



# ALICIA

## Adaptive Learning for User-Centric Applications

ALICIA 3<sup>rd</sup> meeting – Paris October 2<sup>nd</sup> 2014

Bogdan Cautis

Université Paris-Sud – Inria Saclay



# User-centric applications

**Social applications:** user generated content, multi-faceted data, of highly dynamic and heterogeneous nature

**Intelligent crowdsourcing:** data sourcing and micro-task completion

Core tasks (explicit and implicit queries):

- content search / content recommendation
- user search / user recommendation

Common feature: **strongly contextualized** information needs -> must integrate **user profiles, preferences** for query relevance

- Fact: profiles are rarely declared explicitly, not stationary, strongly contextualized.

# ALICIA approach

For strongly contextualized information needs: a **learning** process closing the loop

- account for user feedback, actions, evaluations, interactions
- continuously assess utility of content / users

A system that evolves and learns how to serve information needs: an initial (generic) model is refined with each processed “query” and the observed feedback.

# Broad objectives

Better understand and organize data and users, for improved

- Effectiveness: result quality (may involve completeness, diversity), data quality, task assignment, task completion
- Performance & scalability: may rely on clustering, community detection
- Flexibility / adaptability: performance tradeoff

Requires experimental evaluation, analysis of impact in real use case applications.

# Objectives: Adaptive learning

- Setup and application of adaptive learning techniques in the novel scenarios of user-centric applications
- Extend existing approaches to new demands in these applications

Multi-armed bandits : a natural starting point

- high adaptability, online (by design), optimally manage the exploration/exploitation tradeoff

# Objectives: Multi-armed bandits

Possible resources: pieces of content, workers and experts, producers of content, groups and communities, meta-data, models, MAB algorithms, ...

Extending the state of the art:

- Combination of arms / experts / strategies
- Completeness and compatibility
- Evolving number of arms / experts / strategies
- Scalability

# Objectives: Beyond just lists of results

Answers consisting of several objects

- Crews, packages (composite objects), panoramic view – make it all manageable
  - Complementarity and compatibility
- Covering the different aspects of the query -> a notion of completeness

# Where do we stand ?

- Project (slowly) started in February 2014
- Initially for 42 months -> now ending in January 2018 (40 months remaining)
- Some structural changes: Skyrock left, UPSud joined (replacing the DBWeb@Telecom component), AlephD no longer supporting the Pinterest application
- Taking off with initial hiring, tech. reports, experiments, datasets.



# New task overview

Task	Title	Type	Lead	Start	End
1	<b>Data Models for User-centric Applications</b>		UPSud	0	16/20
1.1	Core Application Data	RD	UPSud	0	7/12
1.2	Higher-order Application Data	RD	LIG	3	16/18
1.3	Physical Data Model	RD	UPSud	4	16/20
2	<b>Algorithms for Adaptive Learning for User Preferences and Expertise</b>		UPS/IMT	4	36/40
2.1	Robustness of MAB Algorithms	RD	IMT	4	22
2.2	Dedicated Bandit Models	RD	UPS	10	36/40
2.3	Algorithms for Learning Complex Objects	RD	LIG	12	36/40
3	<b>Scalable Algorithms for Community Detection, Clustering and Matching</b>		LIG	4	36/42
3.1	Community Detection	RD	LIG	4	26
3.2	Clustering	RD	UPSud	10	28/32
3.3	Matching	RD	UPSud	14	36/42
4	<b>User-Centric Applications</b>		App. partners	0	42/48
4.1	Data Acquisition	RD	UPSud	2	26
4.2	Information Access Applications	RD	Vodkaster / AlephD	14	42/48
4.3	Intelligent Crowdsourcing	RD	Xerox	14	42/48
5	<b>Evaluation</b>		Xerox / UPSud	26	42/48
5.1	Demonstrators for the Project Results	Demo	Vodkaster	26	40/46
5.2	Evaluation and Validation	Eval	Xerox / UPSud	26	42/48
0	<b>Project Management</b>	Mgt	UPSud	0	42/48

# Deliverables (to be slightly shifted too)

Task	Title and substance of the deliverables and milestones	Delivery date, in months starting from T0	Partner in charge of the deliverable
<b>Task 0. Project Management, Organisation</b>			
	T0.1 Progress reports	T0+6, +12, +18, +24, +30, +36, +42	UPSud
	T0.2 Consortium Agreement	T0+9	UPSud
	T0.3 Project Website	T0+4	UPSud
<b>Task 1. Data Models for User-Centric Applications</b>			
	T1.1 Report on modelling core application data	T0+6	UPSud
	T1.2 Report on modeling higher-order application data	T0+16	LIG
	T1.3 Report on physical data organization	T0+16	UPSud
<b>Task 2. Algorithms for Adaptive Learning of User Preferences and Expertise</b>			
	T2.1 Report on robustness of MAB algorithms	T0+18	IMT
	T2.2 Report on dedicated bandit models	T0+24 / T0 +36	UPS/IMT
	T2.2 Toolbox of dedicated bandit models	T0+24 / T0 +36	UPS/IMT
	T2.3 Report on learning complex objects	T0+24 / T0 +36	LIG
	T2.3 Toolbox for learning complex objects with bandits	T0+24 / T0 +36	LIG
<b>Task 3. Scalable Algorithms for Community Detection, Clustering and Matching</b>			
	T3.1 Report on large scale community detection	T0+26	LIG
	T3.1 Toolbox of community detection algorithms	T0+26	LIG
	T3.2 Report on scalable clustering algorithms	T0+28	UPSud
	T3.2 Toolbox of clustering algorithms	T0+28	UPSud
	T3.3 Report on large scale matching algorithms	T0+24 / T0 +36	UPSud
	T3.3 Toolbox of matching algorithms	T0+24 / T0 +36	UPSud
<b>Task 4. User-Centric Applications</b>			
	T4.1 Report on robustness of data acquisition	T0+26	UPSud
	T4.2 Report on information access applications	T0+26 / T0 +42	Vodkaster / AlephD
	T4.2 Prototype on information access	T0+26 / T0 +42	Vodkaster / AlephD
	T4.3 Report on intelligent crowdsourcing	T0+26 / T0 +42	XRCE
	T4.3 Prototype on intelligent crowdsourcing	T0+26 / T0 +42	XRCE
<b>Task 5. Evaluation</b>			
	T5.1 Realizing demonstrators for the techniques resulting from the project	T0+38	Vodkaster
	T5.2 Evaluation report	T0+42	UPSud / XRCE

# The 1<sup>st</sup> project year

- Get to know each other, start collaborations
- Get a good understanding of the problem(s)
  - What application scenarios we want to focus on -> **Tech. report by end of 2014**
  - What kind of information, what level of granularity: core application data, higher-order data, physical data layer...
    - **Shared repository of datasets**
- Defining the ground for the prototype development
  - Architecture and models

# Pressing TODO items

- Website (what is missing?)
- Intellectual property / consortium agreement ?!!
- Logo (friends in graphics?)
- Next meeting (December?, where?)
- ALICIA hiring status
  - UPSud (doc, postdoc), LIG (postdoc?), Toulouse (postdoc?), XRCE (postdoc ?), Telecom ParisTech (doc, postdoc?)
- Datasets: where do we stand ?
- Upload an “updated summary” on the ANR website

Thank you!

