

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/321712569>

Recommender System Through Sentiment Analysis

Conference Paper · December 2017

CITATIONS

30

READS

5,421

7 authors, including:



Amel Ziani

Badji Mokhtar - Annaba University

20 PUBLICATIONS 137 CITATIONS

SEE PROFILE



Nabiha Azizi

Badji Mokhtar - Annaba University

91 PUBLICATIONS 705 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



A new hybrid method combining genetic algorithm and support vector machine classifier: Application to CAD system for mammogram images [View project](#)



Healthcare Information Systems Security and Communication [View project](#)



Badji Mokhtar University,
Annaba, Algeria
LRI, LABGED Laboratories,
Computer science departement



Recommender System Through Sentiment Analysis

Presented By: **Nabiha AZIZI**

*Authors: Amel ZIANI, Nabiha AZIZI, Didier SCHWAB,
Monther ALDWAIRI, Nassira CHEKKAI, Djamel ZENAKHRA,
Soraya CHERIGUENE.*

The 2nd International Conference on Automatic control, Telecommunication and
Signals (ICATS'17)



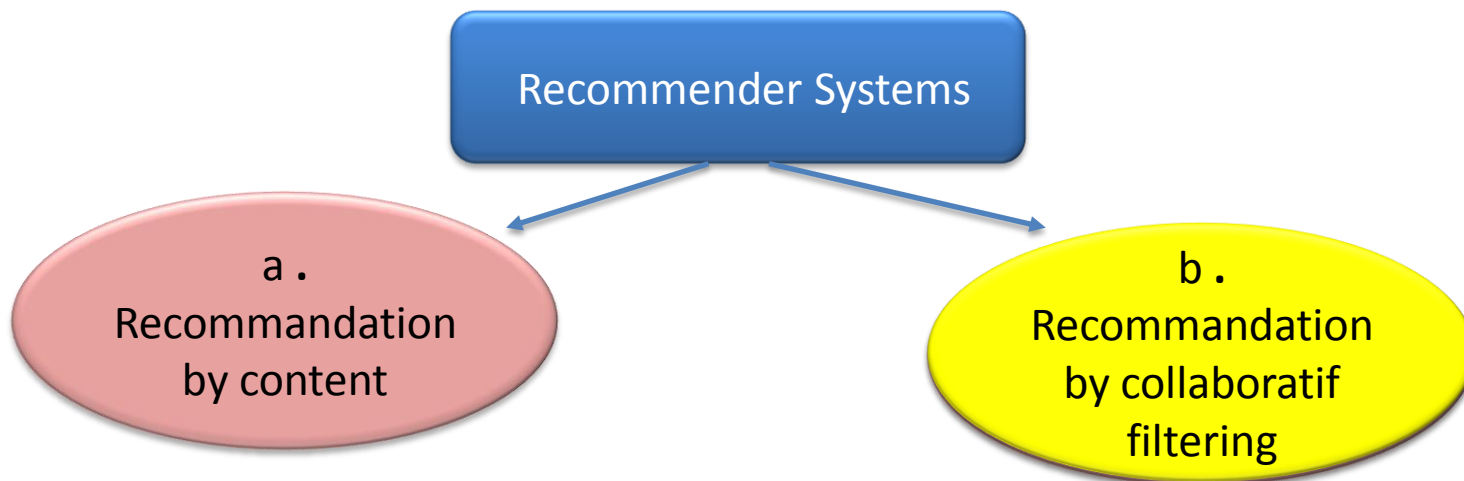
OUTLINE

- Introduction
- Objective
- Opinion mining
- Proposed System
- Feature extraction
- Semi supervised Learning
- Experiment results
- Conclusion

Introduction

What are recommender systems?

- A **recommender system** (RS) helps people that have not sufficient personal experience or competence to evaluate the potentially overwhelming, number of alternatives offered by a Web site.
 - In their simplest form **RSs** recommend to their users personalized and **ranked lists of items**
 - Provide consumers with **information to help them decide** which **items** to purchase,



Introduction



Traditional
methods

Exploit the votes

Ignore the reviews

Vote

- What book should I read?
- What movie should I watch?
- What doctor should I go to?
- What sports news should I read?

AVIS

DESRIPTIF | FICHE TECHNIQUE | AVIS | HAUT DE PAGE ↑

★★★★★ 4 avis - Donnez votre

par Jordan ,
le 6 mai

par Sabire A,
le 26 mars

par Thierry M,
le 17 mars

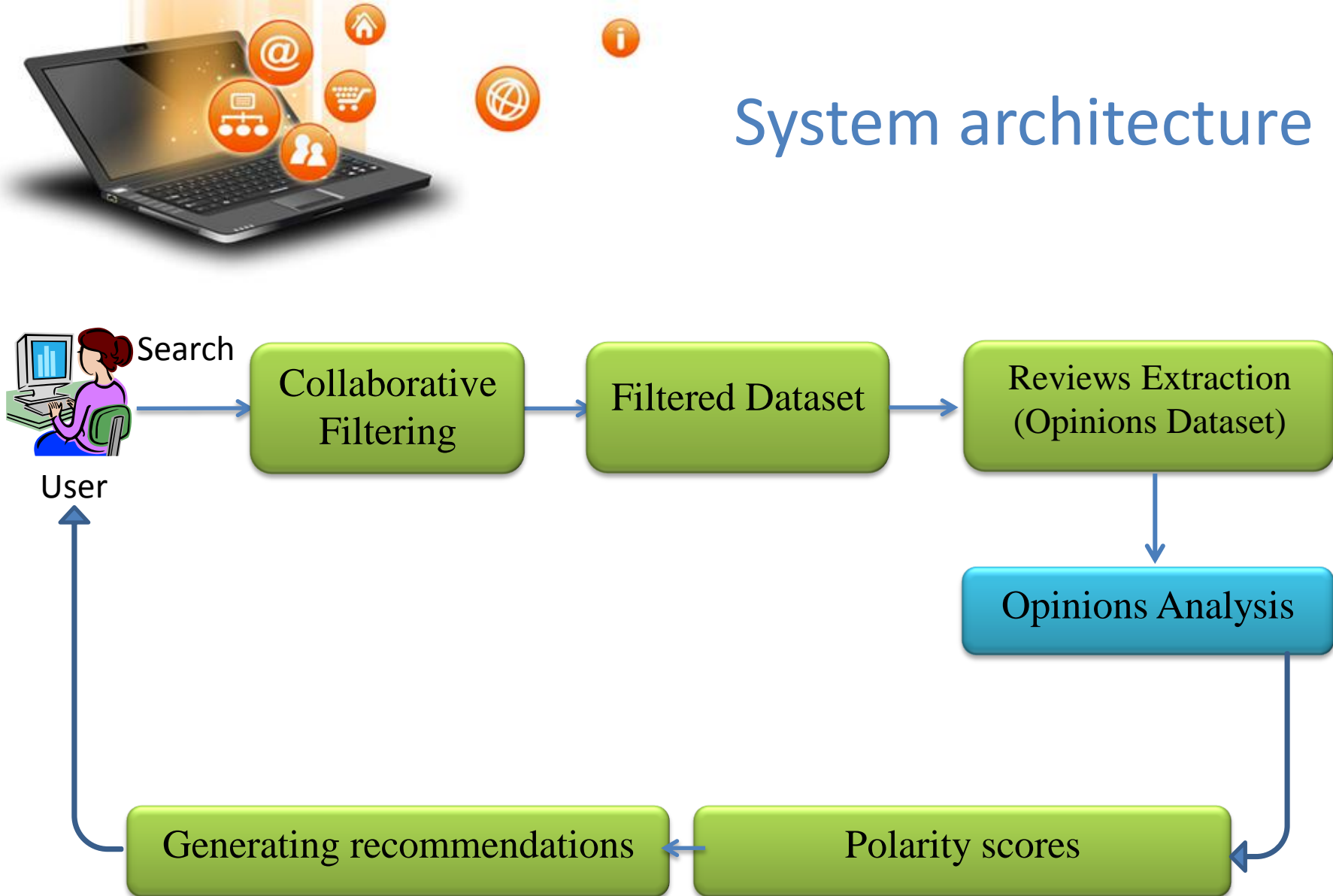
bon...
les bienvenus: Je me suis retrouvé avec 6 Go de libre (sans documents) lorsque Windows a commencé à télécharger les updates. Je retire une étoile pour cela: Prévoir une carte micro-SD pour utiliser ce PC. Sinon, très emballé et répond à mes besoins, surtout pour ce prix! [Lire la suite](#) »



Objective

- The radically innovative objective of our project is to
- formalize the functioning and impact of a recommendation system based on two fields:
- Sentiment analysis and Recommendation using collaborative filtering For Arabic Language.
- Thus, we propose to integrate an opinions analysis system based on the semi supervised classification tested on multilingual recommendation system to improve the recommendation system

System architecture





Why Opinion Mining

- Other's opinions can be crucial to make a decision or choose among multiple options.
- When those choices involve valuable resources , people often rely on their peers' past experiences.

So whom shall we ask?

Before Web

- Friends, relatives and acquaintances
- Consumer Reports



With Web

- Blogs, E-commerce sites
- Review sites
- Discussion forums





Opinion Mining

- **Opinion Mining** (Sentiment Analysis) is a challenging task that combines data mining and Natural Language Processing (NLP) techniques in order to computationally treat subjectivity in textual documents.
- **Sentiment**
A thought, view, or attitude, especially and mainly based on emotion instead of reason.

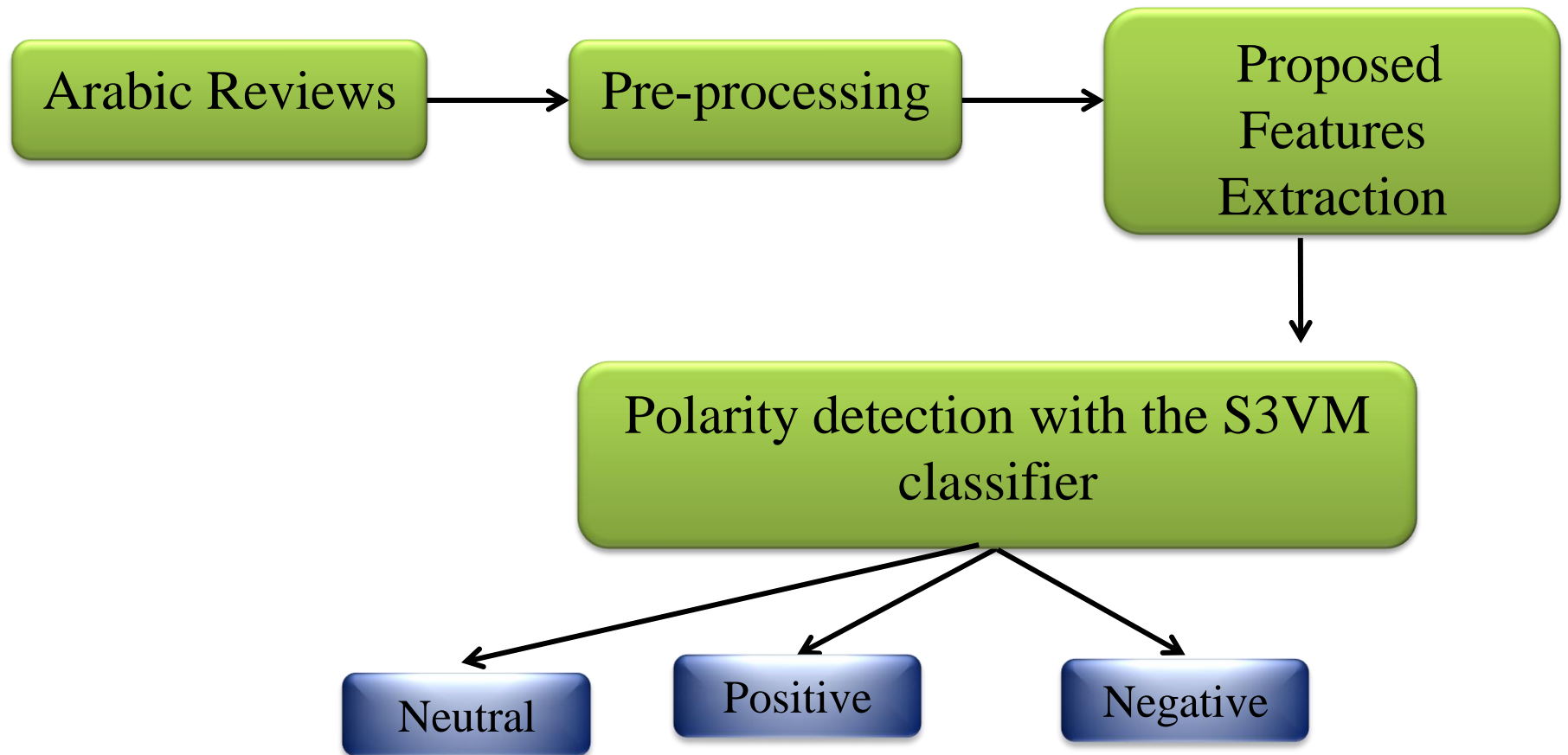


So, what is Subjectivity?

- Subjectivity analysis - is the computational study of affect, opinions, and sentiments expressed in text:
 - Blogs
 - Editorials
 - reviews (of products, movies, books, etc.)
 - newspaper articles



The proposed process for Opinion classification





Features extraction

Sentence	Number of sentences
Positive Words	Number of positive words
Negative Words	Number of negative words
Neutral Words	Number of neutral words
Polarity	Sum of polarity words
Positive polarity	Average of positive polarity words
Negative polarity	Average of negative polarity words
Neutral polarity	Average of neutral polarity words
Predicates	Number of predicates
	Average of predicates
Adverbs	Number of adverbs
	Average of adverbs
Adjectives	Number of adjectives
	Average of adjectives
Emotionalism	The emotionalism of the document
Addressage	The addressage of the document
Reflexivity	The reflexivity of the document



Classification

- ❖ In a another previous work we have analyzed the SVM with his four kernel functions . We have obtained promising results but with the increase of the dataset the accuracy of the systems drop down. The results are shown in the next table:
- ❖ Ziani & al (2015) : “ Recommender System For Sports Articles Based On Arabic Opinions Polarity Detection With A Hybrid Approach RSS-SVM” in **3rd International Conference on Control, Engineering & Information Technology CEIT’2015, IEEE.**

SVM classifier	Precision	Recall	Fmeasure	Accuracy
Linear	41.22	34.32	31.11	28.56
Polynomial	51.01	45.02	50.22	35.01
Gaussian	62.32	64.02	60.00	59.85
Tangent	58.04	49.02	39.03	50.00



- ❑ Later, in **Ziani & al (2015)** : “ Combining Random Sub Space Algorithm and Support Vector Machines Classifier for Arabic Opinions Analysis” in **3rd International Conference on Computer Science, Applied Mathematics and Applications - ICCSAMA 2015, Springer.**

we have analyzed the Random Sub Space (RSS) algorithm behavior with other kind of classifier more stable, so the decision tree classifier is replaced by Support vector machine in order to take into account all features / using at each time a limited number of these primitives.

- ❑ The results of this experiment also were promising and this combination ameliorated the RSS and the SVM behaviors in the Arabic opinion mining classification.



Classification

- Despite all the encouraging obtained results and the improvements in the opinion mining classification system performance, there is always a problem when we increase the dataset volume of reviews or lexicon words ;
- So we can't assure that the SVM can be the most accurate classifier for the opinion mining classification.



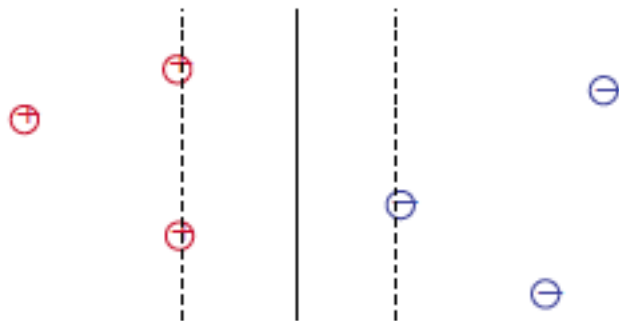
Semi-Supervised Learning

- Can we improve the quality of our learning by taking into account all data set (labeled and unlabeled data)
- Usually a lot more unlabeled data available than labeled
- Assume a set L of labeled data and U of unlabeled data (from the same distribution)
- Focus on Semi-Supervised Classification though there are many other variations
 - Aiding clustering with some labeled data
 - Regression
 - Model selection with unlabeled data (COD)
- Transduction vs Induction

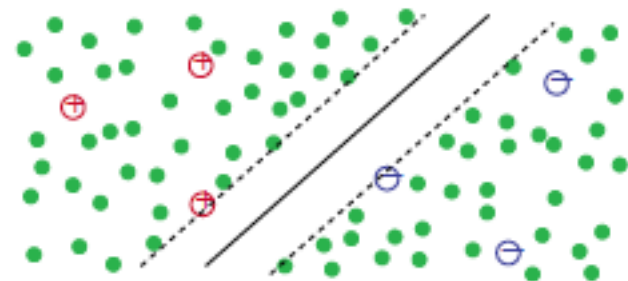


S3VM Classifier

- Semi-Supervised SVM (S3VM) or Transductive SVM (TSVM)
- Maximize margin of both L and U . Decision surface placed in non-dense spaces
 - Assumes classes are "well-separated"
 - Can also try to simultaneously maintain class



(a) SVM decision boundary



(b) S3VM decision boundary



Experiments and results

We carried out multiple experiments to calculate the previous measures.

We have tested our approach using different datasets in multiple languages to confirm the efficiency of our system:



Recommendation

Once the neighbors are obtained and their associated reviews are classified, a weighted average is used to combine the neighbors' item ratings to produce a prediction value for the target user.

As soon as all the similarities of the target user A with respect to the other users are calculated using Spearman similarity and the “ n ” most similar users that constitute the vicinity of this target user are defined, the prediction (P_{Aj}) of the value of an item j evaluated by the user A is calculated using the weighted sum of the estimates of the nearest neighbors who have already estimated the item j as follows:

$$P_{Aj} = \bar{v}_A + \frac{\sum_{i=1}^n \text{sim}(A,i) * (v_{i,j} - \bar{v}_i)}{\sum_{i=1}^n |\text{sim}(A,i)|}$$

$$\text{sim} = \frac{\sum_{i=1}^n ((\text{rang}(x_i - \overline{\text{rang}(x)}))(\text{rang}(y_i - \overline{\text{rang}(y)})))}{\sqrt{\sum_{i=1}^n ((\text{rang}(x_i - \overline{\text{rang}(x)})^2 \sum_{i=1}^n (\text{rang}(y_i - \overline{\text{rang}(y)})^2)}$$



Experiments and results

- **English dataset:** the Restaurant_TijuanaRestaurant dataset which contains 2000 reviews from 50 guests in 40 restaurants. The choice of this benchmark is justified by the fact that it contains all the necessary information on which our contribution is based (Comments and votes).
- **French dataset:** collected from Idlc.com website "HighTech Experience" which is a site for online sale of computer hardware. Our data structure contains 10 users, 5 smartphones and 50 evaluations.
- **Arabic and dialect dataset:** built from the Algerian website dz.jumia.com "JumiaMarket" which is an Algerian website of online shopping in Algeria. Our data structure consists of 10 users, 5 oriental clothing for women and 50 evaluations
- Collect of one thousand reviews in Arabic (MSA), one hundred in dialect and two thousand in French and English. We also have succeeded to collect one million (1 000 000) of unlabeled reviews by applying a web crawler (web robot that systematically browses the World Wide Web).



Experiments and results

After analyzing the results of the three datasets, we have concluded that our method was performing in terms of precision, which guarantees the improvement of the recommendations by proposing to the user good recommendations. Even the recall was always estimated at 100% that can maximize the chance for a user to access the items he would like to have.

	Precision	Recall
English	0.96	1.0
French	1.0	1.0
Arabic and dialect	0.90	1.0



Conclusions and perspectives

- ❑ This study presented a basic tool which can be used to recommend Algerian users by analyzing the reviews and detect its polarity.
- ❑ The results were promising, and that is what encouraged us to continue working along this line.
- ❑ Enrich feature vector with another sort of morphological primitives.
- ❑ Incorporate the system in realtime with recommendation system.

Thank you for your attention



Any suggestions or questions are welcome