DE VINCI RESEARCH CENTER

Tourist's Tour **Prediction** by **Sequential** Data Mining Approach

大连东软信息学院 Dalian Neusoft University of Information | AOMA

Lilia BEN BACCAR, Sonia DJEBALI, Guillaume GUERARD
sonia.djebali@devinci.fr.

21 -23 November 2019

smart - - grid.net

Tourism

1,2 Md tourism in 2017

10% part of GDP

Instagram

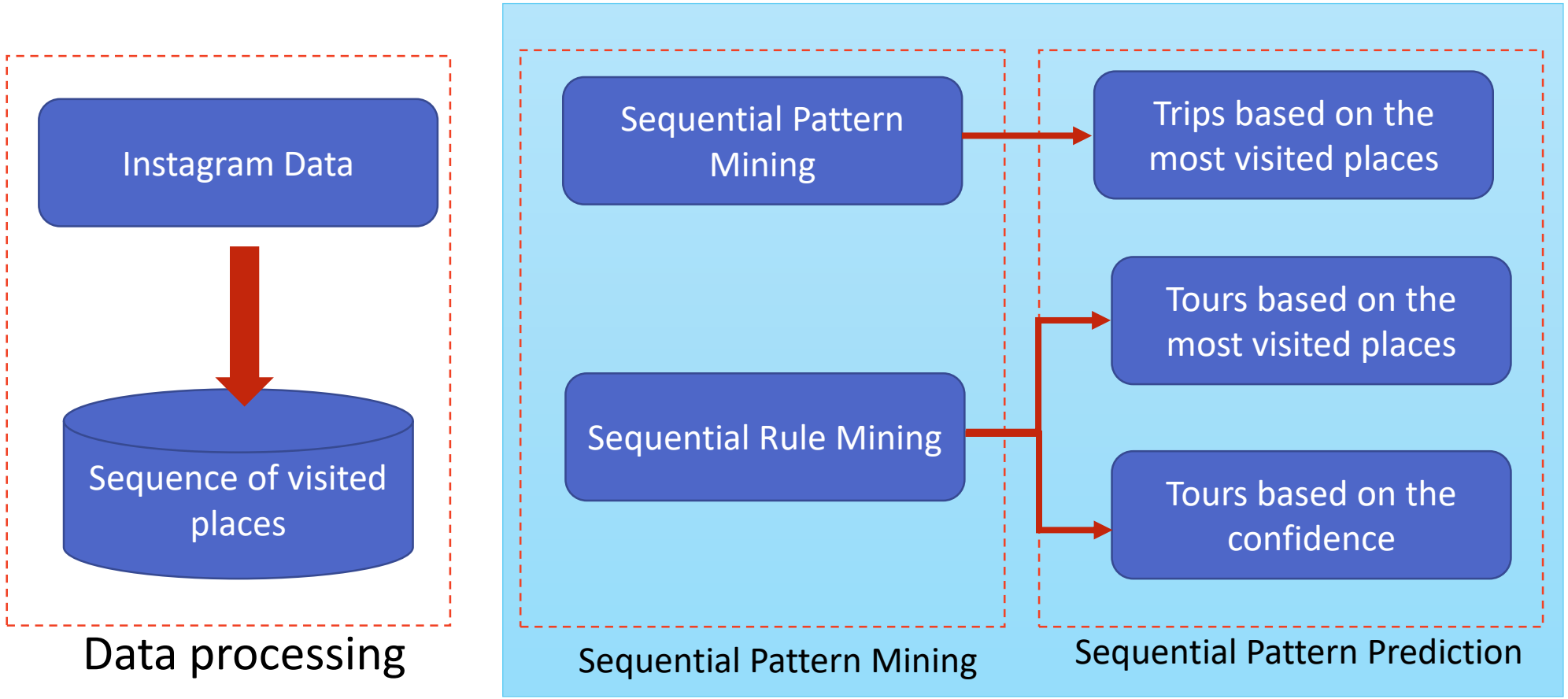
1Md regular active users

500M daily users

*Based on geo-located and time-related information of photographs on Instagram, we propose in this paper an original approach to **determine and to predict** behaviors of tourists by analyzing sequences of places visited during a trip by each tourist.*

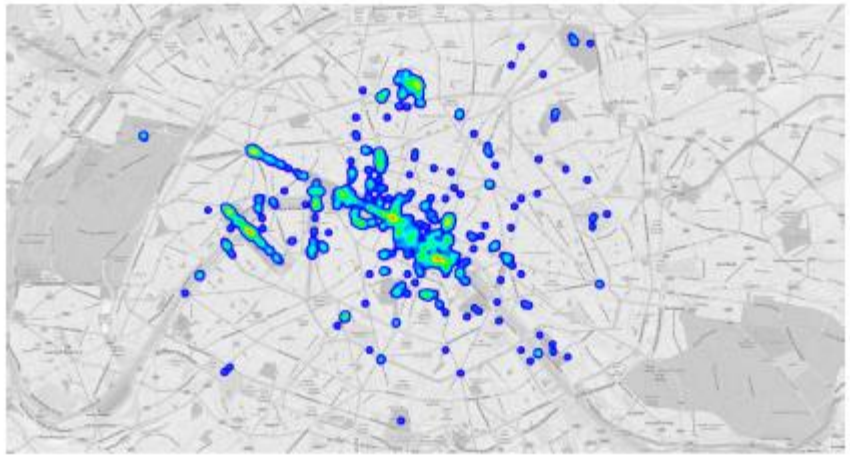
sonia.djebali@devinci.fr

Our approach



Data Processing

Tourist's photos of Paris famous places



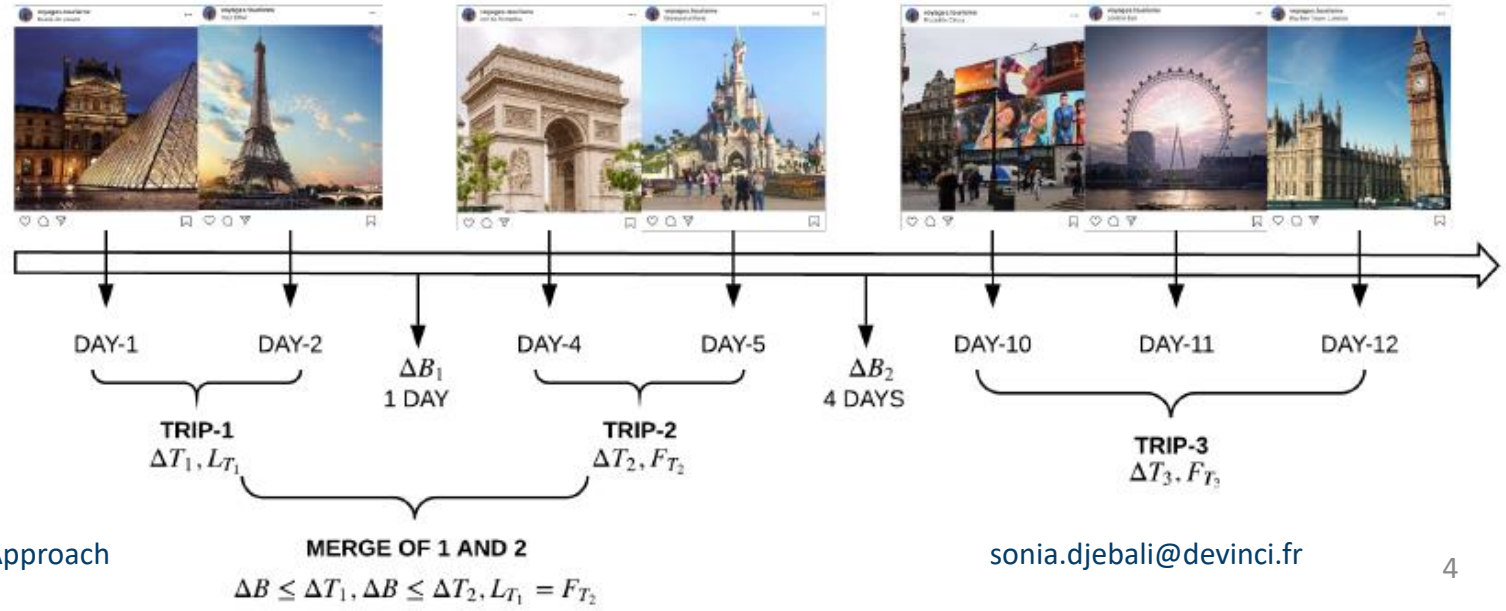
Country	USA	UK	Italy	Russia	Brazil	Spain	Australia
Trips	3 064	1 423	923	742	606	464	330
Average trip duration (days)	≈ 2.64	≈ 2.32	≈ 2.65	≈ 2.91	≈ 2.85	≈ 2.56	≈ 2.77
Average number of photos / trip	≈ 3.76	≈ 3.86	≈ 4.30	≈ 4.87	≈ 4.02	≈ 4.03	≈ 3.70

Sequences

A trip is a succession of days when a non-resident tourist takes at least one photo per day

$$\Delta B \leqslant \Delta T_i \text{ and } \Delta B \leqslant \Delta T_j$$

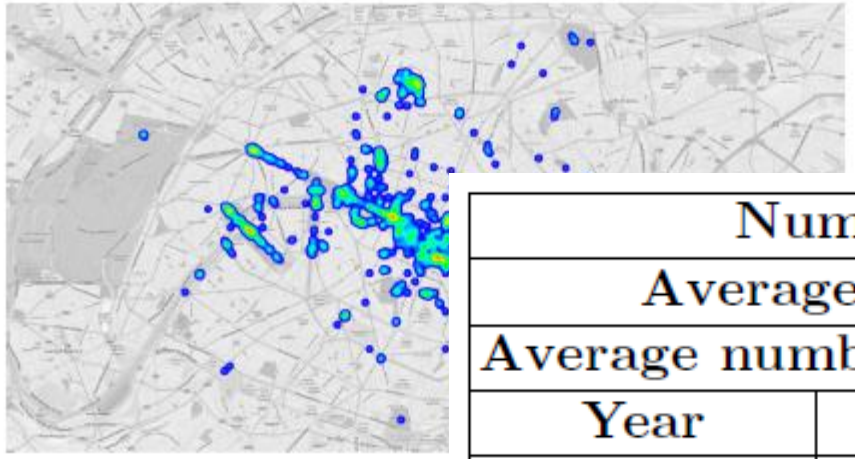
$$\text{and } L_{T_i} = F_{T_j}$$



sonia.djebali@devinci.fr

Data Processing

Tourist's photos of Paris famous places



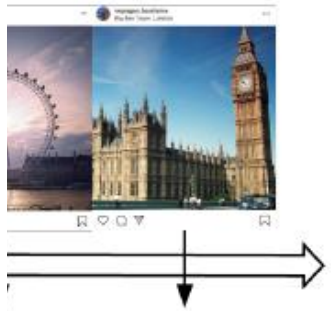
Number of trips		742 trips
Average length of trips		≈ 3 days
Average number of photos per trip		≈ 5 photos
Year	Number of trips in	Best month
2011	1 trip	September (1 trip)
2012	50 trips	September (10 trips)
2013	114 trips	January (15 trips)
2014	253 trips	October (36 trips)
2015	324 trips	September (35 trips)

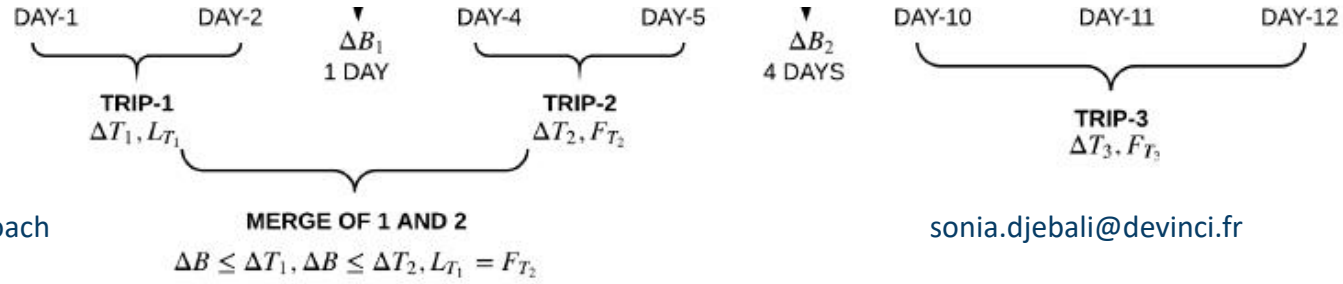
Sequences

A trip is a succession of da
non-resident tourist take
one photo per day

$$\Delta B \leqslant \Delta T_i \text{ and } \Delta L \leqslant \dots$$

$$\text{and } L_{T_i} = F_{T_j}$$





sonia.djebali@devinci.fr

Data Analysis

Trips based on the *most visited* places

Learn the most visited places, no temporal point of view between those places

→ Based on the support of the places

Tours based on the *most visited* places

Learn paths between famous places, build a tour from those rules

→ Build an automaton whose links represent rules above a rule's support threshold

→ Random walk on this automaton

Tours based on the *most plausible paths between places*

Learn the most plausible paths between famous places, build a tour from those rules

→ Build a stochastic automaton whose links represent rules and their confidence

→ Random walk on this stochastic automaton

Trip based on the most visited places

Sequential Pattern Mining constraint-based algorithm: PrefixSpan

[Pei et alii 2004]

Learning phase

Results : minsup = 5%	
Patterns	*Support*
< Tour Eiffel >	30.66%
< Musée du Louvre >	25.83%
< Cathédrale Notre-Dame de Paris >	16.16%
< Avenue des Champs Elysées >	14.09%
< Sacré-Coeur >	11.89%
< Centre Pompidou (CNAC) >	8.01%
< Jardin du Luxembourg >	6.63%
< Galeries Lafayette >	6.08%
< Musée du Louvre, Tour Eiffel >	5.39%
< Montmartre >	5.39 %

Predict phase

A russian tourist want to visit the top 5 places of Paris :

- Tour Eiffel
- Musée du Louvre
- Cathédrale Notre-Dame de Paris
- Champs Elysées
- Sacré-Coeur

sonia.djebali@devinci.fr

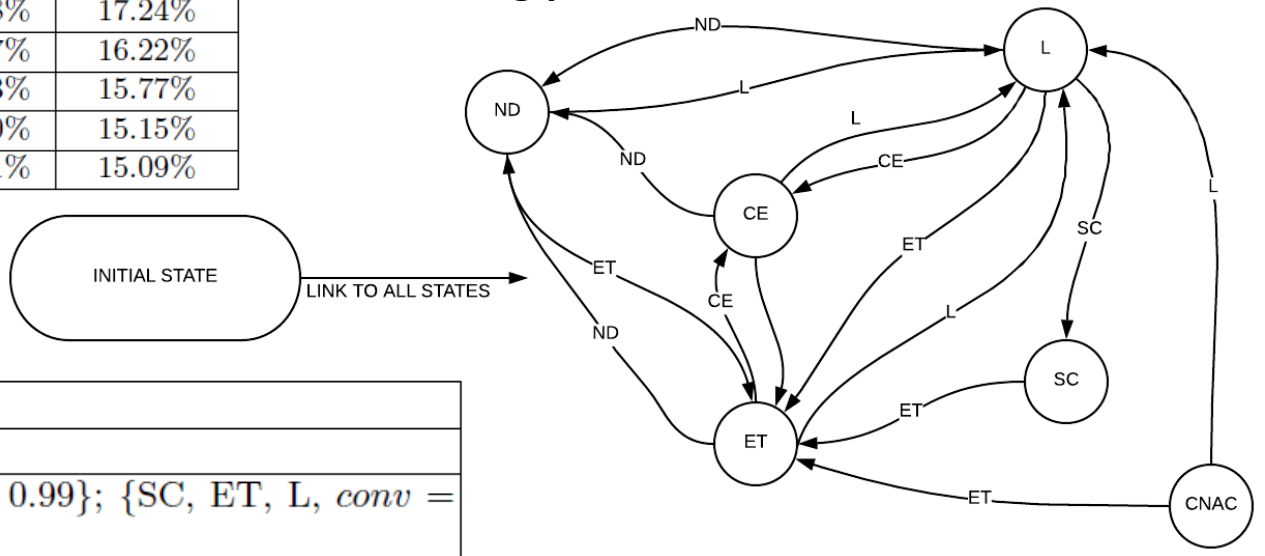
Tours based on the most visited places

Sequential Rule Mining Constraint-Based Algorithm: Rules Growth algorithm

[Viger et alii 2011]

Results : minsup = 1% and minconf = 15%		
Rules	*Support*	*Confidence*
Cathédrale Notre-Dame de Paris ⟶ Musée du Louvre	3.59%	22.22%
Musée du Louvre ⟶ Tour Eiffel	5.39%	20.86%
Cathédrale Notre-Dame de Paris ⟶ Tour Eiffel	3.18%	19.66%
Centre Pompidou (CNAC) ⟶ Musée du Louvre	1.52%	18.97%
Tour Eiffel, Musée du Louvre ⟶ Cathédrale Notre-Dame de Paris	1.80%	18.57%
Centre Pompidou (CNAC) ⟶ Tour Eiffel	1.38%	17.24%
Tour Eiffel ⟶ Musée du Louvre	4.97%	16.22%
Tour Eiffel ⟶ Cathédrale Notre-Dame de Paris	4.83%	15.77%
Sacré-Coeur ⟶ Tour Eiffel	1.80%	15.15%
Tour Eiffel, Cathédrale Notre-Dame de Paris ⟶ Musée du Louvre	1.11%	15.09%

Learning phase



Prediction phase (with conviction)

Length	Sequences
1	{SC, ET, *conv* = 0.82}
2	{SC, ET, ND, *conv* = 1.01}; {SC, ET, CE, *conv* = 0.99}; {SC, ET, L, *conv* = 1.13}
3	{SC, ET, ND, L, *conv* = 1.21}; {SC, ET, CE, ND, *conv* = 1.09}; {SC, ET, CE, L, *conv* = 1.11}; {SC, ET, L, ND, *conv* = 0.97}; {SC, ET, L, CE, *conv* = 0.99}

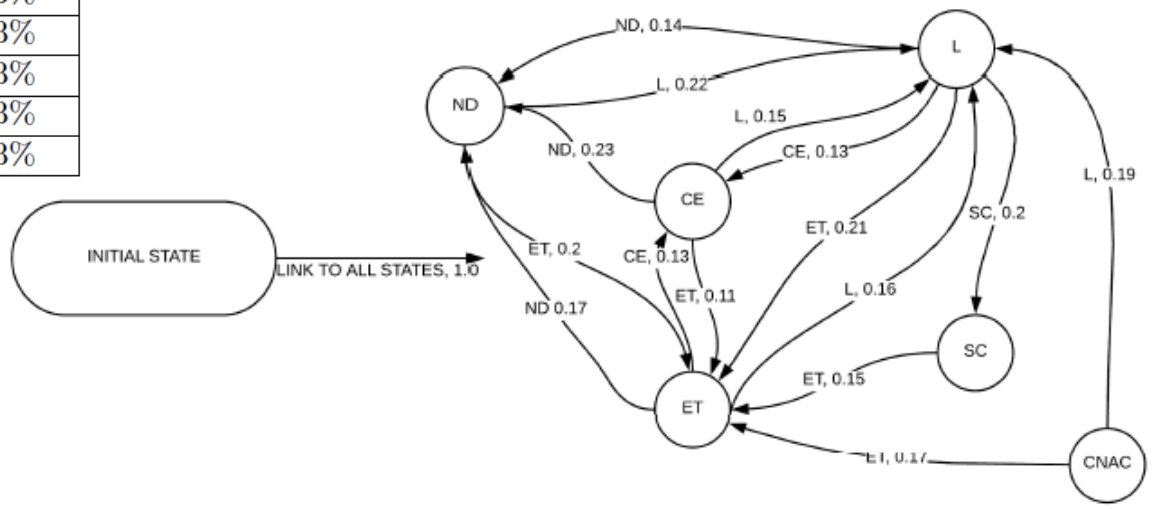
Tours based on the confidence

+ Sequential Rule Mining Preference-Based Algorithm: TNS algorithm, Top-K non-redundant sequential rules

[Viger et alii 2013]

Results : k = 15 and minconf = 33%		
Rules	*Support*	*Confidence*
Galerie Emmanuel Perrotin ⟶ Palais de Tokyo - Musée d'Art Moderne, Louis Vuitton Foundation for Creation	0.41%	100%
Colette ⟶ Tour Eiffel	0.55%	66.66%
Tour Eiffel, Musée du Louvre, Centre Pompidou (CNAC) ⟶ Cathédrale Notre-Dame de Paris	0.69%	41.67%
Musée du Louvre, Galeries Lafayette ⟶ Tour Eiffel	0.55%	36.36%
Ladurée ⟶ Avenue des Champs Elysées	0.97%	33.33%
Musée du Louvre, Centre Pompidou (CNAC) ⟶ Tour Eiffel	0.97%	33.33%
L'Avenue ⟶ Avenue des Champs Elysées	0.41%	33.33%
Trocadéro ⟶ Sacré-Coeur, Jardin du Luxembourg	0.41%	33.33%
L'Avenue ⟶ Avenue Montaigne	0.41%	33.33%
Shangri-La Hotel ⟶ Tour Eiffel	0.41%	33.33%

Learning phase



Prediction phase (with conviction)

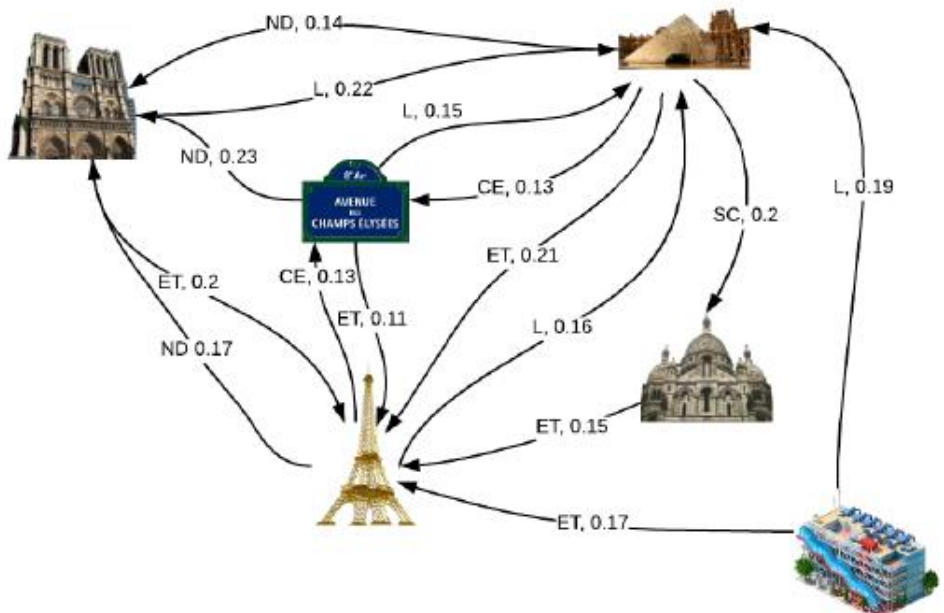
Examples of tours	Confidence of the tour	Latest conviction
{Musée du Louvre, Pont Neuf, Cathédrale Notre-Dame de paris}	13%	7.61
{Saint-Lazare, Galeries Lafayette, Palais Garnier, Musée du Louvre}	11.1%	4.32
{Avenue des Champs Elysées, Arc de Triomphe, Parc Monceau }	9.24%	4.73

Conclusion

Main idea : based on some specific tourists (for example, Russian, no specific states, mid-aged, no specific sex), the proposed method build three kind of trips

- Most visited places (for “lazy” people)
- Most used routes (for “classic” tourists)
- Specific routes (for “atypical” tourists)

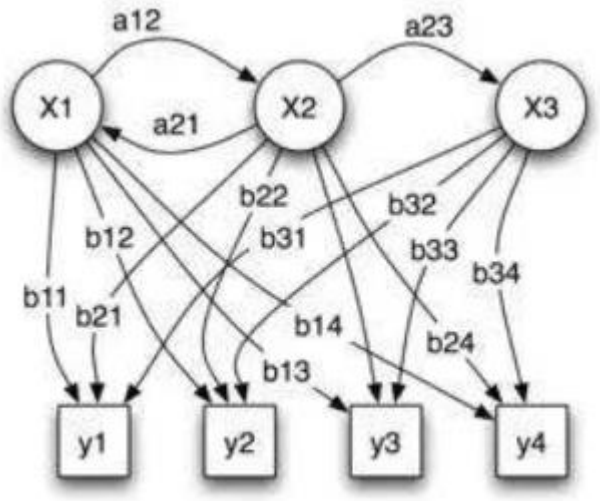
The most probable routes of our methods retrieve the most popular tours (from russian tour operators) and the most popular special/temporary event (among russians).



Future works

Consider a tour as temporal series (Yamata) because Rules respect the order but not the timeline.
Build HMM based on tourists' attributes, i.e. mid-aged Russian from Moscow.

See you again next year ! (maybe at ADMA 2020)



Tourist's Tour Prediction by Sequential Data Mining Approach

Thank you for your attention

smart - - grid.net

Lilia BEN BACCAR, Sonia DJEBALI, Guillaume GUERARD
sonia.djebali@devinci.fr

DE VINCI RESEARCH CENTER

大连东软信息学院 Dalian Neusoft University of Information

21-23 November 2019 i November 2019