

ระบบแนะนำสถานที่ท่องเที่ยวสำหรับนักท่องเที่ยว โดยใช้เทคนิค ดาต้าไมนิ่ง

เรखा โสมพงษ์¹, ธงชัย แก้วกิริยา²

สาขาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ, สถาบันเทคโนโลยีไทย-ญี่ปุ่น

Emails: so.raykha_st@tni.ac.th¹

thongchai@tni.ac.th²

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อนำเสนอแนวคิดการออกแบบระบบแนะนำสถานที่ท่องเที่ยว สำหรับนักท่องเที่ยว ด้วยการหาเหมืองข้อมูล ด้วยการใช้อัลกอริทึมการวิเคราะห์ข้อมูลแบบต้นไม้ตัดสินใจ (Decision Tree) โดยทำการเปรียบเทียบระหว่าง 3 อัลกอริทึม ได้แก่ อัลกอริทึม J48 REPTree และ Simple Cart ซึ่งในงานวิจัยนี้ได้แบ่งสถานที่ท่องเที่ยวออกเป็น 3 ประเภท คือ 1) Natural เป็นสถานที่ท่องเที่ยวประเภท เกาะ แก่ง ทะเล น้ำตก และสถานที่ท่องเที่ยวที่เกิดขึ้นเองโดยธรรมชาติ 2) Historical เป็นสถานที่ท่องเที่ยวประเภท โบราณสถาน และสถานที่ท่องเที่ยวที่มีความเกี่ยวข้องกับประวัติศาสตร์ และ 3) Cultural เป็นสถานที่ท่องเที่ยวประเภทศิลปะ วัฒนธรรม และสถานที่ท่องเที่ยวที่มนุษย์สร้างขึ้น โดยผลการประมวลผลความแม่นยำของระบบแนะนำอัลกอริทึม J48 มีความแม่นยำเท่ากับ 84.73% ในการประเมินความเหมาะสมของกรอบแนวคิดระบบแนะนำสถานที่ท่องเที่ยว ได้ค่าเฉลี่ยเท่ากับ 4.67 ส่วนเบี่ยงเบนมาตรฐานมีค่า 0.58

คำสำคัญ-- Data mining, Decision Tree, J48, Recommendation System, Tourism Attraction

1. บทนำ

การท่องเที่ยว หมายถึง การเดินทางของบุคคล หรือ กลุ่มบุคคล เพื่อไปยังสถานที่ใดสถานที่หนึ่ง ซึ่งไม่ใช่สถานที่พักอาศัยในปัจจุบัน เพื่อการพักผ่อนหย่อนใจ เป็นการชั่วคราว และไม่ใช้การเดินทางที่สืบเนื่องจากการกระทำที่ทำงานรูปแบบของสถานที่สำหรับท่องเที่ยว แบ่งออกได้หลายประเภทเช่น สถานที่ท่องเที่ยวเชิงธรรมชาติ เชิงศิลปวัฒนธรรม เชิงประวัติศาสตร์ เป็นต้น โดยความชื่นชอบในรูปแบบของการท่องเที่ยว สำหรับแต่ละบุคคล จะมีความแตกต่างกัน ทั้งนี้ขึ้นอยู่กับลักษณะ/พฤติกรรมส่วนบุคคล ทักษะการตัดสินใจ รวมถึงอายุ เพศ และการศึกษา ของแต่ละบุคคล นอกจากนี้ทัศนคติของนักท่องเที่ยวเองแล้วสถานที่ท่องเที่ยวเองก็จำเป็นต้องมี เอกลักษณ์ ความแตกต่าง ทั้งในด้านลักษณะของภูมิประเทศ ศิลปะพื้นบ้าน เพื่อแบ่งแยกความเป็นเอกลักษณ์

ของสถานที่ท่องเที่ยวแต่ละที่ออกจากกัน และเพื่อให้สามารถดึงดูดนักท่องเที่ยวให้เข้ามาเที่ยวยังสถานที่ท่องเที่ยววันนั้น

ปัจจุบันจำนวนของนักท่องเที่ยวชาวไทยที่เดินทางท่องเที่ยวภายในประเทศ มีเปอร์เซ็นต์เพิ่มขึ้นจากปี 2553-2556 จำนวน 11% ดังตารางที่ 1 [1] และประเทศไทยมีแหล่งท่องเที่ยวที่ลงทะเบียนกับการท่องเที่ยวแห่งประเทศไทยมากกว่า 4,500 แห่ง ทั่วประเทศ [2] โดยรูปแบบของสถานที่สำหรับท่องเที่ยววันนั้น สามารถแบ่งออกได้หลายประเภท เช่น สถานที่ท่องเที่ยวเชิงธรรมชาติ เชิงศิลปวัฒนธรรม เชิงประวัติศาสตร์ เป็นต้น การเลือกสถานที่ท่องเที่ยวสำหรับนักท่องเที่ยวแต่ละคนนั้น จะแตกต่างกันออกไป ทั้งนี้ขึ้นอยู่กับพฤติกรรม ทักษะการตัดสินใจ รูปแบบการดำเนินชีวิต ของนักท่องเที่ยวแต่ละคน ทางด้านสถานที่ท่องเที่ยวเองก็จำเป็นต้องมีเอกลักษณ์เฉพาะตัว ทั้งในด้านลักษณะของภูมิประเทศ ศิลปะพื้นบ้าน หรือสิ่งดึงดูดเฉพาะตัว เพื่อสร้างความแตกต่างของแต่ละสถานที่และเพื่อให้สามารถดึงดูดนักท่องเที่ยวให้เข้ามาเที่ยวชม ซึ่งในขณะนี้จากการลงทะเบียนสถานที่ท่องเที่ยว จะเห็นได้ว่า สถานที่ท่องเที่ยวในรูปแบบต่างๆ มีเพิ่มมากขึ้น ไม่ว่าจะเป็นสถานที่ท่องเที่ยวที่เกิดขึ้นตามธรรมชาติ หรือเกิดจากมนุษย์สร้างขึ้น เมื่อนักท่องเที่ยวมีความต้องการที่จะเดินทางท่องเที่ยวเพื่อพักผ่อนหย่อนใจ จะใช้เวลาในการหาข้อมูลสถานที่ท่องเที่ยวตามความต้องการของตนเอง จากหลากหลายแหล่งที่มา เพื่อใช้ประกอบการตัดสินใจเลือกสถานที่ที่ต้องการ จึงทำให้ในปัจจุบันมีการนำเทคโนโลยีเข้ามาช่วยในการตัดสินใจ และช่วยอำนวยความสะดวกให้กับผู้เข้าชมมากขึ้น ในด้านของการท่องเที่ยว โดยเฉพาะในเรื่องของการหาข้อมูลสถานที่ท่องเที่ยว คำแนะนำจากบุคคลอื่น บน Social Media หรือ ตามเว็บไซต์ต่าง ๆ เป็นตัวเลือกที่ช่วยในการตัดสินใจ

เทคนิคเหมืองข้อมูล เป็นเทคนิควิธีการในการวิเคราะห์ความสัมพันธ์ของข้อมูลจำนวนมาก เพื่อค้นหารูปแบบ และความสัมพันธ์ที่ซ่อนอยู่ในข้อมูลนั้นๆ เป็นเทคนิคที่นิยมนำมาใช้เพื่อการตลาดในการทำโปรโมชั่น รวมถึงเพื่อสนับสนุน และช่วยในการตัดสินใจของผู้บริหาร นอกจากนี้การทำเหมืองข้อมูล ยังสามารถหาความสัมพันธ์ ความเกี่ยวข้องกันของข้อมูลที่ต้องการนำมาใช้ประโยชน์ เพื่อนำมาประยุกต์ใช้ประโยชน์ทางธุรกิจ และในด้านอื่นๆ

ผู้วิจัยจึงมีแนวคิดในการพัฒนาระบบแนะนำสถานที่ท่องเที่ยวสำหรับนักท่องเที่ยว โดยใช้เหมืองข้อมูล เพื่อเป็นการรวบรวม วิเคราะห์ข้อมูล ที่เกี่ยวข้อง และแนะนำสถานที่ท่องเที่ยวที่เหมาะสมกับนักท่องเที่ยวแบบ Hybrid คือการใช้แบบสอบถามในการเก็บข้อมูล และหาความสัมพันธ์ของข้อมูล แล้วนำมาสร้างเป็นระบบเพื่อใช้ในการแนะนำสถานที่ท่องเที่ยวให้กับนักท่องเที่ยว เพื่อให้

สามารถตอบสนองความต้องการของนักท่องเที่ยวได้อย่างถูกต้อง และเพื่อให้นักท่องเที่ยวผู้ใช้ระบบสามารถตัดสินใจเลือกสถานที่ท่องเที่ยวในรูปแบบที่ต้องการได้

จำนวนผู้เยี่ยมชม	2553	Δ%	2554	Δ%	2555	Δ%	2556	Δ%
กรุงเทพ	26,861,095	10.43	28,867,346	7.47	30,269,692	4.86	31,988,047	6
ภาคกลางไม่รวมกทม.	18,166,914	16.71	16,856,616	-7.21	21,985,621	30.43	22,763,319	4
ภาคตะวันตก	14,910,305	16.53	15,956,118	7.01	17,441,224	9.31	18,688,000	7
ภาคตะวันออก	13,651,875	12.82	14,496,617	6.19	16,028,776	10.57	17,504,675	9
ภาคตะวันออกเฉียงเหนือ	22,163,584	3.95	25,850,236	16.63	26,977,750	4.36	29,947,625	11
ภาคเหนือ	15,197,718	9.29	15,514,321	2.08	19,130,690	23.31	21,115,135	10
ภาคใต้	11,570,623	4.96	15,636,474	35.14	14,449,314	-7.59	19,717,887	36
รวม	122,522,114	10.34	133,177,728	8.7	146,283,067	9.84	161,724,688	11

ตารางที่ 1 จำนวนนักท่องเที่ยวชาวไทย ที่เดินทางท่องเที่ยวภายในประเทศ ระหว่างปี 2553-2556 [1]

2. งานวิจัยที่เกี่ยวข้อง

สำนักสถิติเศรษฐกิจและสังคม สำนักงานสถิติแห่งชาติ เรื่อง "การสำรวจพฤติกรรมการเดินทางท่องเที่ยวของชาวไทย พ.ศ. 2557" พบว่า ในกลุ่มผู้ที่มีระดับการศึกษาสูงจะมีการเดินทางท่องเที่ยวสูงกว่าไม่เดินทาง คือ กลุ่มปริญญาตรีและสูงกว่าปริญญาตรี มีผู้เดินทางท่องเที่ยว ร้อยละ 87.9 ส่วนสถานภาพการทำงานของผู้ที่เดินทางท่องเที่ยว พบว่า กลุ่ม นักเรียน/นักศึกษา เป็นผู้เดินทางท่องเที่ยวร้อยละ 79.3 [3]

นฤพนธ์ พนาวงศ์ และ จักรกฤษณ์ เสน่ห์ นมะหุด ได้ นำเสนอผลการออกแบบระบบแนะนำข้อมูลท่องเที่ยวในประเทศไทย โดยใช้หลักการประเมินค่านำหนักคะแนนแบบหลายปัจจัยและหลายมิติมีวัตถุประสงค์เพื่อใช้แนะนำข้อมูลท่องเที่ยวที่ได้รับความนิยม อีกทั้งยังแนะนำสถานที่ท่องเที่ยวใกล้เคียง ที่พัก ร้านอาหาร ร้านขายของฝาก ร้านขายสินค้าหนึ่งตำบลหนึ่งผลิตภัณฑ์ ที่ตอบสนองความต้องการของผู้ใช้โดยดูจากข้อมูลที่ใช้คนอื่นได้ให้คะแนนความชอบไว้ ทำให้เกิดความรู้สึกไม่เสียเวลาในการตัดสินใจที่จะไปท่องเที่ยวยังสถานที่นั้น ๆ อีกทั้งยังจะได้ความประทับใจอย่างที่คาดคิดไว้ [4]

กองการท่องเที่ยว สำนักวัฒนธรรม กีฬา และการท่องเที่ยว เรื่อง "แผนพัฒนาการท่องเที่ยว กรุงเทพมหานคร รายงานฉบับสมบูรณ์" จากการสำรวจพบว่า พฤติกรรมการท่องเที่ยวในกรุงเทพมหานครของนักท่องเที่ยวชาวไทย พบว่า นักท่องเที่ยวส่วนใหญ่รับทราบข้อมูลการท่องเที่ยวกรุงเทพมหานครจากการบอกต่อของเพื่อนและคนรู้จักร้อยละ 28.9 รองลงมาคือจากวิทยุ [5]

Yongheng Zhao and Yanxia Zhang [6] ได้นำเสนอการเปรียบเทียบการทำงานของอัลกอริทึมของต้นไม้ตัดสินใจในการหาวัตถุบนท้องฟ้าของนักดาราศาสตร์ โดยวิธี REPTree Random Tree Random Forest J48 NBTree และ ADTree ผลที่ได้รับคือ ADTree เป็นวิธีการที่ดีที่สุดหากต้องการความรวดเร็วในการประมวลผล J48 เป็นวิธีการที่ดีที่สุดหากต้องการความถูกต้อง และความ

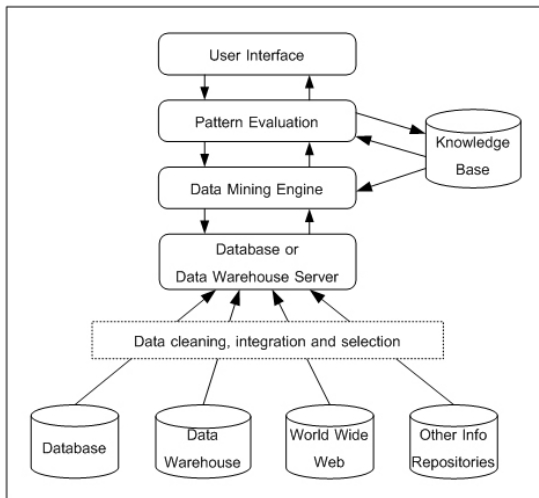
Jehad Ali , Rehanullah Khan , Nasir Ahmad และ Imran Maqsood [7] นำเสนอการเปรียบเทียบระหว่าง Random Forest กับ Decision Tree ผลการจับหมวดหมู่แสดงให้เห็นว่า Random Forest ให้ผลที่ดีกว่าในกรณีของชุดข้อมูลขนาดใหญ่ ในขณะที่ J48 ให้ผลที่ดีกว่าหากข้อมูลมีขนาดเล็ก หรือมีจำนวนข้อมูลไม่มาก โดยในการนี้ได้มีการใช้ข้อมูลที่เพิ่มขึ้นจาก 286 ไปจนถึง 699 ข้อมูล ส่งผลให้จำนวนร้อยละของ Random Forest มีจำนวนเพิ่มขึ้นจาก 69.23% เป็น 96.13%

Sushikumar Kalmegh [8] นำเสนอการเปรียบเทียบการวิเคราะห์อัลกอริทึมในโปรแกรม Weka ได้แก่ REPTree Simple Cart และ Random Tree ในการจำแนกข่าวอินเดียว พบว่าประสิทธิภาพและความแม่นยำของ Random Tree มีมากกว่า REPTree และ Simple Cart

3. ทฤษฎีที่เกี่ยวข้อง

การทำเหมืองข้อมูล (Data Mining)

เหมืองข้อมูล[9] [10] [11] [12] กระบวนการวิเคราะห์ข้อมูลอย่างอัตโนมัติ หรือกึ่งอัตโนมัติ เพื่อแยกประเภท ค้นหารูปแบบ และแนวทางความสัมพันธ์ของข้อมูลจากฐานข้อมูลขนาดใหญ่ หรือคลังข้อมูล ให้อยู่ในรูปแบบของกฎ (Rule) โดยอาศัยหลักสถิติ การรู้จำ การเรียนรู้ของเครื่อง และหลักคณิตศาสตร์ เปรียบเสมือนวิวัฒนาการในการจัดเก็บ ตีความหมายข้อมูล เพื่อนำสารสนเทศที่ได้ไปใช้ในการตัดสินใจ และประยุกต์ใช้กับในงานหลายประเภท ทั้งในด้านธุรกิจ เพื่อประกอบการตัดสินใจของผู้บริหาร ด้านวิทยาศาสตร์ การแพทย์ เป็นต้น



รูปที่ 1 สถาปัตยกรรมเหมืองข้อมูล [10]

แนวคิดด้านการท่องเที่ยว

นักวิชาการหลายท่านได้ให้ความหมายของการท่องเที่ยว (Tourism) [13] [14] [15] โดยสรุปว่า หมายถึงการเดินทางออกจากที่พักเป็นการชั่วคราวเพื่อพักผ่อนหย่อนใจเยี่ยมญาติ หรือวัตถุประสงค์อื่นใดที่ไม่ใช่เพื่อการประกอบอาชีพหรือหารายได้เพื่อให้เกิดความสนุกสนานเพลิดเพลินและผ่อนคลายความเครียดในยามว่างหรือเมื่อเวลาที่ต้องการจะพักผ่อนซึ่งอาจเป็นการท่องเที่ยวที่ให้ความรู้และประสบการณ์จากการเดินทางท่องเที่ยวอีกด้วย นอกจากนี้การท่องเที่ยวยังเป็นการพักผ่อนคลายความตึงเครียด และได้รับความรู้ความเข้าใจวัฒนธรรม และยังเป็นการนำทรัพยากรของประเทศมาใช้ให้เกิดประโยชน์อย่างกว้างขวางโดยความสำคัญของการท่องเที่ยว

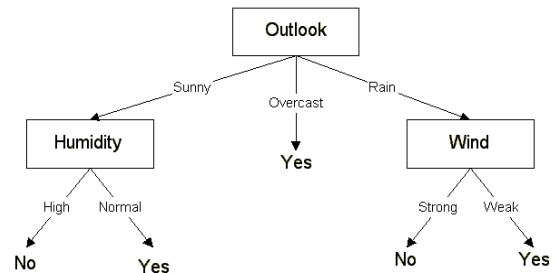
ระบบแนะนำ (Recommendation System)

ระบบแนะนำ [16] [17] [18] [19] [20] เป็นเครื่องมือซอฟต์แวร์ และเทคนิคการให้บริการข้อเสนอแนะสำหรับผู้ใช้งาน โดยข้อเสนอแนะที่เกี่ยวข้องกับการตัดสินใจต่างๆ ในกระบวนการนี้ ส่วนใหญ่จะเป็นการแนะนำในการเลือกซื้อสินค้า เช่น แผ่นซีดีเพลง หนังสือออนไลน์ เป็นต้น และการออกแบบระบบจะใช้กราฟฟิคในส่วนที่ใช้สำหรับติดต่อกับผู้ใช้งานเป็นหลัก เป็นระบบที่สามารถช่วยคัดกรองข้อมูล วัตถุ ต่างๆจำนวนมาก รวมทั้งวิเคราะห์จุดเด่น จุด

ด้อยของแต่ละทางเลือก ให้ตรงกับความต้องการของผู้ใช้ โดยทั่วไปจะจำแนกได้ 3 วิธี คือ 1. การแนะนำแบบ Content-Base คือการแนะนำที่วิเคราะห์จากพื้นฐานความชอบของผู้ใช้คนอื่นๆ ในอดีตที่ผ่านมา เช่น เว็บไซต์หนึ่ง แนะนำบทความที่มีเนื้อหา หรืออยู่ในกลุ่มประเภทบทความเดียวกัน ให้กับผู้อ่าน ร้านขายแผ่นภาพยนตร์ออนไลน์ เมื่อผู้ซื้อเลือกซื้อภาพยนตร์แอคชั่น ก็จะแนะนำภาพยนตร์แอคชั่นเรื่องอื่นให้กับผู้ซื้อเช่นกัน 2. การแนะนำแบบ Collaborative ที่จะวิเคราะห์ความชอบของผู้ใช้อื่นๆ ที่มีลักษณะ พฤติกรรมความชอบที่คล้ายคลึงกัน มาแนะนำ หรือ ทำนายให้กับผู้ใช้ปัจจุบัน เช่น เมื่อเข้าเว็บไซต์จำหน่ายสินค้าออนไลน์ หลังจากที่กดเลือกซื้อสินค้า A จะมีแถบโฆษณาสินค้าตัวอื่นปรากฏขึ้น พร้อมข้อความ เช่น ผู้ซื้อท่านอื่นที่ซื้อ A นี้ เลือกซื้อสินค้า B ด้วย เป็นต้น และ 3. การแนะนำแบบ Hybrid เป็นการผสมผสานการแนะนำทั้ง Content-Base และ Collaborative เข้าด้วยกัน โดยการนำ Collaborative เพิ่มเข้าไปในหลักการของ Content-Base

การตัดสินใจแบบกิ่งก้านสาขา (Decision Tree)

การตัดสินใจแบบกิ่งก้านสาขา หรือ ต้นไม้ตัดสินใจ (decision tree) [21] [22] [23] ในการบริหารธุรกิจ เป็นแผนผังต้นไม้ที่ช่วยในการตัดสินใจ โดยแสดงถึงมูลค่าของทรัพยากรที่จะใช้ ความเสี่ยงในการลงทุนและและผลลัพธ์ที่มีโอกาสเกิดขึ้น ต้นไม้ตัดสินใจสร้างขึ้นเพื่อช่วยการตัดสินใจเพื่อใช้ในการสร้างแผนงาน นิยมใช้มากในการบริหารความเสี่ยง (risk management) ต้นไม้ตัดสินใจ เป็นส่วนหนึ่งของทฤษฎีการตัดสินใจ (decision theory) และ ทฤษฎีกราฟภายในต้นไม้จะประกอบไปด้วยโหนด (node) ซึ่งแต่ละโหนดจะมีคุณลักษณะ (attribute) เป็นตัวทดสอบ กิ่งของต้นไม้ (branch) แสดงถึงค่าที่เป็นไปได้ของคุณลักษณะที่ถูกเลือกทดสอบ และใบ (leaf) ซึ่งเป็นสิ่งที่อยู่ล่างสุดของต้นไม้ตัดสินใจแสดงถึงกลุ่มของข้อมูล (class) หรือนั่นก็คือผลลัพธ์ที่ได้จากการทำนาย โหนดที่อยู่บนสุดของต้นไม้เรียกว่าโหนดราก (root node)



รูปที่ 3. ตัวอย่างแสดงต้นไม้ตัดสินใจ [24]

จะเห็นว่าทุกครั้ง หลังจุดตัดสินใจ (เครื่องหมายสี่เหลี่ยม) จะเป็นทางเลือกต่างๆ ในขณะที่หลังเครื่องหมายวงกลม จะเป็นสภาวะการณ์ต่างๆที่เกิดขึ้น ข้อสังเกตบางประการในการเขียนแผนการตัดสินใจ

อัลกอริทึม J48

J48 [6] เป็นอัลกอริทึมในการสร้างต้นไม้ตัดสินใจ ที่ใช้หลักการของทฤษฎีข่าวสาร ค่าที่วัดได้จะนำมาใช้ในการตัดสินใจว่าจะใช้ตัวแปรใดในการแบ่งข้อมูล มีการใช้ค่าเกน (Gain) และค่าคาดคะเนของข้อมูล (Entropy)

$$I(p,n) = -\frac{p}{p+n} \log_2 \left(\frac{p}{p+n} \right) - \frac{n}{p+n} \log_2 \left(\frac{n}{p+n} \right) \quad (1)$$

ค่าคาดคะเนของข้อมูล (Entropy) ที่แยกโดยการใช้ลักษณะประจำ A ซึ่งกำหนด A คือลักษณะประจำที่แบ่ง S ออกเป็น {S1, S2, ..., Sv} โดยให้ S1 มีตัวอย่างจากคลาส P จำนวน p1 และตัวอย่างจากคลาส N จำนวน n1 ดังสมการ (2)

$$E(A) = \sum_{i=1}^v \frac{p_i+n_i}{p+n} I(p_i, n_i) \quad (2)$$

ดังนั้นค่าเกนข้อมูล (Data Gain) ที่ได้จากการแยกข้อมูลด้วย ลักษณะประจำ A จะได้ดังสมการ (3)

$$\text{Gain}(A) = I(p,n) - E(A) \quad (3)$$

เทคนิค J48 เป็นวิธีการที่พัฒนามาจาก ID3 ซึ่ง J48 มีความสามารถเพิ่มเติมขึ้นจาก ID3 ดังนี้

1. ขนาดของต้นไม้ไม่ใหญ่เกินไป
2. มีการสร้างกฎหลังจากที่ทำการปรับแต่งต้นไม้แล้ว
3. สามารถทำการปรับแต่งต้นไม้ประกอบการตัดสินใจ (Pruning Trees) ในขณะที่สร้างได้ เพื่อลดความผิดพลาด
4. สามารถใช้กับข้อมูลที่มีความต่อเนื่องที่เป็นตัวเลขได้
5. สามารถใช้กับชุดข้อมูลที่มีค่าผิดพลาดได้ (Missing Attribute)

เทคนิค REPTree

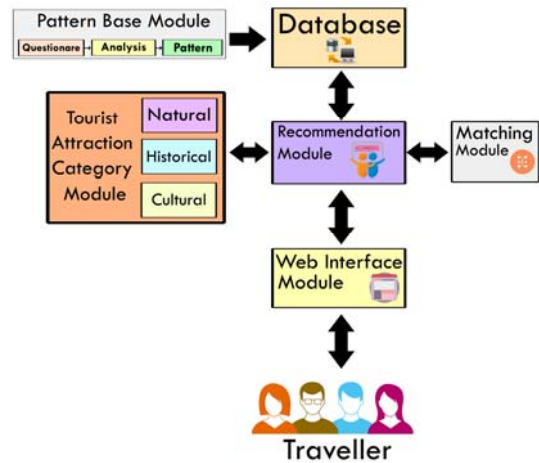
REPTree [8] เป็นเทคนิคที่มีความรวดเร็วในการดำเนินการ จากการสร้างต้นไม้ตัดสินใจด้วยหลักการคำนวณค่าเกน (Gain) และค่าคาดคะเนของข้อมูล แล้วทำการตัดทอนข้อผิดพลาดของข้อมูลระหว่างที่มีการสร้างต้นไม้ แต่อัลกอริทึมนี้สามารถใช้ได้กับข้อมูลจำพวกตัวเลขเท่านั้น

เทคนิค Simple Cart

Simple Cart [8] เป็นเทคนิคในการจัดหมวดหมู่ เพื่อสร้างต้นไม้ตัดสินใจแบบไบนารี คือ มีการสร้างโหนดลูกออกมาเพียง 2 โหนด และใช้ค่าคาดคะเน ในการแบ่งแอทริบิวต์ เทคนิค Simple Cart นี้เหมาะกับการตัดข้อมูลที่ขาดหายไปของแต่ละ record และเป็นเทคนิคที่ดีที่สุดที่แนะนำในการใช้ Training Data

4. กรอบแนวความคิดในการวิจัย

จากแนวโน้มของสถานที่ท่องเที่ยวที่มีเพิ่มมากขึ้น ส่งผลให้นักท่องเที่ยวมีตัวเลือกที่หลากหลายในการตัดสินใจเลือกสถานที่ท่องเที่ยว เพื่อช่วยให้นักท่องเที่ยวสามารถเลือกสถานที่ท่องเที่ยวที่ตรงกับความต้องการของตนเองได้อย่างรวดเร็วยิ่งขึ้น ทางผู้วิจัยจึงได้พัฒนาระบบแนะนำสถานที่ท่องเที่ยวขึ้น จากการวิเคราะห์รูปแบบความสัมพันธ์ของปัจจัยต่างๆ ที่ส่งผลต่อพฤติกรรม ความชื่นชอบ ในเรื่องของการท่องเที่ยว โดยได้นำเสนอกรอบแนวคิดดังรูปที่ 4



รูปที่ 4 กรอบแนวความคิดในการวิจัย

จากกรอบแนวคิดในรูปที่ 4 สามารถอธิบายการทำงานของระบบแนะนำสถานที่ท่องเที่ยว ออกได้เป็น 5 ระบบย่อยดังนี้

ระบบย่อยที่ 1 การศึกษารูปแบบ (Pattern Base Module) เป็นขั้นตอนที่เริ่มจากการศึกษาปัจจัยที่มีผลต่อการเลือกสถานที่ท่องเที่ยว ของนักท่องเที่ยว เช่น เพศ, อายุ, อาชีพ, รายได้ เป็นต้น เมื่อได้ปัจจัยที่ส่งผลต่อการเลือกสถานที่ท่องเที่ยวแล้ว จะนำมาสร้างเป็นแบบสอบถาม เพื่อทำการเก็บข้อมูลจากกลุ่มตัวอย่าง ได้แก่ นักศึกษาระดับปริญญาตรี ของสถาบันเทคโนโลยีไทย-ญี่ปุ่น แล้วนำมาวิเคราะห์ผ่านโปรแกรมการวิเคราะห์รูปแบบความสัมพันธ์ ได้มาซึ่ง รูปแบบพื้นฐาน (Pattern Base) หลังจากนั้น จึงนำข้อมูลจากการวิเคราะห์ที่ได้ เก็บไว้ในฐานข้อมูล

ระบบย่อยที่ 2 การจับคู่ความสัมพันธ์ (Matching Module) เป็นขั้นตอนหลังจากที่ได้ รูปแบบพื้นฐาน และเก็บข้อมูลไว้ในฐานข้อมูลเรียบร้อยแล้ว การจับคู่ความสัมพันธ์นี้จะนำรูปแบบที่ได้จากการวิเคราะห์ มาจับคู่กับข้อมูลที่ได้รับจากนักท่องเที่ยวที่ใช้งานระบบ ผ่านเว็บอินเทอร์เฟซ และ นำข้อมูลที่ได้ไปทำการวิเคราะห์ประมวลผลข้อมูล เพื่อให้ได้รับผลลัพธ์การแนะนำสถานที่ท่องเที่ยว

ระบบย่อยที่ 3 ระบบแนะนำ (Recommendation Module) เป็นส่วนของการประมวลผล เพื่อคำแนะนำในการเลือกสถานที่ท่องเที่ยวให้กับผู้ใช้ระบบ โดยการนำข้อมูลที่ได้จากการกรอกข้อมูลของผู้ใช้ระบบ และ รูปแบบพื้นฐานที่ได้จากการทำเหมืองข้อมูล มาประมวลผลและวิเคราะห์ความสัมพันธ์ให้ได้รูปแบบที่เหมาะสมที่สุด

และนำผลที่ได้ไปใช้ในการให้คำแนะนำแก่ผู้ใช้ระบบต่อไป ผ่านทางหน้าเว็บอินเทอร์เฟซของระบบ

ระบบย่อยที่ 4 ประเภทสถานที่ท่องเที่ยว (Tourist Attraction Category Module) เป็นขั้นตอนการสังเคราะห์ แบ่งประเภทของสถานที่ท่องเที่ยวในประเทศไทย เพื่อให้มีความครอบคลุมถึงสถานที่ท่องเที่ยวทุกประเภท ในการนำข้อมูลไปทำเหมืองข้อมูล เพื่อการวิเคราะห์ความสามารถและนำไปใช้ในการให้คำแนะนำสถานที่ท่องเที่ยวได้ในงานวิจัยนี้ได้ทำการแบ่งประเภทสถานที่ท่องเที่ยวออกเป็น 3 ประเภท ได้แก่ ประเภทธรรมชาติ , ประเภทประวัติศาสตร์ และ ประเภทศิลปวัฒนธรรม

ระบบย่อยที่ 5 เว็บอินเทอร์เฟซ (Web Interface Module) เป็นหน้าจอบริบท สำหรับรับข้อมูลจากผู้ใช้ที่เป็นนักท่องเที่ยวกรอกข้อมูล แล้วทำการส่งข้อมูลเพื่อนำมาเปรียบเทียบกับรูปแบบพื้นฐานที่ได้จากวิเคราะห์ปัจจัยที่ส่งผลต่อการเลือกสถานที่ท่องเที่ยวภายในฐานข้อมูลที่จัดเก็บไว้ และเว็บเพจนี้ทำหน้าที่แสดงผลลัพธ์ให้กับผู้ใช้ได้รับทราบการประมวลผลของระบบ

5. การสร้างรูปแบบพื้นฐาน

งานวิจัยนี้ได้นำโปรแกรมสำเร็จรูป Weka เข้ามาใช้เพื่อหาความสัมพันธ์ของข้อมูลดังกล่าว เมื่อการวิเคราะห์ข้อมูลเสร็จสิ้นผลลัพธ์ที่ได้ออกมาคือ รูปแบบความสัมพันธ์กันของปัจจัย ตัวแปรต่างๆที่ส่งผลต่อการเลือกสถานที่ท่องเที่ยว ทำให้ได้กฎพื้นฐาน (Pattern Base) ที่จะนำไปใช้ในการวิเคราะห์ออกมา ด้วยการใช้อัลกอริทึมแบบกิ่งก้านสาขา (Decision Tree)

เมื่อได้ผลการวิเคราะห์รูปแบบการแนะนำสถานที่ท่องเที่ยวแล้ว ขั้นตอนนี้เป็นกรณำรูปแบบ (Pattern) การแนะนำมาทำการแปลงให้อยู่ในรูปแบบที่สามารถนำไปพัฒนาเป็นระบบแนะนำ เพื่อประเมินผลได้ โดยเมื่อนำมาแปลงเป็นรูปแบบกฎพื้นฐานแล้วได้ ผลลัพธ์ของกฎพื้นฐานที่ได้รับจำแนกตามตัวแปรตามที่กำหนดไว้แต่ละตัว มีจำนวนดังนี้ สถานที่ท่องเที่ยวเชิงธรรมชาติ มีจำนวน 21 กฎ สถานที่ท่องเที่ยวเชิงประวัติศาสตร์ มีจำนวน 13 กฎ สถานที่ท่องเที่ยวเชิงศิลปวัฒนธรรม มีจำนวน 5 กฎ รวมจำนวนกฎพื้นฐานทั้งสิ้น 39 กฎพื้นฐาน ดังตารางที่ 2 ตัวอย่างกฎพื้นฐาน

ตารางที่ 2 ตัวอย่างกฎพื้นฐาน

ลำดับ	กฎพื้นฐาน	ประเภท
1	If Q6>4 and Q3>4 and Q13>4	ศิลปวัฒนธรรม
2	If Q6>4 and Q3>4 and Q13<=4 and Q11>3	ธรรมชาติ

6. การประเมินผล

ผลการประเมินของผู้เชี่ยวชาญที่มีต่อความเหมาะสมของกรอบแนวคิดการออกแบบระบบแนะนำสถานที่ท่องเที่ยว สำหรับนักท่องเที่ยวในแต่ละรายการ พบว่า ความเหมาะสมของภาพรวม

โมเดลที่สังเคราะห์ขึ้น มีค่าเฉลี่ยเท่ากับ 4.67 ส่วนเบี่ยงเบนมาตรฐาน 0.58 รายการที่มีระดับความคิดเห็นในระดับที่เหมาะสมมากที่สุด ได้แก่ ความเหมาะสมของ Pattern Base Module ความเหมาะสมของ Web Interface Module ซึ่งมีค่าเฉลี่ยเท่ากับ 4.67 ส่วนเบี่ยงเบนมาตรฐาน 0.58 เท่ากันทั้ง 2 Module ดังตารางที่ 3

โดยการประเมินใช้เกณฑ์คะแนนประเมินความเหมาะสมของกรอบแนวคิดดังนี้ [24]

4.50-5.00 : หมายถึง มีความเหมาะสมมากที่สุด

3.50-4.49 : หมายถึง มีความเหมาะสมมาก

2.50-3.49 : หมายถึง มีความเหมาะสมปานกลาง

1.50-2.49 : หมายถึง มีความเหมาะสมน้อย

1.00-1.49 : หมายถึง มีความเหมาะสมน้อยที่สุด

ตารางที่ 3 ผลการประเมินความเหมาะสมของกรอบแนวคิด

รายละเอียดการประเมิน	ค่าเฉลี่ย	S.D.	ระดับ
1. ความเหมาะสมของ Pattern Base Module	4.67	0.58	มากที่สุด
2. ความเหมาะสมของ Matching Module	4.33	0.58	มาก
3. ความเหมาะสมของ Recommendation Module	4.33	1.15	มาก
4. ความเหมาะสมของ Tourist Attraction Category Module	3.33	0.58	ปานกลาง
5. ความเหมาะสมของ Web Interface Module	4.67	0.58	มากที่สุด
6. ความเหมาะสมของภาพรวมโมเดลที่สังเคราะห์ขึ้น	4.67	0.58	มากที่สุด
รวม	4.33	0.68	มาก

งานวิจัยนี้ใช้อัลกอริทึมในการวิเคราะห์รูปแบบการแนะนำสถานที่ท่องเที่ยวทั้งสิ้น 3 อัลกอริทึม เพื่อเปรียบเทียบความแม่นยำในการแนะนำนักท่องเที่ยว ด้วยโปรแกรม Weka 3.6 ซึ่งประกอบไปด้วยอัลกอริทึมดังตารางที่ 4

ตารางที่ 4 ผลการเปรียบเทียบความแม่นยำในการแนะนำ

ลำดับที่	อัลกอริทึม	% ความแม่นยำในการแนะนำ
1	J48	84.73
2	Simple Cart	83.84
3	REPTree	82.77

จะเห็นว่าผลการวิเคราะห์ข้อมูลโดยใช้อัลกอริทึมแบบ J48 มีความถูกต้องแม่นยำในการแนะนำสูงที่สุด เท่ากับ 84.73% อัลกอริทึมที่มีความแม่นยำในการทำนายรองลงมาได้แก่ Simple Cart และ REPTree ซึ่งมีความแม่นยำในการทำนายเท่ากับ 83.84% และ 82.77% ตามลำดับ

7. สรุป

งานวิจัยนี้ได้นำเสนอแนวคิดในการออกแบบระบบแนะนำการเลือกสถานที่ท่องเที่ยว สำหรับนักท่องเที่ยว ด้วยการทำเหมืองข้อมูล โดยผลของการแนะนำพบว่า อัลกอริทึม J48 มีความแม่นยำในการแนะนำมากที่สุด โดยมีเปอร์เซ็นต์ความแม่นยำเท่ากับ 84.73% ระบบแนะนำนี้ขึ้นเพื่อช่วยในการเลือกสถานที่ท่องเที่ยว และให้คำแนะนำกับนักท่องเที่ยวที่มีความต้องการในการท่องเที่ยวภายในประเทศ ให้สามารถตัดสินใจเลือกสถานที่ท่องเที่ยวในแบบที่ตนเองต้องการได้

ข้อเสนอแนะ ควรมีการเก็บข้อมูลจากกลุ่มตัวอย่างที่มีจำนวนมาก และแพร่หลายตรงกับเรื่องที่ต้องการแนะนำ เพื่อความถูกต้องแม่นยำของระบบแนะนำ

เอกสารอ้างอิง

- [1] ศูนย์วิจัยด้านตลาดการท่องเที่ยว (TAT Intelligence Center) การท่องเที่ยวแห่งประเทศไทย. Retrieved January 20, 2015, from <http://marketingdatabase.tat.or.th/>
- [2] ระบบฐานข้อมูลแหล่งท่องเที่ยว กระทรวงการท่องเที่ยวและกีฬา. Retrieved January 23, 2015, <http://61.19.236.136:8090/dotr/>
- [3] สำนักสถิติเศรษฐกิจและสังคม สำนักงานสถิติแห่งชาติ. "การสำรวจพฤติกรรมการท่องเที่ยวของชาวไทย พ.ศ.2557". หน้า 6
- [4] นฤพนธ์ พินางค์ และ จักรกฤษณ์ เสน่ห์ นมะหุด. "การออกแบบระบบแนะนำข้อมูลท่องเที่ยวในประเทศไทย โดยใช้หลักการประเมินค่าน้ำหนักคะแนนแบบปัจจัย และหลายมิติ". การประชุมสัมมนาวิชาการนำเสนอผลงานวิจัยระดับชาติ ครั้งที่ 10. หน้า 259-271
- [5] กองการท่องเที่ยว สำนักวัฒนธรรม กีฬา และการท่องเที่ยว กรุงเทพมหานคร. รายงานฉบับสมบูรณ์ แผนการพัฒนาการท่องเที่ยว กรุงเทพมหานคร ปี 2555-2558. หน้า 52
- [6] Yongheng Zhao and Yanxia Zhang. (2007). Comparison of decision tree methods for finding active objects. *Advances of Space Research*.
- [7] Sushikumar Kalmegh. (2011). Analysis of WEKA Data Mining Algorithm REPTree, Simple Cart and RandomTree for Classification of Indian News. *International Journal of Innovative Science, Engineering & Technology*, Vol. 2 Issue 2.
- [8] Jehad Ali et al. (2012). Random Forests and Decision Trees. *IJCSI International Journal of Computer Science Issues*, Vol. 9.
- [9] Oracle® Data Mining Concepts. (n.d.). Retrieved January 31, 2015, from http://docs.oracle.com/cd/B28359_01/datamine.111/b28129/proc_ess.htm
- [10] ความหมายและความสำคัญของ ฐานข้อมูล - Data Mining. (n.d.). Retrieved March 20, 2015, from <https://sites.google.com/a/bumail.net/data-mining/khwam-hmay>
- [11] มารูจิก Data Mining กันเถอะ (Introduction to Data Mining) | แหล่งรวบรวมความรู้ทางด้าน data mining. (n.d.). Retrieved March 20, 2015, from <http://open-miner.com/2009/11/03/introduction-datamining/>
- [12] อุดลย์ ยี่มงาม. (2014, September 30). การทำเหมืองข้อมูล (Data Mining) [Education]. Retrieved from http://compcenter.bu.ac.th/index.php?option=com_content&task=view&id=75&Itemid=172
- [13] วิชาเทคโนโลยีสารสนเทศ HQ102 คณะมนุษยศาสตร์ สาขาการท่องเที่ยว มหาวิทยาลัยหอการค้าไทย. Retrieved from <http://peenlonie.blogspot.com/p/blog-page.html>
- [14] บทที่ 2 การทบทวนแนวคิดและทฤษฎีเกี่ยวกับการท่องเที่ยวเชิงนิเวศ. Retrieved April 27, 2015, from <http://masterplan.andamanecotourism.com>
- [15] บทที่ 2-ทบทวนแนวคิดทฤษฎี และเอกสารงานวิจัยที่เกี่ยวข้อง. (n.d.). Retrieved April 27, 2015, from <http://chiangkhan.com>
- [16] Recommend System - InMymind. (n.d.). Retrieved May 27, 2015, from <https://sites.google.com/site/kanna7332/project/recommend-system>
- [17] Recommender system. In Wikipedia, the free encyclopedia. Retrieved May 27, 2015, from https://en.wikipedia.org/w/index.php?title=Recommender_system&oldid=668739243
- [18] การใช้ระบบแนะนำสนับสนุนการตัดสินใจ, นลินี โสพิศสถิตย์ ปีงบประมาณ 2555 มหาวิทยาลัยราชภัฏสวนสุนันทา คณะวิทยาศาสตร์และเทคโนโลยี
- [19] Maarten Devillé , Peter Vrancx. VUB ARTIFICIAL INTELLIGENCE LAB Retrieved June 27, 2015, from <http://ai.vub.ac.be/node/1152>
- [20] Daniel Rodriguez. (2013). Recommender Systems. University of Alcalá.
- [21] Puidui. System Analysis Techniques and Design for Business: ต้นไม้การตัดสินใจ ช่วยอ่านด้วยนะ. Retrieved May 20, 2015, from http://sasdkmitl09.blogspot.com/2009/07/blog-post_23.html
- [22] sit. (n.d.). ขั้นตอนการสร้างโมเดล Decision Tree. Retrieved June 20, 2015, from <http://dataminingtrend.com/2014/decision-tree-model/>
- [23] ชัดชัย แก้วตา และ อัจฉรา มหาวีร์วัฒน์. การวินิจฉัยคดีด้วยเทคนิคต้นไม้ตัดสินใจ. NCIT2010
- [24] Thomas Debray. (2009). Classification in Class Imbalanced Datasets. Maastricht University.
- [25] Susan Jamieson. Likert Scales: how to (ab)use them. *Medical Education* vol. 38. Issue 12. pages 1217-1218. 2004.