CMSC5733 Social Computing

Tutorial III: Introduction to Project

Shenglin Zhao

The Chinese University of Hong Kong

slzhao@cse.cuhk.edu.hk

Overview

- Goal
  - Hands-on experience on social computing
- Requirement
  - Topics should be related to social computing, e.g., social network analysis, graph theory, recommender systems, Q&A, opinion mining, human computation, etc.
- Group rule
  - Up to four students
  - Each member should undertake specific work

Grading Criteria

- Project proposal and final report: 60%

- Presentation: 40%

We will measure each member's contribution of one group. Members in a group may not gain equal marks.

Grading Criteria

- Factors to be considered :
  - the novelty and utility of your deliverables;
  - the relevance to the course;
  - the challenges you have to solve (i.e., technical contributions);
  - the quality of presentation/writing.

Four Types

- Survey
- Algorithm comparison
- System
- Theoretical paper

Survey

- "A survey is a paper that summarizes and organizes recent research results in a novel way that integrates and adds understanding to work in the field. A survey article assumes a general knowledge of the area; it emphasizes the classification of the existing literature, developing a perspective on the area, and evaluating trends."

As described by ACM Computing Surveys

Your survey paper should …

- Summarize the research in several papers on a particular topic
- Include your own commentary on the significance of the approach and the solutions presented in each paper
- Provide a critical assessment of the work that has been done
- Include a discussion on future research directions
- **REMEMBER**
  - Everything you write in this survey paper has to be in your own words
  - All ideas, paraphrases of other people's words must be correctly attributed in the body of the paper and in the references

- Add quotations

"A location-based social network (LBSN) does not only mean adding a location to an existing social network so that people in the social structure can share location-embedded information, but also consists of the new social structure made up of individuals connected by the interdependency derived from their locations in the physical world as well as their location-tagged media content, such as photos, video, and text. Here, the physical location consists of the instant location of an individual at a given timestamp and the location history that an individual has accumulated in a certain period. Further, the interdependency includes not only that two persons co-occur in the same physical location or share similar location histories but also the knowledge, e.g., common interests, behaviors, and activities, inferred from an individual's location (history) and location-tagged data."

How To Pick Articles – In General

- When picking papers to read - try to:
  - Pick a recent survey of the field so you can quickly gain an overview,
  - Pick a paper that you can easier understand – book chapters often give easier understandable materials and lengthy explanation that may give you a head start, although they may not be as up-to-date as papers,
  - Pick papers that are related to each other in some ways and/or that are in the same field so that you can write a meaningful survey out of them,
  - Favour papers from well-known journals and conferences,
  - Favour "first" or "foundational" papers in the field (as indicated in other people's survey paper),
  - Favour more recent papers,
  - Once you have identified an interesting technology to report upon, follow developments in that strand of technology (e.g. time-wise and technology-wise developments).
  - Find relationships with respect to each other and to your topic area (classification scheme/categorization)

Article Structure

- Title
- Abstract
- Introduction
- Body of paper
- Conclusion/Future Work
- References

Article Structure

- Introduction
  - Significance of the topic
  - Discuss the background and target audience
  - Summarize the surveyed research area and explain why the surveyed area has been studied
  - Summarize the taxonomy scheme you used to do the survey
  - Summarize the surveyed techniques with the above taxonomy scheme

Article Structure

- Survey details/Body of paper
  - Present the surveyed techniques using the taxonomy scheme in detail
  - Identify the trends in the surveyed area. Give evidences for your decision
  - Identify some representative work
  - Identify the open problems/difficulties, and future research issues

Article Structure

- Conclusions/Future work
  - Summarize the conclusions of your survey
  - List open problems and possible future work directions
- References
  - List all the citations referenced in your paper

Figures and Tables

- Take figures from papers including appropriate citations
- Draw your own figures to show taxonomy or structure of the survey
- Use tables to organize comparisons between applications/systems/etc

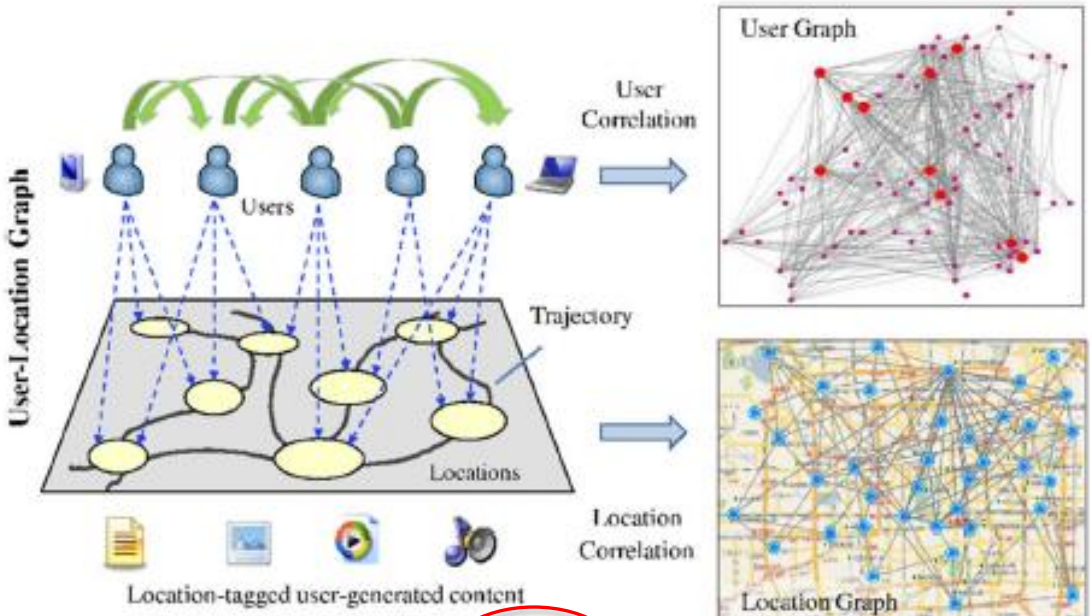


Fig. 1 Concept of location-based social networks [137]

Table 2 Statistics on literatures related to recommendations in LBSNs

Names	2008	2009	2010	2011	2012	2013
Conferences						
WWW	0	2	3	2	1	1
MDM	1	1	1	1	0	2
KDD	0	0	1	4	3	3
ACM-GIS	1	1	2	3	2	2
UbiComp	0	0	4	1	0	2
LBSN	N/A	3	3	5	4	N/A
RecSys	0	0	2	1	1	2
Journals						
VLDB	0	0	2	0	1	0
ACM-TIST	0	0	1	1	4	2
ACM TWEB	0	0	0	1	0	0
Total numbers	2	7	19	19	16	14

Table 2: Features of different systems

System	Computation Task	Computation Mode	Platform	Software
MapReduce	General	Synchronous & Asyschronous[1]	Cluster	Hadoop [3]
Dryad	General	Synchronous & Asyschronous[2]	Cluster	Dryad&DryadLink [6]
Pregel	Graph	Synchronous	Cluster	Giraph [2]
Spark	Iterative or interactive applications	—	Cluster	Spark [9]
GraphLab	Graph	Synchronous & Asyschronous	Cluster	GraphLab [8]
GraphChi	Graph	Synchronous & Asyschronous	PC	GraphChi [7]
TurboGraph	Graph	Synchronous & Asyschronous	PC	TurboGraph [10]

Examples

Algorithm Comparison

- Same report paper structure
  - Title
  - Abstract
  - Introduction
  - Body of paper
  - Conclusion/Future Work
  - References

Most important:

Final report should contain runnable code

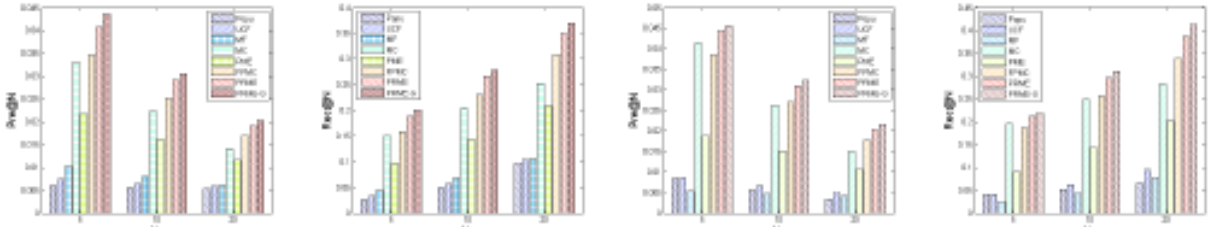
Algorithm Comparison

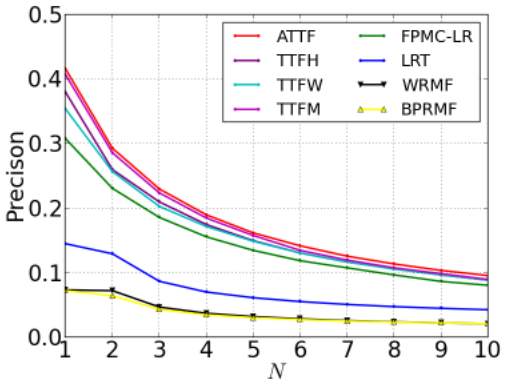
- Introduction
  - Same: Topic significance, background, etc.
  - Extra: why you choose these algorithms

- Body of paper
  - Explain your implemented algorithms
  - Data source
  - Experiment
    - Results comparison, pros. vs. cons.
    - Sensitivity analysis

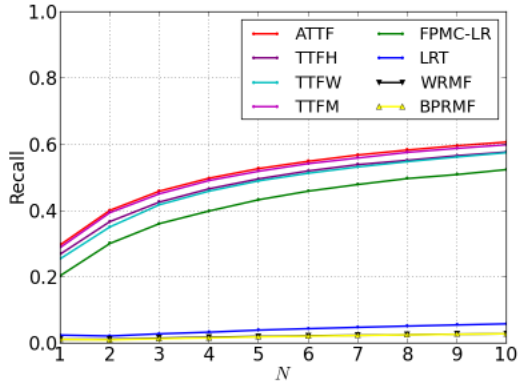
Algorithm Comparison

- Conclusion/Future Work
  - Summarize, which one is better? Advantages and disadvantages.
  - Is it possible to improve the algorithms? How to improve?

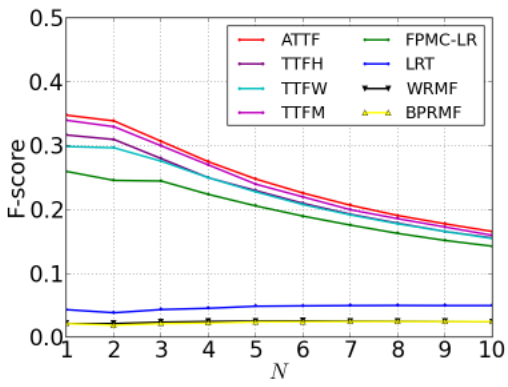




(a) Precision.



(b) Recall.



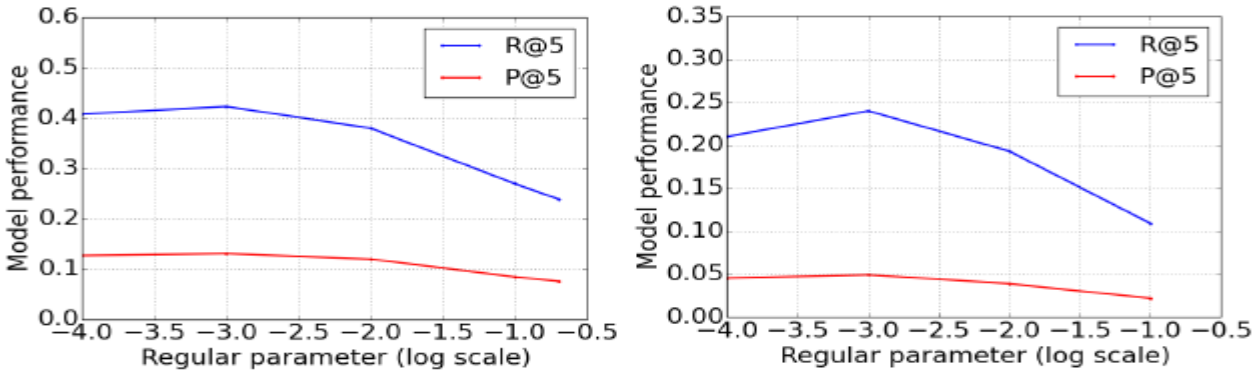
(c) F-score.

Table 2: Performance Comparison

		BPRMF	WRMF	LRT	FPMC−LR	TLAR	SLAR	STELLAR
Gowalla	P@5	0.025	0.031	0.033	0.048	0.053	0.050	**0.059**
	R@5	0.020	0.025	0.030	0.167	0.204	0.197	**0.226**
Foursquare	P@5	0.031	0.033	0.061	0.109	0.119	0.114	**0.129**
	R@5	0.027	0.028	0.053	0.347	0.373	0.368	**0.425**

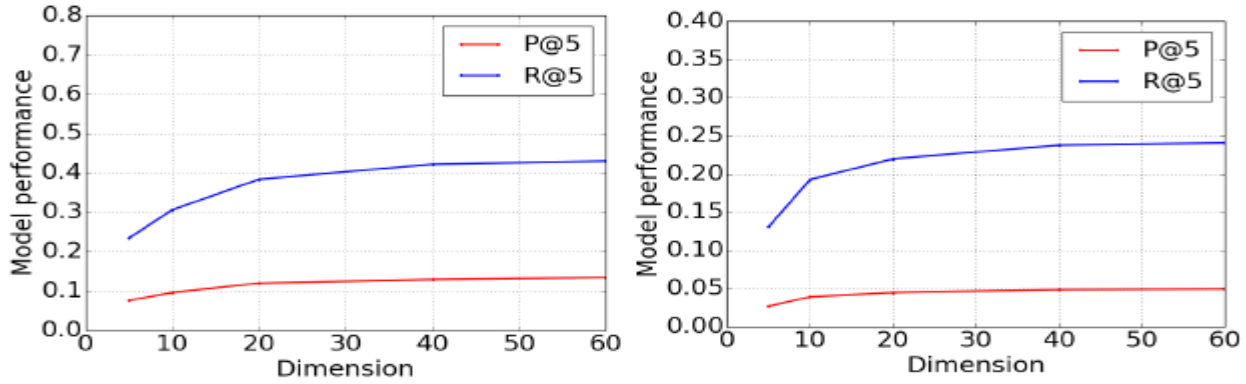
Dataset	**Metric**	**SCP**	**LC**	**BNMF**	**BNMTF**	**BigCLAM**	**PNMF**
Dolphins	M	0.3049	0.6538	0.5067	0.5067	0.4226	**0.9787**
Les Misérables	M	0.3066	0.7730	0.1247	0.1031	0.5395	**1.1028**
Books about US politics	M	0.4955	0.8507	0.4613	0.4924	0.5290	**0.8640**
Word adjacencies	M	0.0707	0.2705	0.2539	0.2677	0.2312	**0.6680**
American college football	M	0.6050	0.8907	0.5584	0.5733	0.5175	**1.0492**
Jazz musicians	M	0.0114	1.1424	0.1133	0.1118	**1.1438**	0.9357
Network science	M	0.7286	0.9558	0.6607	0.7413	0.5026	**1.6570**
Power grid	M	0.0439	0.3713	0.3417	0.3682	1.0097	**1.1051**
High-energy theory	M	0.5427	**0.9965**	0.5648	0.6004	0.9636	0.9725
DBLP	F_1	0.0967	0.0402	-	-	0.0390	**0.0985**
Amazon	F_1	0.0315	0.0070	-	-	**0.0441**	0.0419
YouTube	F_1	0.0445	-	-	-	0.0194	**0.0605**

Table 2: **Experimental results in terms of modularity (M) and $\mathbf{F_1}$ score (F_1).**



(a) Fousquare (b) Gowalla

Figure 4: The effect of regularization



(a) Foursquare (b) Gowalla

System

- Same report paper structure
  - Title
  - Abstract
  - Introduction
  - Body of paper
  - Conclusion/Future Work
  - References

Most important: demonstrable

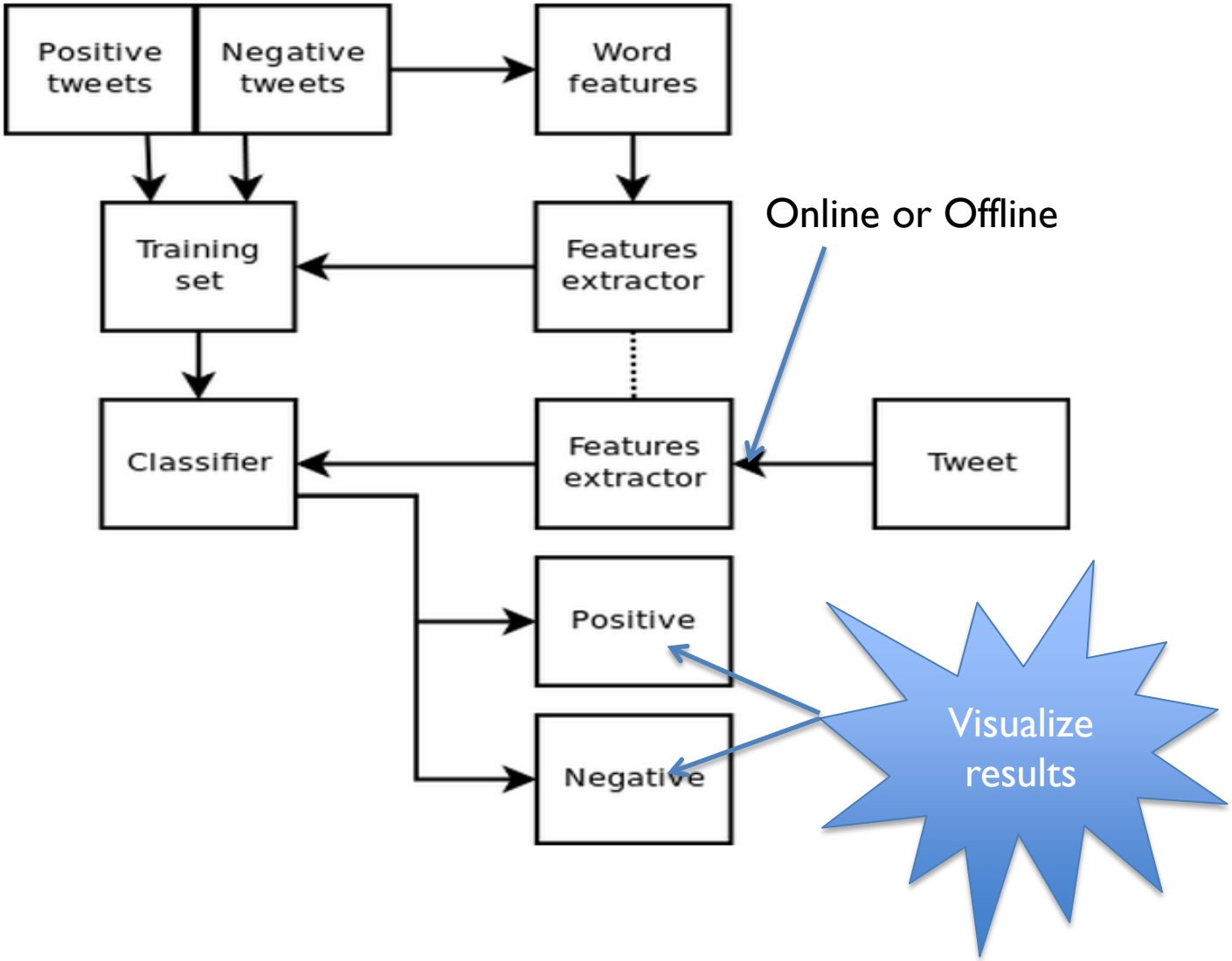
System

- Body of paper
  - system overview
  - interface
    - what kind of data format
    - Interfaces between different modules (e.g., data transition, data store, etc.)
  - algorithm
  - visualization

Example

- A light project to demonstrate twitter sentiment analysis
  - Given a twitter record, demonstrate sentiment analysis
  - Online or offline

System Overview

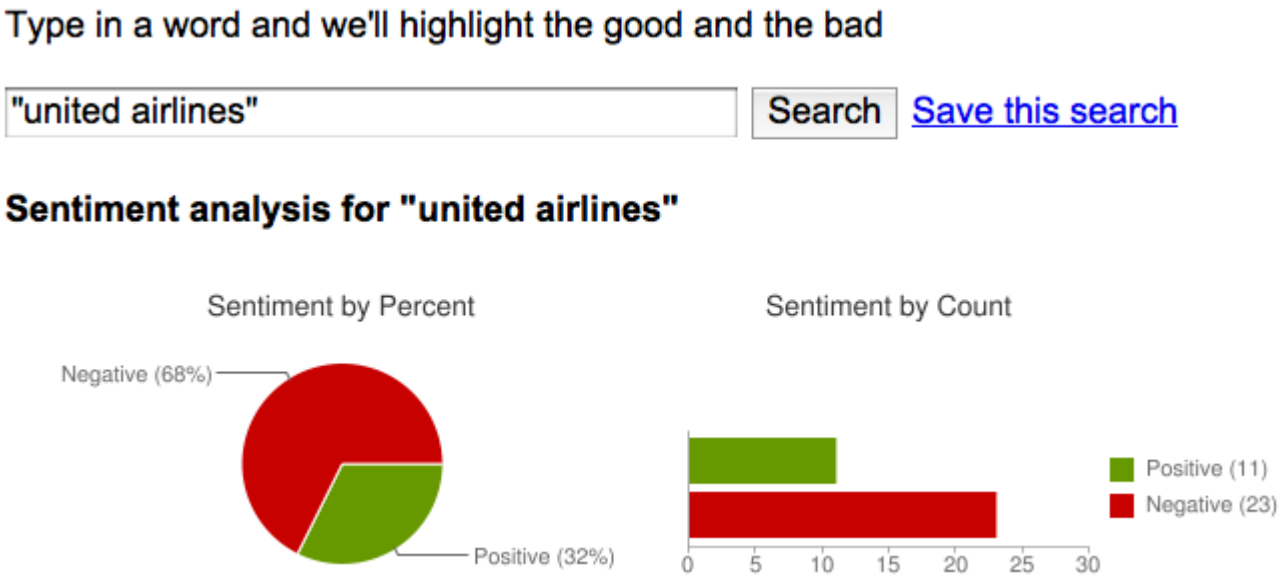


- Interface
  - data format: string
  - data store: local file system
  - Easy to define the interfaces

- Algorithm
  - Naïve Bayesian Classifier
  - SVM

Visualization Example

– Offline

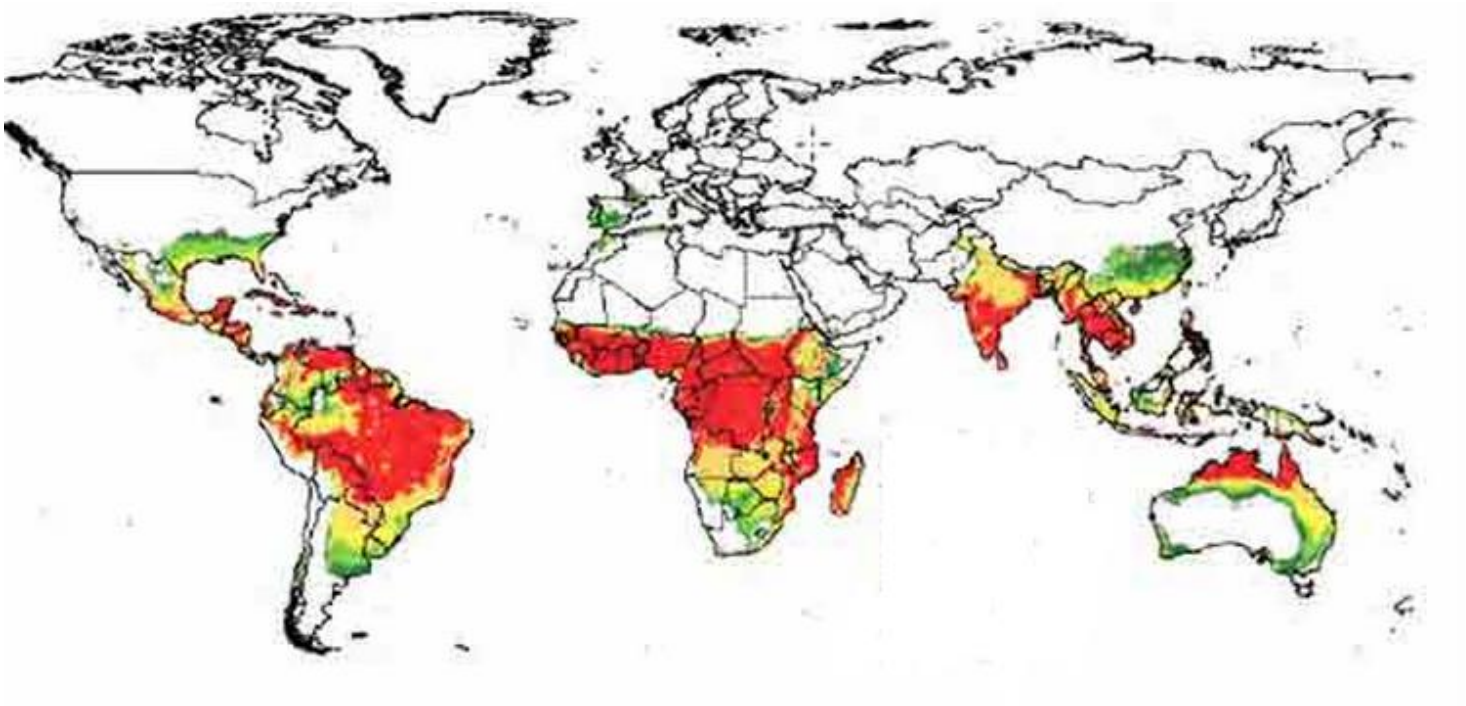


Visualization Example

– Online

Record	Result
I love this car.	
This view is amazing.	
He is my best friend.	
I do not like this car.	
This view is horrible.	
He is my enemy.	

Visualization – Geo Distribution



http://www.mangifera.org/images/geo.jpg

Visualization – Gender Distribution

https://www.clicky.co.uk/wp-content/uploads/2009/07/facebook-dem-gender-1.jpg

Visualization

- Age Distribution
- Nationality Distribution
- Education Distribution
- Change Distribution

Real Example

- http://www.streamcrab.com/results?keyword=mac

Theoretical Paper

- Four steps:
  - Find a familiar topic
  - Find an open problem
  - Resolve the problem
  - Write your paper

Theoretical Paper

- Paper structure
  - Title
  - Abstract
  - Introduction
  - Related Work
  - Model
  - Experiment
  - Conclusion/Future Work
  - References

Most important: novel

Theoretical Paper

- Abstract
  - Background, topic significance
  - Motivation, the problem
  - Your method
  - Result

Theoretical Paper

- Introduction
  - Background, topic significance
  - Motivation, the problem
  - Your method (brief)
  - Your contributions
- Related Work
  - Review the literature
  - Clarify your difference

Theoretical Paper

- Model
  - Demonstrate your model
  - Show data analysis if your model bases on it
- Experiment
  - Data source
  - Experiment setting
    - How to train your model
    - How to test your model
    - Metrics
  - Comparison with state-of-the-art methods
  - Discussion
    - Advantages and disadvantages, and why

Theoretical Paper

- Conclusion and Future Work
  - Summarize this paper
  - Possible directions to improve your model

Example

Data Sources

- http://snap.stanford.edu/data/links.html
- http://snap.stanford.edu/data/index.html
- http://socialcomputing.asu.edu/pages/datasets
- http://archive.ics.uci.edu/ml/index.html
- https://www.kaggle.com/c/FacebookRecruiting/data
- http://tianchi.aliyun.com/

References

- http://www.laurentluce.com/posts/twitter-sentiment-analysis-using-python-and-nltk/
- http://www.laurentluce.com/posts/python-twitter-statistics-and-the-2012-french-presidential-election/
- What is a Survey Paper? 2009, JenniferWong. http://www.cise.ufl.edu/~qxiao/csqjxiao_files/What%20is%20a%20Survey%20Paper.ppt
- http://xiaoxinzhou.blog.163.com/blog/static/207045386201322032428 30/
- Zhang, Hongyi, Irwin King, and Michael R. Lyu. "Incorporating Implicit Link Preference Into Overlapping Community Detection." *Twenty-Ninth AAAI Conference on Artificial Intelligence*. 2015.
- Yu, Yonghong, and Xingguo Chen. "A Survey of Point-of-Interest Recommendation in Location-Based Social Networks." *Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence*. 2015.
- Bao, Jie, et al. "A survey on recommendations in location-based social networks." *ACM Transaction on Intelligent Systems and Technology* (2013).