as 0.1mM sodium EDTA and diethylene-triaminepentaacetic acid (DTPA), prevented benzene formation; 100mM ethanol solutions decrease benzene formation by 90% (Chang and Ku, 1993). Nyman et al. (2010) investigated the influence of accelerated thermal conditions (40 °C or 60 °C for 24 hours or 40 °C for 15 days) and U.V. light (known wavelength and intensity for 24 hours or 7 days) on liquid models containing 0.025 % ascorbic acid and 0.04 % sodium benzoate or processed beverage samples containing one or more benzene precursors (0.04 % sodium benzoate and ascorbic acid are added to

cranberry juice). Polyethylene terephthalate (PET) bottles stabilized with Tinuvin (U.V. light filter) were used to determine how well this U.V light filter inhibits benzene formation. Benzene formation was more significant in sample heated to 40 °C. than in those exposed to U.V. or visible light. However, stabilized PET bottles reduced benzene formation in beverages containing orange juice only marginally. Among the samples not exposed to U.V. light or heat, only those with cranberry juice contained benzene (2.6 ng / g) after the storage period, possibly because benzoic acid is naturally found in cranberries. Casado et al. (2011) studied pickled green olive, cucumbers, and caper berries containing ascorbic acid and benzoic acid under three experimental conditions to determine the chemical and sensory changes induced by these additives: (0) without additives; (1) sodium benzoate; and (2) sodium benzoate and ascorbic acid. The samples were stored in plastic bottles at room temperature (20 °C – 24 °C) or refrigeration (6 °C - 9 °C). Benzene only formed in treatment 2 under storage at room temperature and plastic pouches. For refrigerated or short storage periods (<14weeks),benzene concentrations were similar to those found by McNeal et al. (1993) for pickled vegetables. Green olive stored in plastic bottles had 1.5 g / L of benzene at the end of 67-week period. Depending on the vegetable matrix and packing material, benzene concentration in samples stored for more than one year can exceed 10 g / L, which is the maximum the acceptable concentration of benzene in water

Methods

1. Effect of benzoate and ascorbic acid on benzene formation.

25% orange juice was prepared by using the same recipe as practiced in the commercial factory (water 83.5%, conc.orange juice 5%, fructose syrup 5.4%, sucrose syrup 5.8%, sodium

benzoate 0.02%, citric acid 0.04 %, orange flavor 0.15 %, sunset yellow 0.005 %), pasteurized at 100 °C 10 min., cool to 70 °C then fill in PET. bottle and cool to room temperature (30 °C). Testing models consisted of sodium benzoate (0, 200 ppm) and ascorbic acid (0, 60, 120 ppm) was added to orange juice before pasteurization. Finished products were analyzed for benzene (Central Laboratory (Thailand) Co., Ltd, 2010), benzoic acid (Techakriengkrai and Surakarnkul, 2006), ascorbic acid (Kafkas et al., 2006), copper and iron (William, 2002) after storage for 24 hours. The amount of benzoate and ascorbic acid added to orange juice, which forms the highest benzene, were chosen to continue in experiment 2.

2. Effect of pH on benzene formation.

25% Orange juice was prepared by the same method as 1. The amount of sodium benzoate and ascorbic acid, which formed the highest benzene (result from 1) were added. pH of orange juice was adjusted to 2, 3, 3.7 (commercial product) and 4 with citric acid. Finished products were analyzed for benzene, benzoic acid and ascorbic acid after storage for 24 hours. pH, which formed the highest benzene, was chosen to continue in experiment 3.

3. Effect of pasteurized conditions on benzene formation.

25% Orange juice was prepared by the same method as 1. Sodium benzoate and ascorbic acid were added the same amount as 1, pH of orange juice was adjusted to the result as 2. Time and Temperature of pasteurization was varied from 100 °C, 10, and 20 min.; 90 °C, 15, 20, and 30 min. Finished products were analyzed for benzene, benzoic and ascorbic acid after storage for 24 hours. Pasteurized conditions which form the highest benzene was chosen to continue in experiment 4.

4. Effect of storage conditions on benzene formation.

Prepared pasteurized orange juice the same method as 1. Sodium benzoate 200 ppm and ascorbic acid 120 ppm were added, pH of orange juice was adjusted the same as 3, then pasteurized at Temperature and time which form the highest benzene (result from 3). Products were stored at 4 °C, 35 °C, 45 °C. with and without light exposure. Benzene were analyzed at 1, 15, 30, 45, 60, 75, 90 days of storage.

5. Effect of sodium benzoate and ascorbic acid on benzene formation in pasteurized 25 % orange juice stored at 45 °C for a different period.

Prepared pasteurized orange juice the same method as 1. Sodium benzoate 200 ppm and 120 ppm of ascorbic acid were added, pH was adjusted to 2 (result from 2), pasteurized at 100 °C 20 minutes (result from 3). Products were stored at 45 °C with light exposure (result from 4). Benzene, benzoic, and ascorbic acid were analyzed at 1, 30, 60, 90 days of storage.

6. Effect of metal ions and chelating agents on benzene formation.

Prepared pasteurized orange juice, then added with sodium benzoate and ascorbic acid, adjusted pH, pasteurized and storage at the conditions which formed the highest benzene. Chelating agent tests were disodiumethylenediaminetetraacetic acid (disodium EDTA) and sodium hexametaphosphate (SHMP), each of which varied from 75, 150, 500, 1000 ppm of orange juice. Metal ions were added in the form copper sulfate ($\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$) and ferrous sulfate ($\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$) calculated as Cu 5 ppm and Fe 15 ppm of orange juice. These two levels were the maximum allowance limits of Thai regulation (FDA, 2013). Products were stored at 45 °C with light exposure.

Benzene, benzoic acid, and ascorbic acid in the products were analyzed at 1 and 30 days of storage.

7. Statistical analysis

All experiments were performed in duplicate and results are presented as mean $\pm$ S.D. Primary data and regression analyses were performed using Microsoft Office Excel (Microsoft Corp., USA). Comparison of means was carried out using Minitab version 14 (Minitab Inc., USA) to determine significant differences ($P \leq 0.05$) between treatment tested.

Results

1. Effect of benzoate and ascorbic acid on benzene formation

Benzene was not detected in 25% pasteurized orange juice added with 200 ppm sodium benzoate and ascorbic acid 60 and 120 ppm (Table 1). The analysis result found benzoic acid 165.00 ± 2.00 - 170.00 ± 2.05 ppm in a finished product, which was in the range of 97.36 - 100.31% of original benzoic acid, means that no significant loss occurs in pasteurized juice. Benzoic acid present in the product was within Thai legal limit of beverage in a sealed container (200 ppm) (FDA, 2013). Ascorbic acid remained in pasteurized juice were 32.00 ± 3.00 - 35.00 ± 2.00 ppm from 60 ppm added; and 73.00 ± 8.00 - 81.00 ± 9.00 ppm from 120 ppm added, percent losses were 25.00 - 45.83. Copper was not detected; only iron found but not exceed the legal limit (15 ppm) (FDA, 2013).

2. Effect of pH on benzene formation

25 % pasteurized orange juice added with 200 ppm of sodium benzoate and 120 ppm of ascorbic acid, pH were adjusted to 2, 3, 3.7 and 4. pH did not change after pasteurization. Significantly ($p < 0.05$) Higher benzene concentration was found at

pH 2 (5.6 ± 0.24 ppb) than pH 3 (4.40 ± 0.28 ppb) and did not detect at pH 3.7 and 4 (Table 2). It showed that pH influence the benzene formation in food. Benzoic acid in products did not change by pH, but more destruction of ascorbic acid occurred at higher pH (Table 2)

3. Effect of pasteurized conditions on benzene formation

25% pasteurized orange juice was prepared as 2, adjusted pH to 2 (highest benzene formation), then pasteurized at 100°C for 10, 20 minutes and 90°C for 15, 20, 30 minutes. Benzene, benzoic acid, and ascorbic acid was analyzed in the finished product. Benzene was detected only when pasteurized at 100°C , found 5.32 ± 0.16 ppb for 10 minutes and 8.82 ± 0.22 ppb for 20 minutes, benzene formation was significantly ($P \leq 0.05$) more significant at the longer heating time and higher Temperature (Table 3). The amount of benzoic and ascorbic acid retained in pasteurized juice were not significantly different ($P \leq 0.05$) (Table 3).

4. Effect of storage conditions on benzene formation

25% pasteurized orange juice was prepared as 3, pasteurized at 100°C for 20 minutes (highest benzene formation), then storage at 4°C , 35°C , 45°C ; with and without light exposure. Benzene were analyzed every 15 days until 90 days of storage. Benzene in orange juice samples stored at 4°C with and without light exposure showed non-significant different ($P > 0.05$) and did not change with the increase of storage time (Table 4). Benzene in the samples store at 35°C and 45°C . with and without light exposure increased significantly ($P \leq 0.05$) as storage time increased. Benzene formed in the samples stored at 45°C was higher than at 35°C and 4°C respectively.

5. Effect of sodium benzoate and ascorbic acid on benzene formation in pasteurized 25% orange juice stored at 45°C for a different period.

Benzene formation in the sample contains both benzoic, and ascorbic acid was significantly ($p \leq 0.05$) more than in the sample containing only benzoic acid or ascorbic acid. When stored at 45°C for 30, 60 and 90 days. (Table 5).

6. Effect of metal ions and chelating agents on benzene formation

Two chelating agents were used : disodium EDTA 75 and 150 ppm, which is the maximum allowance in multivitamin solution (<https://www.fda.gov/food/ingredientspackaginglabeling/foodadditivesingredients/ucm091048.htm>, 2018) and SHMP 500 and 1000 ppm, which is the maximum allowance in water-based flavored drink by codex alimentarius (2017). Metal ions used were $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ and $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, calculated as copper 5 ppm and iron 15 ppm which is the maximum allowance in beverages of Thai FDA (2013). Benzene concentration in orange juice samples added with iron 15 ppm (31.67 ppb) and copper 5 ppm (35.38 ppb) was not significant different ($p > 0.05$) but significantly much more than ($p \leq 0.05$) in the sample without metal added (26.64 ppb) (table 6). The conditions without metal ions, with or without disodium EDTA and SHMP were not influenced on benzene formation at 1 day storage, but at 30 days, the amount of benzene in the sample added with disodium EDTA was higher than SHMP and without chelating agents (Table 7). Without the addition of metal ion, copper was not detected but found iron $1.87 \pm 0.01 - 2.19 \pm 0.07$ ppm (Table 1) in orange juice sample. After storage for 30 days, the addition of SHMP 500 and 1000 ppm did not cause a significant different in

benzene formation. The same result also occurred with the addition of disodium EDTA 75 and 150 ppm, but benzene formation was significantly less when added SHMP than without chelating agent and when added disodium EDTA 75 and 150 ppm, but benzene formation was significantly less when added SHMP than without chelating agent and when added disodium EDTA respectively (table 8). Addition of copper 5 ppm, SHMP 500 and 1000 ppm did not cause benzene formed significant

different, the same pattern also occurred with the addition of disodium EDTA 75 and 150 ppm, but the addition of SHMP at two concentrations caused significant less benzene formation than added disodium EDTA and without chelating agent, respectively. This finding means that disodium EDTA can chelate with copper better than iron, but SHMP is the good chelating agent for both iron and copper.

Table 1 Benzene formation, benzoic acid, ascorbic acid, Fe and Cu content in 25 % pasteurized orange juice^a added sodium benzoate and ascorbic acid

Sodium benzoate added (ppm)	Benzoic acid after heating (ppm)	Ascorbic acid added (ppm)	Ascorbic acid after heating (ppm)	Benzene (ppb)	Fe (ppm)^{ns}	Cu (ppm)
0	Not detected	0	Not detected	Not detected	1.87 ± 0.01	Not detected
200 (169.48)	168 ± 2	0	Not detected	Not detected	1.92 ± 0.02	Not detected
0	Not detected	60	32 ± 3	Not detected	1.91 ± 0.03	Not detected
0	Not detected	120	73 ± 8	Not detected	1.98 ± 0.01	Not detected
200 (169.48)	165 ± 2	60	35 ± 2	Not detected	2.04 ± 0.02	Not detected
200 (169.48)	170 ± 2	120	81 ± 9	Not detected	2.19 ± 0.07	Not detected

^a pH 3.7, pasteurized at 100 °C 10 min., stored at room temp for 24 hrs. before benzene analysis

Number in parenthesis calculated as benzoic acid

ns in the same column indicate non significant different (p > 0.05)

Table 2 Effect of pH on benzene formation, benzoic acid, ascorbic acid content in 25% pasteurized orange juice^a

pH	Benzene (ppb)	Benzoic acid (ppm) ^{ns}	Ascorbic acid (ppm)
2	5.60 ^a ±0.24	176 ± 4	89 ^a ± 10
3	4.40 ^b ±0.28	171 ± 6	71 ^b ± 6
3.7 (commercial practiced)	Not detected	177 ± 2	65 ^{bc} ± 6
4	Not detected	177 ± 5	52 ^c ± 3

^a added 200 ppm sodium benzoate and 120 ppm ascorbic acid, pasteurized at 100 °C 10 min., stored at room temp for 24 hrs. before benzene analysis.

ns or same letter under the same column indicate non significant different (p>0.05)

Table 3 Effect of processing conditions on benzene formation, benzoic and ascorbic acid content in 25% pasteurized orange juice^a

Processing conditions	Benzene (ppb)	Benzoic acid (ppm) ^{ns}	Ascorbic acid (ppm) ^{ns}
90 °C 15 min. (commercial practiced)	Not detected	171 ± 4	80 ± 9
90 °C 20 min.	Not detected	176 ± 2	75 ± 9
90 °C 30 min.	Not detected	172 ± 3	71 ± 7
100 °C 10 min.	5.32 ^b ±0.16	176 ± 2	72 ± 10
100 °C 20 min.	8.82 ^a ±0.22	169 ± 2	65 ± 6

^a added 200 ppm sodium benzoate and 120 ppm ascorbic acid, adjusted pH to 2, stored at room temperature for 24 hrs. before benzene analysis.

ns or same letter under the same column indicate non significant different (p>0.05)

Table 4 Effect of storage conditions on benzene formation in 25% pasteurized orange juice^a

Storage time (day)	Benzene (ppb)					
	Stored at 4 °C		Stored at 35 °C		Stored at 45 °C	
	Dark ^{ns}	Light ^{ns}	Dark ^{ns}	Light ^{ns}	Dark ^{ns}	Light ^{ns}
1	7.31 ± 0.07	7.31 ± 0.07	7.31 ± 0.07	7.31 ± 0.07	7.31 ± 0.07	7.31 ± 0.07
15	6.13 ± 0.39	6.87 ± 0.58	10.56 ± 0.58	10.28 ± 0.37	19.46 ± 0.92	18.89 ± 0.94
30	5.95 ± 0.34	6.53 ± 0.07	12.23 ± 0.24	12.52 ± 0.48	31.89 ± 0.66	36.01 ± 2.29
45	5.53 ± 0.44	7.17 ± 0.19	14.66 ± 0.29	14.98 ± 0.50	36.16 ± 1.92	35.48 ± 3.07
60	6.82 ± 0.58	7.43 ± 0.32	17.96 ± 0.17	18.4 ± 1.47	44.31 ± 2.68	39.14 ± 0.71
75	7.29 ± 0.89	7.55 ± 0.97	24.62 ± 4.51	24.78 ± 2.83	44.54 ± 1.35	45.65 ± 2.68
90	7.23 ± 0.18	7.84 ± 0.43	29.15 ± 2.73	30.17 ± 3.70	43.16 ± 0.43	47.17 ± 3.95

^a added 200 ppm sodium benzoate and 120 ppm ascorbic acid, adjusted pH to 2, pasteurized at 100 °C 20 min.

ns benzene concentration stored at each Temperature was non significant different (p>0.05) between dark and light exposure condition

Table 5 Effect of sodium benzoate and ascorbic acid on benzene formation in 25% pasteurized orange juice^a stored at 45 °C for different period of time

Additive (ppm)		Benzene (ppb)			
Sodium benzoate	Ascorbic acid	1 day storage	30 days storage	60 days storage	90 days storage
0	0	Not detected	Not detected	Not detected	Not detected
200	0	10.35 ^a ± 0.34	21.82 ^b ± 1.12	37.28 ^b ± 0.01	40.45 ^b ± 0.42
200	120	7.31 ^b ± 0.07	36.01 ^a ± 2.29	39.14 ^a ± 0.71	47.17 ^a ± 3.95
0	120	Not detected	Not detected	Not detected	Not detected

^a pH 2, pasteurized at 100 °C 20 min.

Same letter under the same column indicate non significant different (p>0.05)

Table 6 Effect of metal catalysts on benzene formation in 25% pasteurized orange juice^a

Iron (ppm)	Copper (ppm)	Benzene (ppb)	
		1 day storage ^{ns}	30 days storage
0	0	8.19 ± 1.67	26.64 ^b ± 0.36
15	0	8.12 ± 0.39	31.67 ^a ± 1.26
0	5	10.44 ± 0.24	35.38 ^a ± 0.51

^a added 200 ppm sodium benzoate and 120 ppm ascorbic acid, adjusted pH to 2, pasteurized at 100 °C 20 min. stored at 45 °C.

ns or same letter under the same column indicate no significant different (p>0.05)

Table 7 Effect of chelating agents on benzene formation in 25% pasteurized orange juice^a

Chelating agents (ppm)		Benzene (ppb)	
Disodium EDTA	SHMP	1 day storage ^{ns}	30 days storage
0	0	8.19 ± 1.67	26.64 ^b ± 0.36
75	0	7.04 ± 0.18	28.77 ^a ± 0.25
150	0	5.84 ± 0.35	28.39 ^a ± 0.25
0	500	6.71 ± 0.83	25.35 ^c ± 0.21
0	1000	6.64 ± 1.26	24.77 ^c ± 0.18

^a added 200 ppm sodium benzoate and 120 ppm ascorbic acid, adjusted pH to 2, pasteurized at 100 °C 20 min. stored at 45 °C.

ns or same letter under the same column indicate no significant different (p>0.05)

Table 8 Effect of chelating agents on benzene formation in pasteurized 25% orange juice added with metal catalysts^a

Chelating agent (ppm)		Benzene (ppb) at 1 day storage		Benzene (ppb) at 30 days storage	
Disodium EDTA	SHMP	Fe 15 ppm	Cu 5 ppm ^{ns}	Fe 15 ppm	Cu 5 ppm
0	0	8.12 ^b ± 0.39	10.44 ± 0.24	31.67 ^b ± 1.26	35.38 ^a ± 0.51
75	0	12.42 ^a ± 0.16	9.61 ± 0.94	38.31 ^a ± 0.80	30.45 ^b ± 0.35
150	0	10.76 ^a ± 0.42	8.16 ± 0.5	38.82 ^a ± 0.95	32.36 ^b ± 1.13
0	500	7.87 ^b ± 1.05	10.13 ± 0.50	25.29 ^c ± 4.05	23.09 ^c ± 0.15
0	1000	8.44 ^b ± 0.66	7.47 ± 1.91	22.59 ^c ± 0.30	22.45 ^c ± 0.64

^a added 200 ppm sodium benzoate and 120 ppm ascorbic acid, adjusted pH to 2, pasteurized at 100 °C 20 min. stored at 45 °C.

ns or same letter under the same column indicate no significant different (p>0.05)

Discussion

Benzene was not detected in all products consisted of sodium benzoate (0, 200 ppm) and ascorbic acid (0, 60, 120 ppm), maybe due to a short storage time (24 hours) before analysis. Sodium benzoate can break down into benzoic acid in acidic conditions such as fruit juice (Aprea et al., 2008); 200 ppm benzoate equivalent to 169.48 ppm benzoic acid. The analysis result found benzoic acid 165.00 ± 2.00 -170.00 ± 2.05 ppm in the finished product, which within Thai legal limit of beverage in a sealed container (200 ppm) (FDA, 2013), means that no significant loss occur in pasteurized juice which due to the heat stability of sodium benzoate (melting point at 249 °C) (Aprea et al., 2008). Losses of ascorbic acid caused by oxygen, higher pH, and heat in the

pasteurization process (Cvetkovic and Jokanovic, 2009). At low acidic condition (high pH), ascorbic acid will change to unstable L-dehydroascorbic acid, then to diketogulonic acid which has no ascorbic acid activity (Lawrence, 2010). Medeiros et al. (2011) described benzene formation from the decarboxylation of benzoate to be pH - dependent and favored at acidic pH. This is why benzene formation was lower or not detected at higher pH (with a sharp drop from pH 3 - 5). When pasteurized samples at 90 °C and 100 °C, benzene was detected significantly ($P \leq 0.05$) more significant at the higher Temperature and longer heating time. It showed that more heat would accelerate the reaction to form more benzene in the product (Gardner and Lawrence, 1993; Aprea et al., 2008). The samples storage at

35 and 45 °C with and without light exposure increased significantly ($P \leq 0.05$) as storage time increased. Benzene formed in the samples stored at 45 °C was higher than at 35 °C and 4 °C respectively. The same pattern also found by Aprea et al. (2008) studied the influence of Temperature on benzene formation using aqueous models with the same ascorbic acid and sodium benzoate concentrations as those found in processed beverages. The solutions were stored at 25 °C and 45 °C for 72 hours. Benzene formation remained constant for the first 12 hours ($<0.1 \mu\text{g/L}$) but increased to $0.44 \mu\text{g/L}$ after 70 hours; in the solution stored at 45 °C, $118.5 \mu\text{g/L}$ of benzene formed after 24 hours, increased to $125 \mu\text{g/L}$ after 48 hours. In the experiment by Morsi et al. (2012), soft drink samples were incubated at 20, 45 and 90 °C for 21 days before the determination of benzene, resulted in a significant increase in samples subjected to Temperature of 45 °C (ranging from 5.5 ppb to 7.4 ppb) and 90 °C (ranging from 25 ppb to 55.1 ppb) compared to the stored samples at 20 °C (ranging from 0.7 ppb to 1.5 ppb). Nyman et al. (2010) investigated the influence of accelerated Temperature (40 °C or 60 °C for 24 hours or 40 °C for 15 days) and U.V. light (known wavelength and intensity for 24 hours or 7 days) on liquid models containing 0.025 % ascorbic acid and 0.04 % sodium benzoate. Benzene formation was greater in samples heated to 40 °C than in those exposed to U.V. or visible light. Benzoic acid can react with the ascorbic acid in a metal catalyzed hydroxyl radical reaction (copper and iron) to form benzene (Gardner and Lawrence, 1993) or could be formed from benzoic acid change directly to benzene or change to benzaldehyde first then to benzene later (Lange et al., 2002). Disodium EDTA can react with ascorbic acid and form benzene ([\[/facts5007478-whatdisodium-edta.html\]\(http://facts5007478-whatdisodium-edta.html\), 2010\). After storage for 30 days the addition of SHMP 500 and 1000 ppm and disodium EDTA 75 and 150 ppm did not cause a significant difference in benzene formation. Disodium EDTA can chelate with copper better than iron, but SHMP is an excellent chelating agent for iron and copper. This could be explained by the stability constant \(log K\) value of copper \(18.8\) is higher than iron \(14.3\); metal ion with higher stability constant value can form a complex with chelating agent better than metal an ion with lower stability constant value \(<http://www.georgebyresearch.com/html/stabilityconstants.htm>, 2011., Igoe and Hui, 2001\)](http://www.ehow.com -</p>
</div>
<div data-bbox=)

Acknowledgments

We would like to express our sincere thanks to the Research and Development Institute Ramkhamhaeng University for funding this project

References

- Aprea, E., Biasioli, F., Carlin, S., M`ark, T.D. and Gasperi, F. 2008. Monitoring Benzene Formation from Benzoate in Model Systems by Proton Transfer Reaction-Mass Spectrometry, *International J. of Mass Spectrometry*. 275(1-3): 117–121.
- ATSDR (Agency for Toxic Substances and Diseases Registry). 2018. Toxic Substance Portal: Benzene (C.A.S.ID#: 71432). <https://www.atsdr.cdc.gov/substances/toxsubstance.asp?toxid=14>.
- Casado, F.J., Sanchez, A.H., de Castro, A., Rejano, L., Beato, V.M. and Montano, A. 2011. Fermented Vegetables Containing Benzoic and Ascorbic Acids as Additives: Benzene Formation During Storage and Impact of Additives on Quality Parameters,” *J. of Agricultural and Food Chemistry*. 59(6): 2403-2409.

- Central Laboratory (Thailand) Co.,Ltd., 2010. In House Method based on J. of. Chromatography A1035(2004)17-22 and detected by GC/MS with Headspace Technique.
- Chang, C. and Ku, K. 1993. Studies on Benzene Formation in Beverages. J. of Food and Drug Analysis. 1 : 385-393.
- Codex Alimentarius, FAO/WHO Foods Standards. 2017.FoodAdditiveGroupDetails:Phosphates. <http://www.fao.org/gsfaonline/groups/details.html?id=18>).
- Cvetkovic, B.R. and Jokanovic, M.R. 2009. Effect of Preservation Method and Storage Condition on Ascorbic Acid Loss in Beverages. BIBLID: 1450-7188. 40,1-7.
- Dos Santos, V.P.S., Salgado, A.M., Torres, A.G. and Pereira, K.S. 2015. Review Article: Benzene as a Chemical Hazard in Processed Foods. International J. of Food Science., p. 1-7.
- FDA (Food and Drug administration). 2013. Notification of the Ministry of Public Health (No.356) on Beverage in Sealed Container, Ministry of Public Health, Thailand.
- Gardner, L. K. and Lawrence, G. D. 1993. Benzene production from decarboxylation of benzoic acid in the presence of ascorbic acid and a transition-metal catalyst. J. of Agricultural and Food Chemistry. 41(5): 693-695.
- <http://www.ehow.com/facts5007478-what-disodium-edta.html>. Retrieved March 5, 2010.
- <http://www.george-eby-research.com/htm/stabilityconstants.htm>.StabilityConstantsofVariousMetalChelates.RetrievedJanuary25,2011.
- <https://www.fda.gov/food/ingredientspackaginglabeling/foodadditivesingredients/ucm091048.htm>. Food AdditiveStatus List. U.S. Food and DrugAdministration. Department of Health and Human Service.Retrieved April 20, 2018.
- Igoe, R.S. and Hui, Y.H. 2001. Dictionary of Food Ingredients 4th ed. U.S.A.: An Aspen Publication. In House Method based on J. of Chromatography A 1035 (detected by GC/MS with Headspace Technique. 2004. 17-22.
- IARC (International Agency for Research on Cancer). 2018. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. World Health Organization : <http://monographs.iarc.fr/ENGhttp://monographs.iarc.fr/ENG/Classification/>.
- Kafkas, E., Kosar, M., Turemis, N. and Baser, K.H.C. 2006. Analysis of Sugars, Organic Acids and Vitamin C Contents of Blackberry Genotypes from Turkey. Food Chemistry 97 : 732-736.
- Lange, M.W., Ommen, J.G.V. and Lefferts.,L. 2002. Deoxygenation of Benzoic Acid on Metal Oxides, 2. Formation of Byproducts. Applied Catalysis A : General, 231 : 17-26.
- Lawrence, J. 2010. Vitamin C : Stability and Quantitative Estimate of L-Ascorbic Acid.<http://www.uwsp.edu/chemistry/jlawreice/Content/Chem%20260%20Content/Chem%20260%20Labs/Vitamin%20C%20Lab.pdf> University of Wisconsin-Stevens Point.
- McNeal, T.P., Nyman, P.J., Diachenko, G.W. and Hollifield, H.C. 1993. Survey of Benzene in Foods by Using Headspace Concentration Techniques and Capillary Gas Chromatography. J. of AOAC International. 76 (6): 1213-1219.
- Medeiros, V.R., De, M. B., Andjelkovic, M., Canfyn, M., Van, O.I. and Van, L, J. 2011. Factors InfluencingBenzene Formation from the Decarboxylation of Benzoate in Liquid Model System. J. Agr. Food Chem. 59(24): 12975-12981.

- Morsi, R.M.Y., EL-Tahan, N.R. and El-Tobgy, K. 2012. Probability of Benzene Forming in Egyptian Non-Alcohol Carbonated Soft Drinks,” Australian J. of Basic and Applied Sciences. 6(3): 271–278.
- Nyman, P.J., Wamer, W.G., Begley, T.H., Diachenko, G.W. and Perfetti, G.I. 2010. Evaluation of accelerated U.V. and Thermal Testing for Benzene Formation in Beverages Containing Benzoate and Ascorbic Acid. J. of Food Science. 75(3) : 263-267.
- Techakriengkrai, I. and Surakarnkul, R. 2006. The Analysis of Benzoic Acid and Sorbic Acid in Thai Rice Wines and Distillates by Solid-Phase Sorbent Extraction and High-Performance Liquid Chromatography. J. of Food Composition and Analysis. 20: 220-225.
- William, H. 2002. Official Method of Analysis of AOAC International. 17th ed. Gaithersburg, Md : Washington DC.

การตรวจสอบฤทธิ์ต้านแบคทีเรียของสารสกัดหยาบจากเปลือกผลไม้ตระกูลส้มห้าชนิด Investigation for the Antibacterial Activity of Five Citrus Fruits' Peel Extracts

Tuangrat Tunvongvinis¹ and Wimon Chanchaem²



บทคัดย่อ

โรคติดเชื้อจากแบคทีเรียเป็นปัญหาที่สำคัญของประชากรโลกที่ทำให้หลายประเทศสิ้นเปลืองงบประมาณเป็นจำนวนมาก อย่างไรก็ตามในช่วงหลายปีที่ผ่านมาได้มีรายงานการพบสารต้านจุลินทรีย์ในพืชสมุนไพรรวมทั้งผลไม้ อย่างต่อเนื่องและคาดว่าสารดังกล่าวจะมีศักยภาพในการช่วยบำบัดโรคติดเชื้อสำคัญๆ ได้มากมาย การศึกษาครั้งนี้มีวัตถุประสงค์ในการนำเปลือกผลไม้ตระกูลส้ม 5 ชนิด ที่ได้รับความนิยมในการบริโภคสูง ได้แก่ ส้มโอ ส้มเขียวหวาน ส้มสายน้ำผึ้ง มะกรูด และมะนาว มาศึกษาฤทธิ์ต้านแบคทีเรียชนิดที่สำคัญ โดยใช้วิธีการสกัด 4 วิธี ได้แก่ การต้มสกัด การหมักด้วยเอทิลแอลกอฮอล์ การปั่นผงแห้งเปลือกผลไม้ในน้ำและการปั่นเปลือกผลไม้สด นำสารสกัดหยาบที่ได้มาทดสอบฤทธิ์ต้านแบคทีเรีย 3 ชนิด ได้แก่ *Escherichia coli*, *Staphylococcus aureus* และ *Pseudomonas aeruginosa* โดยวิธี agar disk diffusion ผลการศึกษาพบว่าสารสกัดหยาบจากเปลือกส้มทั้ง 5 ชนิด โดยวิธีการหมักด้วยเอทิลแอลกอฮอล์มีฤทธิ์ต้านแบคทีเรีย *S. aureus* ได้ สารสกัดหยาบจากการสกัดด้วยวิธีการปั่นผงแห้งเปลือกผลไม้ในน้ำมีฤทธิ์ต้านแบคทีเรีย *P. aeruginosa* ได้ดี ส่วนสารสกัดหยาบจากการสกัดด้วยวิธีการต้มสกัดและการปั่นเปลือกผลไม้สดไม่พบฤทธิ์ต้านแบคทีเรียที่นำมาทดสอบ นอกจากนี้ยังพบว่าฤทธิ์ต้านแบคทีเรียของสารสกัดจากเปลือกผลไม้ตระกูลส้มขึ้นกับปัจจัยหลายประการ ได้แก่ สารละลายที่ใช้ในการสกัด วิธีการสกัด และชนิดของแบคทีเรียที่นำมาทดสอบ โดยสารสกัดหยาบจากเปลือกผลไม้ตระกูลส้ม 5 ชนิดที่นำมาศึกษาในระดับความเข้มข้น 250 mg/200 μ L มีฤทธิ์ต้านแบคทีเรีย *P. aeruginosa* ได้ในระดับใกล้เคียงยาที่ใช้เป็นตัวควบคุมบวก จึงควรศึกษาต่อไปถึงชนิดและปริมาณของสารออกฤทธิ์ เพื่อให้ได้ข้อมูลที่ชัดเจนสำหรับการนำเปลือกส้มเหล่านี้ไปประยุกต์ใช้เป็นสารต้านจุลินทรีย์ในผลิตภัณฑ์ต่าง ๆ ต่อไป

คำสำคัญ: ฤทธิ์ต้านแบคทีเรีย โรคติดเชื้อจากแบคทีเรีย ผลไม้ตระกูลส้ม สารสกัดหยาบจากเปลือกผลไม้

ABSTRACT

Bacterial infections are very important problems of world population. Many countries have to pay large budgets dealing with these diseases. However, there were lots of studies currently reported that fruits and folk plants were rich sources of medicinal compounds those have high potential as antimicrobial agents for alternative treatments. In this study, fruit peels of five citrus fruits including *Citrus maxima*, *Citrus reticulata* Blanco, *Citrus reticulata* Blanco cv. Sainampueng, *Citrus hystrix* DC and *Citrus aurantifolia* were extracted by

¹ Department of Biology Faculty of Science Ramkhamhaeng University

² Assistant Professor Department of Biology Faculty of Science Ramkhamhaeng University

using four methods including boiling, maceration in absolute ethanol, blending of fruit peels powder in water and blending of fresh fruit peels in water. The antibacterial activities of citrus fruits' peels were evaluated using the agar disk diffusion method towards *Escherichia coli*, *Staphylococcus aureus* and *Pseudomonas aeruginosa*. The results showed that crude extracts of these five fruits' peels from ethanolic extraction showed antibacterial activity to *S. aureus* and those using dry powder blending extraction had good antibacterial activity to *P. aeruginosa*. No inhibition zone was found in all five extracts that used boiling extraction and fresh peel blending extraction. This study also showed that the activities of those extracts depended on many factors such as species of fruits, solvent, and the method of extraction. This preliminary study demonstrated that these citrus fruit peels crude extracts at concentration 250 mg/200 μ L had antibacterial activity against *P. aeruginosa* comparable to that of the positive control. Further study for the type and quantity of those important substances should be performed. This will lead to proper applications in pharmaceuticals or other antimicrobial products in future.

Keywords: Antibacterial activity, Bacterial infections, Citrus fruits, Peel extracts

Introduction

Infectious diseases are nowadays major threats of world population. They caused by a vast variety of pathogens with difference virulence levels (Ramadan et al., 2017). Those pathogens may be bacteria, fungi, protozoa or viruses. For bacterial groups there are lots of important pathogens such as *Staphylococcus aureus*, *Streptococcus* sp., *Salmonella* sp., *Shigella* sp., *Escherichia coli*, *Bacillus cereus* and *Listeria monocytogenes* (Ranibar et al., 2018; Barco et al., 2014; Choi et al., 2014; Bloch et al., 2012; Campion et al., 2004).

The rapidly emergence of antibiotic resistant bacterial pathogens has made infectious diseases be more terrible to human and animals than previously (Bengtsson-Palme et al., 2018; Sharma et al., 2018; WHO, 2014; Levy and Marshall, 2004; Anderson, 1999). Chemical drugs have been generally used for treatment of infectious diseases but adverse side effects in some of them and high cost still be problems (Tagel et al., 2018; Tayel et al., 2013). So there has been much attempt to search for another sources of antimicrobial agents. Searching for high

potential but lower problem drugs is a big challenge that gives benefit to human. In Thailand, a great deal of plants are interesting as new candidates for sources of antibacterial agents (Ramadan et al., 2018; Van Vuuran and Holl, 2017; Panda et al., 2016). In this study we aimed to investigate for the antibacterial activity of fruit peels of five citrus fruits those were most common and widely consumed in Thailand including *Citrus maxima* (Pomelo), *Citrus reticulata* Branco (Mandarin orange), *Citrus reticulata* Branco cv. Sainumpueng (Sainumpueng orange), *Citrus hystrix* DC (Kaffir) and *Citrus aurantifolia* (Lime). The research data will be very useful for efficient uses of these fruit peels as antimicrobial agents in future.

Methods

Bacterial strains

Antibacterial activity of five citrus fruits' peels was investigated against three bacterial strains including *Staphylococcus aureus* ATCC25923, *Escherichia coli* ATCC25922, *Pseudomonas aeruginosa* ATCC27853. All strains were maintained in trypticase soy broth containing 15 percent glycerol at -70°C and subcultured onto blood agar before used.

Plant materials and preparations

Plant materials to be studied were fruit peels of five citrus fruits from local markets in Bangkok and nearby areas. The species of those citrus fruits were *Citrus maxima* (Pomelo), *Citrus reticulata* Branco (Mandarin orange), *Citrus reticulata* Branco cv. Sainumpueng (Sainumpueng orange), *Citrus hystrix* DC (Kaffir) and *Citrus aurantifolia* (Lime).

Those fruit peels were washed in tap water and rinsed with distilled water followed by cutting into thin pieces with 0.2 cm long. After being dried by a hot air oven at 50°C for 8 hrs, then they were made to powder by using electrical blender and kept at 4°C in sterile airtight bottles.

Extract preparation

In order to compare the biological activity of fruit peels using different extraction methods, we prepared crude extracts by using four methods including boiling in water, maceration in absolute ethanol, blending of fruit peel powder in water and blending of fresh fruit peels in water.

For boiling method, 40 g of fruit peel powder was soaked in 200 mL distilled water for 20 minutes followed by boiling for 45 minutes then evaporated to get the crude extracts. In maceration method 40 g of peel powder was soaked 120 mL absolute ethanol for 7 days with everyday agitation then collected the extracts and evaporated. The last two methods also used 40 g of peel materials mixed with 200 mL sterile distilled water followed by blending for 5 minutes.

Assay for antibacterial activity of citrus fruits' peel extract

The antibacterial activity was evaluated using agar disk diffusion method (CLSI, 2012). Bacterial suspension at the concentration of 1.5×10^8 cells/mL was spread on Mueller Hinton agar and allowed to dry for 5 minutes. Sterile 6 mm diameter paper discs, each containing 25 μ L of each extract dissolved in dimethyl sulfoxide

(DMSO) at 250 mg/200 μ L, were put on surface of the culture. A positive control (Chloramphenicol, 30 μ g) and a negative control (10% DMSO) were included. All culture plates were incubated at 37°C for 24 hrs then examined for inhibition zone and recorded the zone diameter.

Statistical analysis

All experiments were replicated three times and results were repressed as means $\pm$ SD

Results

Fruit peels' crude extracts prepared from those four methods including boiling in water, maceration in absolute ethanol, blending of fruit peels powder in water and blending of fresh fruit peels in water were assayed for the antibacterial activity using chloramphenicol 30 μ g and 10% DMSO as positive control and negative control, respectively. The results summarized in Table 1 showed that antibacterial activity of citrus fruit peels depended on the species of citrus and methods of extractions and these factors should also be concerned in further applications of citrus fruit peels in order to get best activity

Table 1 Antibacterial activity as determined by inhibition zone diameter of citrus fruit peels' crude extracts from different extraction methods against *E. coli*, *S. aureus* and *P. aeruginosa*

Citrus peel Samples	Inhibition zone diameter (mm)											
	<i>E. coli</i>				<i>S. aureus</i>				<i>P. aeruginosa</i>			
	boiling	abs ethanol	blending	fresh peel blending	boiling	abs ethanol	blending	fresh peel blending	boiling	abs ethanol	blending	fresh peel blending
Pomelo	-	-	-	-	-	9.0±0.0	-	-	-	-	14.3±4.0	-
Mandarin orange	-	-	-	-	-	10.0±0.0	-	-	-	-	16.7±3.5	-
Sainumpu - eng orange	-	-	-	-	-	8.0±0.0	-	-	-	-	8.3±0.6	-
Kaffir	-	-	-	-	-	11.0±1.0	-	-	-	-	10.0±0.0	-
Lime	-	-	-	-	-	11.3±1.5	-	-	-	-	11.7±2.5	-
Pos. ctrl.	28.8	28.8	28.8	28.8	27	27	27	27	8.5	8.5	8.5	8.5
Neg. ctrl.	-	-	-	-	-	-	-	-	-	-	-	-

(-): No inhibition zone

abs ethanol: Absolute ethanol

Pos ctrl: Positive control

Neg ctrl: Negative control

The ethanolic extracts of *Citrus aurantifolia* , *Citrus hystrix* DC and *Citrus reticulata* Branco exhibited antibacterial activity against *S. aureus* with the inhibition zone diameter of 11.3 mm, 11.0 mm and 10.0 mm respectively. Whereas the blending extraction of peels' powder of *Citrus reticulata* Branco and *Citrus maxima* showed very good antibacterial activity against *P. aeruginosa* with the inhibition zone diameter of 16.7 mm and 14.3 mm respectively and was higher than the positive control. However none of crude extract had antibacterial activity against *E. coli*

Figure 1 showed the comparison of antibacterial activity of different citrus fruits' peel ethanolic extracts (Figure 1A) and blending extracts (Figure 1B) against *S. aureus*, *E. coli* and *P. aeruginosa*. Antibacterial activity expressed by inhibition zone diameter obviously showed that two ethanolic extracts and three blending extracts had a good antibacterial activity against *S. aureus* and *P. aeruginosa* respectively.

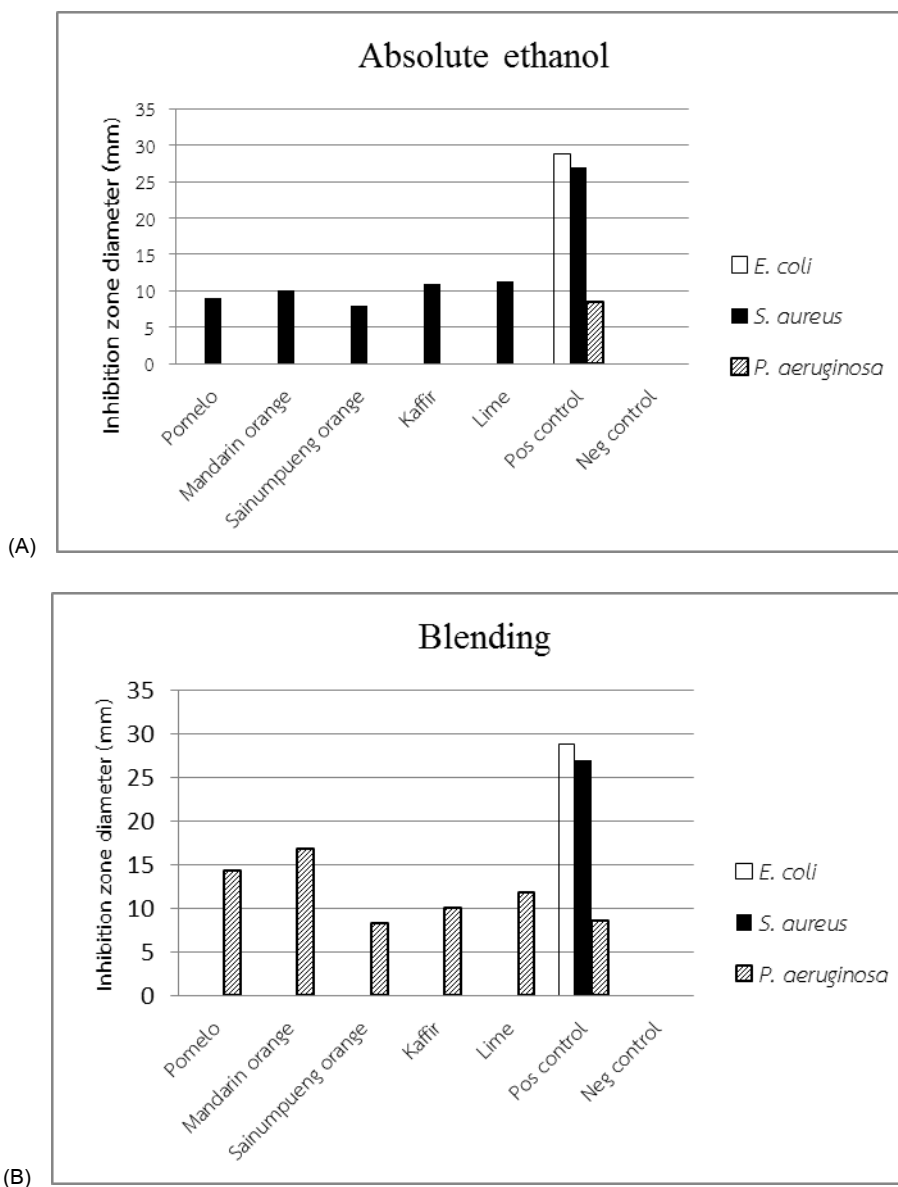


Figure 1 Antibacterial activity as determined by inhibition zone of five citrus fruits' peel ethanolic extracts(A) and blending extractsof peels' powder (B) against *E. coli*, *S. aureus* and *P. aeruginosa*

Discussion and Conclusion

Traditional medicine of Thailand has used citrus fruits for some treatments for long time. Citrus peels have also been used in many aspects such as antimicrobial agents and anti inflammatory agents. The extract of Pamelos' peels was recorded for treatments of some skin infections, abscesses and inflammation. Other citruses such as lime and kaffir were also recorded for traditional uses (Otang and Afolayan, 2016, Suryawanshi,

2011). Scientific research data is necessary for supporting the used of all citruses in future.

In this study fruit peel crude extract of *Citrus reticulata* Branco, one citrus species that was used in traditional medicine to treat some skin infections, showed the outstanding antibacterial activity to *P. aeruginosa* which was an important skin infectious agents at inhibition zone diameter of 16.7 mm, higher than the positive control. Furthermore it was also found that peel extract of

Pamelo, another citrus species that was also used in traditional medicine, showed an interesting antibacterial activity against *P. aeruginosa* at inhibition zone diameter of 14.3 mm. Ethanolic peel extracts of these two citrus fruits also showed antibacterial activity to *S. aureus*, a very important skin infectious agent, corresponding to that reported by Chen et al. (2018) who reported the significant antibacterial effect of Pamelo oil to *S. aureus*. Lime and kaffir are also citrus fruits of interest those harbor antibacterial activity to both *S. aureus* and *P. aeruginosa*, consistent with that reported by Makni et al. (2018) on the significant inhibition of lemon peels against these two pathogenic species.

However, the level of activities were found to depend on species of citrus fruits, method of extraction and the species of tested bacteria. Many factors revealed to be concerned in efficient uses of these plants. The consistent of traditional applications and bioactivity data will provide more confirmation for treatments uses and for the potential of citrus fruits' peels as antimicrobial agent in developments of household products, medical products and other value added applications.

Acknowledgement

We thank The Department of Biology, Faculty of Science Ramkhamhaeng University for providing research facilities.

References

- Anderson, R.M. 1999. The pandemic of antibiotic resistance. *Nat Med.* 5: 147-149.
- Barco, L., Ramon, E., Cortini, E., Longo, A., DallaPozza, M.C., Lettini, A.A., et al. 2014. Molecular characterization of *Salmonella enterica* serotype 4,[5],12:I :DT193 ASSuT strains from two outbreaks in Italy. *Foodborne Pathog Dis.* 11(2): 138-144.
- Bengtsson-Palme, J., Kristiansson, E., Joakim Larsson, D.G. 2018. Environmental factors influencing the development and spread of antibiotic resistance. *FEMS Microbiol Rev.* 42(1): 68-80.
- Bloch, S.K, Felczykowska, A., Nejman Falenezyk, B. 2012. *Escherichia coli* O104: H4 outbreak: have we learnt a lesson from it. *ActaBiochem Pol.* 59(4): 483-488.
- Campion, J.J., McNamara, P.J., Evans, M.F. 2004. Evaluation of ciprofloxacin-resistant *Staphylococcus aureus* in in vitro pharmacokinetic environment. *Antimicrob Agents Chemother.* 48(12): 4733-4744.
- Chen, G.W., Lin, Y.H., Lin, C.H., Jen, H.C. 2018. Antibacterial activity of emulsified Pomelo (*Citrus grandis* Osbeck) peel oil and water soluble Chitosan on *Staphylococcus aureus* and *Escherichia coli*. *Molecules*.23: 840.doi: 10.3390/ molecules 23040840.
- Clinical and Laboratory Standards Institute [CLSI]. 2012. Reference method for dilution antimicrobial susceptibility test for bacterial that grow aerobically. Approved Standard M7-A6.National Committee for Clinical and Laboratory Standards. Wayne, Penn, USA.
- Choi, M.J., Jackson, K.A., Medus, C., Beal, J., Rigdon, C.E., Cloyd, T.C., et al. 2014. Notes from the fields: multistage outbreak of listeriosis linked to soft-ripened cheese-United States 2013. *Morb Motal Wkly Rep.* 63(13): 294-295.
- Levy, S.B., Marshall, B. 2004. Antibacterial resistance worldwide: causes, challenges and responses. *Nat Med.* 10:S122-S129.
- Makni, M., Jemai, R., Kriaa, W., Chtourou, Y., Fetoui, H. 2018. *Citrus limon* from Tunisia: Phytochemical and Physicochemical properties and Biological activities. *BioMed Research International*. Article ID 6251546.

- Otang, W.M. and Afolayan, A.J. 2016. Antimicrobial and antioxidant efficacy of *Citrus limon* L. peel extracts used for skin diseases by Xhosa tribe of Amathole District, Eastern Cape, South Africa. *S Afr J Bot.* 102:46-49.
- Panda, S.K., Mohanta, Y.K., Padhi, L., Park, Y.H., Mohanta, T.K., Bae, H. 2016. Large scale screening of ethnomedical plants for identification of potential antibacterial compounds. *Molecules.* 21(293): 1-20.
- Ramadan, E.M., Abou-Taleb, K.A., Galal, G.F., Abdel-Hamid, N.S. 2017. Antibacterial, antibiofilm and antitumor activities of grape and mulberry leaves ethanolic extracts towards bacterial clinical strains. *Ann Agric Sci.* 62:151-159.
- Ranibar, R., Dehkordi, F.S., Shahreza, M.H.S., Rahimi, E. 2018. Prevalence, identification of virulence factors, O-serogroups and antibiotic resistance properties of Shiga-toxin producing *Escherichia coli* strains isolated from raw milk and traditional dairy products. *Antimicrobial Resistance and traditional dairy products. Antimicrobial Resistance and Infection Control.* 7:53.doi <http://doi.org/10.1186/s13756-018-0345-x>.
- Sharma, C., Rokana, N., Chandra, M., Singh, B.P., Gulhane, R.D., Gill, J.P.S., et al. 2018. Antibacterial resistance: Its surveillance, impact, and alternative management strategies in dairy animals. *Front Vet Sci.* 4:237. doi:10.3389/fvets.2017.00237.
- Suryawanshi, J.A.S. 2011. An overview of *Citrus aurantium* used treatment of various diseases. *Afr J Plant Sci.* 5(7): 390-395.
- Tagel, A.A., Shaban, S.M., Moussa, S.H., Elguindy, N.M, Diab, A.M., Mazrou, K.E., et al. 2018. Bioactivity and application of plant seeds' extracts to fight resistant strains of *Staphylococcus aureus*. *Ann Agric Sci.* 63: 47-53.
- Tayel, A.A., Abdel Monem, O.A., Moussa, S.H., Al Turki, A.I. 2013. Plant extracts as antimicrobials: prospects in food safety and health protection, In: Gionano, A., Costs, A. editors. *Plant Extracts: Role in Agriculture, Health Effects and Medical Applications.* NewYork: Nova Science Publishers, p 311-326.
- Van Vuuren, S. and Holl, D. 2017. Antimicrobial natural product research: a review from a South African perspective for the year 2009-2016. *J Enthnopharmacol.* 208: 236-252.
- World Health Organization [WHO]. 2014. *Antimicrobial resistance: Global report on surveillance 2014.* Geneva, Switzerland: WHO

หลักเกณฑ์และรูปแบบการส่งต้นฉบับบทความเพื่อตีพิมพ์ใน

วารสารวิจัยรามคำแหง

(Ramkhamhaeng Research Journal)



หลักเกณฑ์ในการพิจารณาตีพิมพ์บทความ

1. ผลงานทางวิชาการที่ส่งมาเพื่อตีพิมพ์ต้องไม่เคยผ่านการเผยแพร่ที่ไหนมาก่อน
2. ผลงานทางวิชาการที่ส่งมาเพื่อตีพิมพ์ต้องไม่อยู่ระหว่างการพิจารณาของวารสารอื่น
3. ผลงานทางวิชาการที่ส่งมาเพื่อตีพิมพ์ต้องเป็นบทความที่มีคุณค่าทางวิชาการ คือ เกิดขึ้นจากผู้เขียนได้ทำการทดลองสร้างสรรค์ สังเคราะห์ หรือมีส่วนเกี่ยวข้องกับงานโดยตรง หรือเป็นบทความที่เสนอถึงความคิดหรือหลักการใหม่ที่เป็นไปได้และมีทฤษฎีประกอบหรือสนับสนุนอย่างเพียงพอ มีประโยชน์ต่อการศึกษาและการวิจัย
4. ผลงานทางวิชาการที่ส่งมาเพื่อตีพิมพ์ต้องไม่ได้ลอกเลียนหรือดัดทอนมาจากผลงานวิจัยของผู้อื่นหรือจากบทความอื่นโดยไม่ได้รับอนุญาตหรือปราศจากการอ้างอิงที่ถูกต้อง
5. ผู้เขียนต้องจัดเตรียมต้นฉบับตามรูปแบบข้อกำหนดในการส่งต้นฉบับอย่างเคร่งครัด
6. ผู้เขียนได้แก้ไขความถูกต้องของบทความที่ส่งมาตีพิมพ์ตามข้อเสนอแนะของคณะผู้ทรงคุณวุฒิ (Peer review) แล้ว
7. บทความจะต้องผ่านการตรวจสอบความถูกต้องจากกองบรรณาธิการแล้วเท่านั้น

รูปแบบการจัดเตรียมต้นฉบับ

1. ให้พิมพ์โดยใช้กระดาษ **A4** พิมพ์หน้าเดียว
2. จัดพิมพ์ด้วยโปรแกรม **Microsoft Word for Windows**
3. ใช้ตัวอักษรแบบ **Browallia UPC/New**
4. ระยะห่างระหว่างบรรทัดใช้ **Double Space** โดยมีความยาวไม่เกิน 30 หน้า (รวมเอกสารอ้างอิง)
5. การตั้งค่าหน้ากระดาษ
 - ระยะขอบบน (Top margin) 1" หรือ 2.54 เซนติเมตร
 - ระยะขอบล่าง (Bottom margin) 1" หรือ 2.54 เซนติเมตร
 - ระยะขอบซ้าย (Left margin) 1" หรือ 2.54 เซนติเมตร
 - ระยะขอบขวา (Right margin) 1" หรือ 2.54 เซนติเมตร

รายละเอียดการจัดเตรียมต้นฉบับ

ชื่อเรื่อง ต้องมีทั้งภาษาไทย และภาษาอังกฤษ จัดให้อยู่ชิดซ้ายหน้ากระดาษ ชื่อภาษาอังกฤษ ขึ้นต้นคำให้พิมพ์ด้วยตัวพิมพ์ใหญ่ และให้ใช้ตัวอักษรขนาด 18 ตัวหนา

ชื่อผู้เขียน ให้ใช้ตัวอักษรขนาด 12 ตัวปกติ และให้จัดชิดซ้ายของหน้ากระดาษ โดยให้กำกับหมายเลขยกกำลังไว้ต่อท้ายด้วยสำหรับชื่อของหน่วยงานให้พิมพ์ไว้ในส่วนของเชิงอรรถ (หน้า 1) โดยพิมพ์ชื่อหน่วยงานต้นสังกัดให้ตรงกับตัวเลขยกกำลังที่กำกับไว้ในหน้าเดียวกัน

บทคัดย่อ และ ABSTRACT ใช้ตัวอักษรขนาด 18 ตัวหนา และให้จัดกึ่งกลางหน้ากระดาษ สำหรับเนื้อความให้ใช้ตัวอักษรขนาด 14 ตัวปกติ จัดพิมพ์เป็น 1 คอลัมน์ โดยเนื้อหาต้องครอบคลุมถึงบทนำ วิธีดำเนินการวิจัย ผลการวิจัย สรุปและวิจารณ์ผล เป็นลำดับ

คำสำคัญ (Keywords) : ทั้งภาษาไทย และภาษาอังกฤษ ควรเลือกคำสำคัญที่เกี่ยวข้องกับบทความ ประมาณ 3 - 5 คำ โดยพิมพ์ต่อจากส่วนเนื้อหาของบทคัดย่อ และ Abstract ให้ใช้ตัวอักษรขนาด 14 ตัวปกติ และให้จัดชิดซ้ายของหน้ากระดาษ

เนื้อหา ให้จัดพิมพ์เป็น 1 คอลัมน์ **หัวข้อใหญ่** ใช้ตัวอักษรขนาด 16 ตัวหนา **หัวข้อย่อย** ใช้ตัวอักษรขนาด 14 ตัวหนา จัดชิดซ้ายคอลัมน์ เนื้อความใช้ตัวอักษรขนาด 14 ตัวปกติ โดยให้บรรทัดแรกของทุกย่อหน้าเยื้อง 0.5 นิ้ว ของบรรทัดถัดไป โดยเรียงหัวข้อตามลำดับดังนี้ บทนำ วิธีดำเนินการวิจัย ผลการวิจัย สรุปและวิจารณ์ผล กิตติกรรมประกาศ และเอกสารอ้างอิง

รูปภาพ จัดให้ชิดซ้ายของคอลัมน์ ความละเอียดของรูปภาพไม่น้อยกว่า 600x600 dpi คำบรรยายรูปภาพให้พิมพ์ไว้ใต้รูปภาพ ให้ใช้ตัวอักษรขนาด 12 ตัวปกติ โดยให้แนบไฟล์รูปภาพที่ประกอบในเนื้อเรื่องมาต่างหาก จากไฟล์เอกสารปกติ

ตาราง จัดให้ชิดซ้ายของคอลัมน์ รูปแบบของตารางให้ใช้แบบ Table classic I คำบรรยายตารางให้จัดพิมพ์ไว้ด้านบนของหัวข้อตาราง และใช้ตัวอักษรขนาด 12 ตัวปกติ

หมายเหตุ ต้นฉบับบทความที่น่าส่งจะต้องถูกต้องตามหลักเกณฑ์การเขียนที่กำหนดเท่านั้น จึงจะได้รับการพิจารณาดำเนินการประเมินบทความก่อนตีพิมพ์

การอ้างอิงและการเขียนเอกสารอ้างอิง

• การอ้างอิงในเนื้อเรื่อง

ให้วงเล็บชื่อผู้แต่ง (ภาษาไทย) ชื่อสกุลผู้แต่ง (ภาษาต่างประเทศ) และปีที่พิมพ์ของเอกสารที่อ้างถึง ต่อท้ายข้อความที่ต้องการอ้าง ตัวอย่าง

..... (มณี และคณะ, 2550) หรือ (Archawaranon et al., 2003)
Archawaranon et al. (2003) หรือ มณี และคณะ (2550)

• การอ้างอิงท้ายเรื่อง

1. วารสาร

ก. ภาษาไทย

ชื่อผู้เขียน (ให้เขียนชื่อเต็ม ตามด้วยชื่อสกุล). ปีที่พิมพ์. ชื่อบทความ. ชื่อวารสาร (ใช้ชื่อเต็ม). ปีที่ (ฉบับที่): หน้า
ปรากฏบทความ. เช่น

พุทธชาติ โปธิบาล และ ธนพันธ์ ตรงดี. 2541. สถานะของภาษาดากไบในภาษาถิ่น. วารสาร
สงขลาครินทร์ ฉบับสังคมศาสตร์และมนุษยศาสตร์. 4(2): 167-187.

ข. ภาษาอังกฤษ

ใช้เช่นเดียวกับภาษาไทย แต่ชื่อผู้เขียนใช้ชื่อสกุลขึ้นก่อน, ตามด้วยตัวอักษรย่อของชื่อต้น. และวารสารใช้ชื่อตัวย่อ
ตามเกณฑ์ที่ใช้กัน เช่น

Archawaranon, M. 2003. The impact of human interference on Hill Mynahs *Gracula religiosa*
breeding in Thailand. Bird Conserv. Intl. 13 (2): 139-149.

2. หนังสือ

ชื่อผู้แต่ง. ปีที่พิมพ์. ชื่อหนังสือ. ครั้งที่พิมพ์. สถานที่พิมพ์. ผู้จัดพิมพ์. เช่น

มณี อัครานนท์. 2549. นกขุนทอง: งานวิจัยเพื่อการอนุรักษ์นกในเขตร้อน. กรุงเทพฯ. อมรินทร์พริ้นติ้งแอนด์พับลิชชิ่ง.

Sharp, W.F. 1985. Investment. 3rd ed. New Jersey. Prentice-Hall.

3. บทความ/เรื่อง/ตอน ในหนังสือรวมเรื่องหรือรายงานประจำปี

ชื่อผู้แต่ง. ปีที่พิมพ์. ชื่อบทความ. ใน ชื่อบรรณาธิการ. ชื่อเรื่อง. (ฉบับพิมพ์ ถ้ามี), หน้าปรากฏบทความ.

ผู้จัดพิมพ์. สถานที่พิมพ์. เช่น

เสรี ลีลาชัย. 2542. เศรษฐกิจชาตินิยมในประเทศกำลังพัฒนาและสถานการณ์ในประเทศไทย. ใน ณรงค์ เพ็ชรประเสริฐ
(บรรณาธิการ). 1999 จุดเปลี่ยนแห่งยุคสมัยใหม่ (หน้า 90-141). ศูนย์ศึกษาเศรษฐศาสตร์การเมือง คณะ
เศรษฐศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย. กรุงเทพฯ.

4. การอ้างอิงจากวารสารออนไลน์

ชื่อผู้แต่ง. ปีที่พิมพ์. ชื่อบทความ. ชื่อวารสาร. ปีที่และฉบับที่พิมพ์: หน้าปรากฏบทความ. ที่มา:

สถานที่มาของสารสนเทศ. เช่น

Overby, J.M. 1996. Ozone brings better water. Water Technology [Online]. 19, no. 5:
62-64. Abstract from Dialog File: Water Resources Abstract (117) Item: 00798344

5. วิทยานิพนธ์

ชื่อผู้แต่ง. ปีที่พิมพ์. ชื่อวิทยานิพนธ์. สถานที่พิมพ์. ชื่อสถาบันการศึกษา

พรชัย วงศ์วาสนา. 2543. การศึกษาชีววิทยาประชากรของนกขุนทองในประเทศไทย. วิทยานิพนธ์

ปริญญาวิทยาศาสตรมหาบัณฑิต. กรุงเทพฯ. บัณฑิตวิทยาลัย มหาวิทยาลัยรามคำแหง.

การส่งต้นฉบับ

ส่งต้นฉบับ จำนวน 4 ชุด พร้อมแผ่น CD จำนวน 1 แผ่น ไปยัง กองบรรณาธิการวารสารวิจัยรามคำแหง สถาบันวิจัยและ
พัฒนา อาคารสุขโขทัย ชั้น 12 มหาวิทยาลัยรามคำแหง ถนนรามคำแหง แขวงหัวหมาก เขตบางกะปิ กรุงเทพฯ 10240 โทรศัพท์ 0-2310-
8696, โทรสาร 0-2310-8119 Email: ruresearch@ru.ac.th