What did I do Wrong in my MOBA Game?: Mining Patterns Discriminating Deviant Behaviours

O. Cavadenti, V. Codocedo, J-F Boulicaut, and **M. Kaytoue**

IEEE/ACM DSAA GDS 2016
DSAA Special Session on Game Data Science

October 18th, 2016

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON LIRIS ACTEMIUM

Context

The video game industry

- Millions (billions!) of players worldwide,
- at any-time on any device

The rise of eSports and Streaming

- Teams and sponsors
- Twitch.tv and TVs

Challenge: games shall be hard for pros, enjoyable for casual players

G. Cheung and J. Huang.
Starcraft from the stands: understanding the game spectator.
In *SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2011, pp. 763–772.

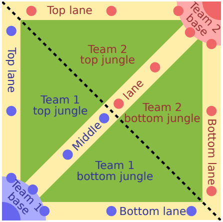
M. Kaytoue, A. Silva, L. Cerf, W. Meira Jr. et C. Raïssi
Watch me playing, i am a professional: a first study on video game live streaming.
In *WWW 2012 (Companion Volume), pages 1181–1188. ACM*, 2012.

T. L. Taylor
Raising the Stakes:E-Sports and the Professionalization of Computer Gaming.
In *MIT Press*, 2012.

Discovering the habits and weaknesses of a MOBA player

Multi-player Online Battle Arena games

- For this talk: DOTA2
- 2 teams playing some kind of rugby
- Equilibrium gets easier to break with time
- Large heroes pool with different roles and style



Requires practice, knowledge ... and advice

- Positioning
- Build order, items
- Experience and gold rates
- Trigger/coordinate team fights, estimating enemy positions
- Micro management

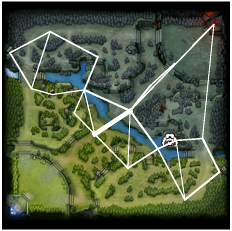
How can I learn from my mistakes?
Can I discover weaknesses from my enemy?

Key Idea: encode game traces and mine patterns

Available information: positioning, build, items, ... and models?

Leagues Of Legends

Mirana (DOTA2)



Pudge (DOTA2)

Describe, compute deviation for mining frequent patterns that discriminate victory, deviation from a standard positioning, ...

pid	Trajectory a	Description	Description	Outlier Score	Victory?
1	$\langle 1,4,7,5,7,5,7 \rangle$	$\{buy_X, buy_Y\}$	$\{ab_{A_1}, ab_{B_2}\}$	0.33	yes
2	$\langle 1,2,3,5,3,5,3 \rangle$	$\{buy_X, buy_Y\}$	$\{ab_{A_1}, ab_{B_2}\}$	0.33	yes
3	$\langle 1,5,7,5,7,5 \rangle$	$\{buy_X\}$	$\{ab_{A_1}, ab_{B_2}\}$	0.40	yes
4	$\langle 1,2,3,5,3,6,3 \rangle$	$\{buy_X, buy_Z\}$	$\{ab_{A_1}, ab_{C_2}\}$	**0.66**	**no**
5	$\langle 1,2,3,5,6,3 \rangle$	$\{buy_Z\}$	$\{ab_{A_1}, ab_{C_2}\}$	**0.80**	**no**

Frequent Pattern Mining

Principle

- A set of items $\mathcal{I}$: an action, the first item bought, ...
- A transaction $t \subseteq \mathcal{I}$ describes the trace of a player
- A transaction database $\mathcal{D} = \{t_1, t_2, ...\}$
- An itemset $X \subseteq \mathcal{I}$ appears in some transactions
- An itemset is frequent if it appears more that a given threshold

id	transaction
t_1	$\{a, b, c\}$
t_2	$\{a, b, c\}$
t_3	$\{c\}$
t_4	$\{a, b, e\}$
t_5	$\{a, e\}$

Example

$supp_{\mathcal{D}}(\{a, b, c\}) = 2$, $supp_{\mathcal{D}}(\{a, b\}) = 3$, $freq_{\mathcal{D}}(\{a, b\}) = 0.6$ and $freq_{\mathcal{D}}(\{a, b, c\}) = 0.2$. If we set the minimal frequency threshold $\sigma = 0.3$, we have that $\{a, c\}$ is frequent while $\{a, b, c\}$ is not a frequent itemset.

Mining Discriminant Patterns

Principle

- A label/class is attached to each transaction
- Find the itemsets that mostly cover a label and not the other

$$\phi(X) = \frac{|supp_{\mathcal{D}^+}(X)| - |supp_{\mathcal{D}^-}(X)|}{|supp_{\mathcal{D}^+}(X)| + |supp_{\mathcal{D}^-}(X)|}$$

id	transaction	class(t)
t_1	$\{a, b, c\}$	$+$
t_2	$\{a, b, c\}$	$+$
t_3	$\{c\}$	$+$
t_4	$\{a, b, e\}$	-
t_5	$\{a, e\}$	-

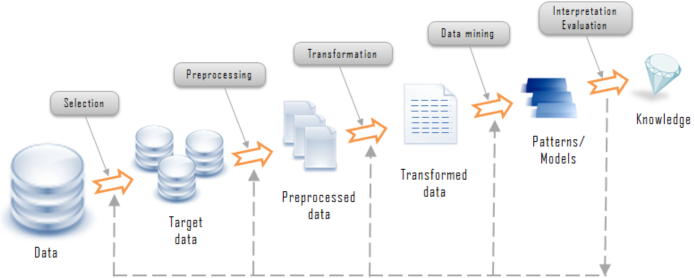
Example

$\phi(\{a\}) = (2-2)/(2+2) = 0$, $\phi(\{a,b\}) = (2-1)/(2+1) = 0.33$, $\phi(\{a,b,c\}) = (2-0)/(2+0) = 1$ and $\phi(\{e\}) = (0-2)/(0+2) = -1$. Consequently, choosing a, b and c can be interesting for a player as it discriminates victory and as it was played relatively often ($freq_{\mathcal{D}^+}(\{a,b,c\}) = 66.66\%$, $freq_{\mathcal{D}^-}(\{a,b,c\}) \; = \; 20\%$).

Pattern Mining for Knowledge Discovery in MOBAs

The different steps of KDD

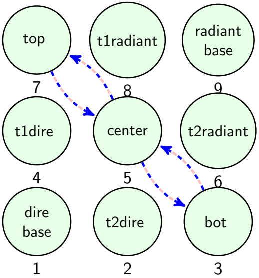
- Select the base to study (a season, a player, a hero, ...)
- Encoding the traces into itemsets
- Choose a pattern domain (itemsets, sequential patterns, ...)
- Determine the labels to discriminate, such as victory, or even a player
- Measure the level of player, his behavior w.r.t standards, ...



Computing a reference behavior graph for DOTA2

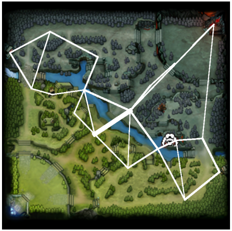
Principle

- Select a set of references player game traces
- Select a set of POIs (towers, shops,...)
- Compute the movement frequencies
- Filter out unfrequent edges
- Store the resulting graph



Leagues Of Legends

Mirana (DOTA2)



Pudge (DOTA2)

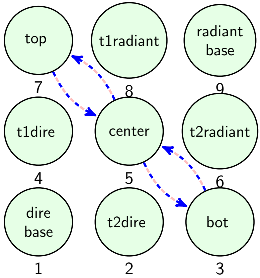
Computing the deviation from a reference model

pid	Trajectory a	Description	Description	Outlier Score	Victory?
1	$\langle 1,4,7,5,7,5,7 \rangle$	$\{buy_X, buy_Y\}$	$\{ab_{A_1}, ab_{B_2}\}$	0.33	yes
2	$\langle 1,2,3,5,3,5,3 \rangle$	$\{buy_X, buy_Y\}$	$\{ab_{A_1}, ab_{B_2}\}$	0.33	yes
3	$\langle 1,5,7,5,7,5 \rangle$	$\{buy_X\}$	$\{ab_{A_1}, ab_{B_2}\}$	0.40	yes
4	$\langle 1,2,3,5,3,6,3 \rangle$	$\{buy_X, buy_Z\}$	$\{ab_{A_1}, ab_{C_2}\}$	**0.66**	**no**
5	$\langle 1,2,3,5,6,3 \rangle$	$\{buy_Z\}$	$\{ab_{A_1}, ab_{C_2}\}$	**0.80**	**no**

Given a trace t and a Reference Model matrix representation M, the outlier score is defined as:

$$\mu(t, M) = \frac{\sum_{i=0}^{i=|trajectory(t)|-1} M(t_i, t_{i+1})}{|trajectory(t)| - 1}$$

where $|.|$ counts the number of POIs



Mining emerging patterns

$$\mathcal{D}^+ = \{description(t) \mid t \in \mathcal{T}, \mu(t, M) \leq \theta\}$$
$$\mathcal{D}^- = \{description(t) \mid t \in \mathcal{T}, \mu(t, M) > \theta\}$$
$$\phi(X) = \frac{|supp_{\mathcal{D}^+}(X)| - |supp_{\mathcal{D}^-}(X)|}{|supp_{\mathcal{D}^+}(X)| + |supp_{\mathcal{D}^-}(X)|}$$

pid	Description	Description	class
1	$\{buy_X, buy_Y\}$	$\{ab_{A_1}, ab_{B_2}\}$	+
2	$\{buy_X, buy_Y\}$	$\{ab_{A_1}, ab_{B_2}\}$	+
3	$\{buy_X\}$	$\{ab_{A_1}, ab_{B_2}\}$	+
4	$\{buy_X, buy_Z\}$	$\{ab_{A_1}, ab_{C_2}\}$	-
5	red$\{buy_Z\}$	$\{ab_{A_1}, ab_{C_2}\}$	-

Example

With $\theta = 0.5$: $\mathcal{D}^+ = \{d(t_1), d(t_2), d(t_3)\}$ and $\mathcal{D}^- = \{d(t_4), d(t_5)\}$. With $min\_sup = 2$, $X_1 = \{buy_X\}$, $X_2 = \{buy_Z\}$, $X_3 = \{buy_X, buy_Y\}$ are frequent

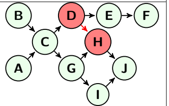
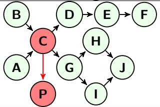
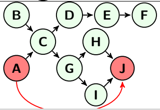
$$\phi(\{buy_X\}) = (3-1)/(3+1) = 0.5$$
$$\phi(\{buy_Z\}) = (0-2)/(0+2) = -1$$
$$\phi(\{buy_X, buy_Y\}) = (2-0)/(2+0) = 1$$

G. Dong, J. Li
Efficient mining of emerging patterns: discovering trends and differences.
KDD 1999

Managing a logistic network with OpenTTD

Video game are also great benchmark datasets!

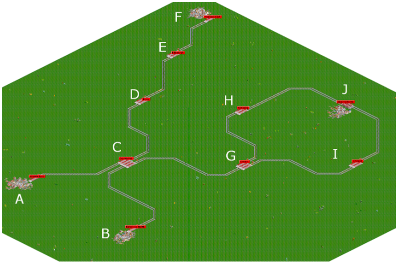
- Managing transportation of (transformed) products and passengers
- FUI Tracaverre (14–17, French ministry of the Industry): unitary traces of products moving in a network with thieves, grey market, ...
- EPCIS Data Generator: `https://github.com/AnesBendimerad/EPCIS-Events-Generator-Based-On-OpenTTD`

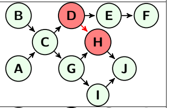
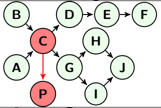
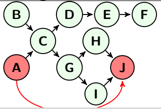
#	$\lvert\mathcal{D}^-\rvert$	$\lvert\mathcal{D}^+\rvert$	$\lvert\mathcal{I}\rvert$	Erreur insérée
1	22	823	319	
2	43	1449	462	
3	109	1826	836	

Managing a logistic network with OpenTTD

Video game are also great benchmark datasets!

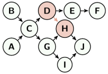
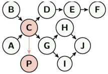
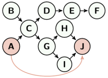
- Managing transportation of (transformed) products and passengers
- FUI Tracaverre (14–17, French ministry of the Industry): unitary traces of products moving in a network with thieves, grey market, ...
- EPCIS Data Generator: `https://github.com/AnesBendimerad/EPCIS-Events-Generator-Based-On-OpenTTD`



#	$\|\mathcal{D}^-\|$	$\|\mathcal{D}^+\|$	$\|\mathcal{I}\|$	Erreur insérée
1	22	823	319	
2	43	1449	462	
3	109	1826	836	

Results with OpenTTD

- Encoding: visited sites, days, resource type
- Data1: 4281 patterns, Data2 : 2842 patterns, Data3: 2930 patterns

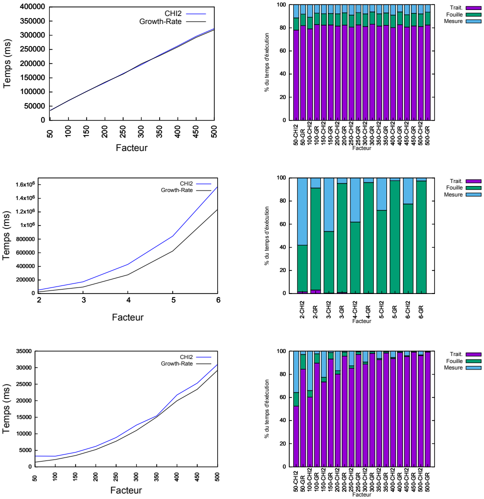
#		Support	Pattern	Score
1	-0.98	**25**	**{*mail*, *Dijon*, *Hamburg*, 27/2/2033, 18/3/2033, 12/3/2033, 6/4/2033, 5/3/2033, 25/3/2033, 31/3/2033, 13/4/2033, 19/4/2033, 26/4/2033}**	
2	-0.98	26	{*mail*, 18/3/2033, 25/3/2033}	
3	-0.98	26	{*Hamburg*, 6/4/2033, 25/3/2033}	
4	-0.98	26	{*mail*, *Dijon*, 18/3/2033, 5/3/2033}	
5	-0.98	26	{*Hamburg*, 13/4/2033, 19/4/2033, 26/4/2033}	
1	-1.0	**43**	**{*Concepcion*, *passenger*, *Problemopolis*}**	
2	-0.65	522	{*Concepcion*, *passenger*}	
3	-0.02	6	{*Concepcion*, *passenger*, *Hamburg*, *Problemopolis*}	
4	-0.015	5	{*Concepcion*, *passenger*, *Edinburgh*, *Problemopolis*}	
5	-0.015	5	{*Concepcion*, *Dijon*, *passenger*, *Problemopolis*}	
1	-0.97	**155**	**{*AtlantaEast*, *JakartaNorth*}**	
2	-0.31	**95**	**{*passenger*, *AtlantaEast*, *JakartaNorth*}**	
3	-0.27	192	{*passenger*, *AtlantaEast*}	
4	-0.18	385	{*passenger*, *Jakarta*, *North*}	
5	-0.16	**60**	**{*mail*, *AtlantaEast*, *JakartaNorth*}**	

$$\theta = 0.003\% \text{ et } min\_freq = 0.001\%$$

Results with OpenTTD

Experimental protocol

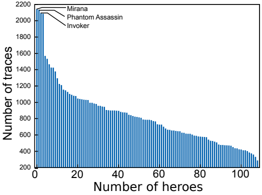
- 845 traces with 319 boolean properties
- Scaling?
  1. Number of traces (x50, ..., x500)
  2. Number of properties (x2,...,x6)
  3. Number of nodes (x50, ..., x500)

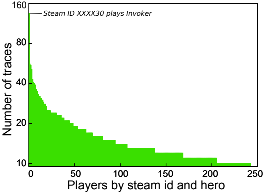


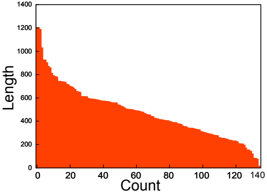
Data & problem settings

Replay collection parsed with Clarity 2.0

- 20,000 DOTA2 replays nicely given by R. Jackson (Dotabank)
- 3,000 replays in Captain's mode
- Split by heroes, focus on mostly played heroes





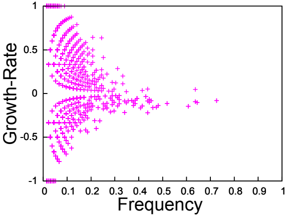


Scenario 1: Patterns that discriminate the game outcome

Experimental protocol

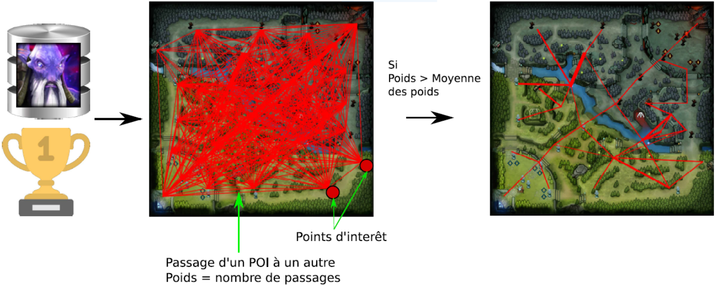
- One single player with *Invoker*: 135 game, 66W/69L (balanced)
- Encoding: bought items and skills taken
- Loosing patterns: items never taken for this class according to *Dotabuff.com*

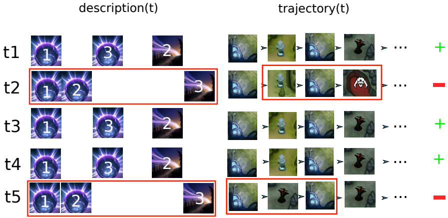
#	X	$supp(X)$	$\phi(X)$
1	{*tpscroll*, *force_staff*, *blade_mail*}	3	-1.0
2	{*tpscroll*, *staff_of_wizardry*, *blade_mail*}	3	-1.0
3	{*tpscroll*, *healing_salve*, *gloves*, *power_treads*}	3	-1.0
4	{*boots*, *tpscroll*, *healing_salve*, *blade_mail*}	3	-1.0
5	{*tango*, *tpscroll*, *force_staff*, *blade_mail*}	2	-1.0

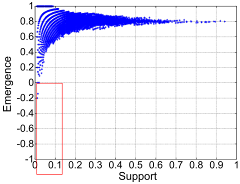


Tradeoff between frequency and win discriminating power

Scenario 2: Patterns of traces deviating from a reference







θ = 0.24

Scenario 2: Patterns of traces deviating from a reference

Discovering strategy errors

- 500 traces of a unique heroes
- Encoding: enemies, skills, visited POIs, ...
- 193 026 frequent patterns, 16 patterns with a negative measure

#	Measure	Support	Pattern
1	-0.66	0.012	{*enemy_queenofpain*, **no_comp_4_level_11**}
2	-0.60	0.01	{*enemy_nyxassassin*, **no_comp_4_level_6**, **poi_infrequent_bot_shop**}
3	-0.42	0.014	{*enemy_rubick*, *no_comp_4_level_11*}
4			{*enemy_nyxassassin*, **poi_infrequent_bot_shop**}
5	-0.33	0.012	{$>$ *_40_dire_fountain*}
6	-0.25	0.016	{*enemy_furion*, **poi_infrequent_bot_shop**}
7	-0.19	0.01	{*enemy_lifestealer*, *enemy_keeperofthelight*, **no_comp_4_level_6**, *no_dire_fort*}
8			{*enemy_medusa*, **no_comp_4_level_6**}
9			{*enemy_chen*, *enemy_gyrocopter*}
10			{*enemy_queenofpain*, *enemy_gyrocopter*}

Top-10 patterns with $\theta = 22\%$ and $min\_sup = 1\%$.

Advice: Skill4 has not been taken at level 6 and shop was not visited

Conclusion

A preliminary work

- Discovering frequent patterns in MOBA data, discriminating a player, victory, a deviation w.r.t. a reference, ...
- Use of basics from closed pattern mining and formal concept analysis
- An expert shall be in the loop (descriptive analytics)

Improvable in many directions

- Each step of the KDD process can be tuned: game selection, reference/target construction/selection, replay encoding, pattern language, ...
- Time shall be take into account: the reference cannot be global
- Items heroes role is more important than hero (carry, ganker, ...)
- Towards a usable tool, many scenarios to be deeply studied

One major limitation is the limited availability of data for some scenario

Other work of the authors related with Game Data Science

Avatar prediction and "smurf" detection in StaCraft II

O. Cavadenti, V. Codocedo, J.-F. Boulicaut, M. Kaytoue
When Cyberathletes Conceal Their Game: Clustering Confusion Matrices to Identify Avatar Aliases.
IEEE DSAA 2015

Discovering and describing balance issues in StaCraft II

G. Bosc, C. Raïssi, J.-F. Boulicaut, P. Tan, M. Kaytoue
A Pattern Mining Approach to Study Strategy Balance in RTS Games
IEEE Transactions on Computational Games and Artificial Intelligence (in press, Dec. 2015).

Thanks for your attention!