

Enhanced Information Retrieval by Relevance Feedback Incorporation in WordSeer

INFO 256 Final Project
Yee Tung (Alice) Man & Robert (Alex) Matevish

Introduction

- WordSeer: a text analysis environment for humanities scholars for retrieving relevant information more easily and accurately.
- To overcome the problem using keyword search that being difficult to find conceptually linked passages of text, the relevance feedback method is incorporated in WordSeer.
- The current relevance feedback algorithm incorporated in WordSeer is the standard Rocchio Algorithm. In this project, we incorporate some other variations of Relevance Feedback algorithms to compare the recall rates.

Algorithm Comparisons

- Rocchio

$$Q_{new} = Q_{old} + \beta \sum_{n1 \text{ rel docs}} \frac{D_i}{n_1} - \gamma \sum_{n2 \text{ nonrelevant docs}} \frac{D_i}{n_2}$$

- Ide dec-chi

$$Q_{new} = Q_{old} + \sum_{all \text{ relevant}} D_i - \sum_{one \text{ top-most nonrelevant}} D_i$$

- Ide regular

$$Q_{new} = Q_{old} + \sum_{all \text{ relevant}} D_i - \sum_{all \text{ nonrelevant}} D_i$$

Rocchio

$$Q_{new} = Q_{old} + \beta \sum_{n1 \text{ rel docs}} \frac{D_i}{n_1} - \gamma \sum_{n2 \text{ nonrelevant docs}} \frac{D_i}{n_2}$$

 relevant

 Non-relevant

- Add reduced term weights to query
- Normalize by dividing the number of documents used for retrieval

$$\beta + \gamma = 1.0$$

Ide dec-chi and Ide Regular

• Ide dec-chi

$$Q_{new} = Q_{old} + \sum_{all\ relevant} D_i - \underbrace{\sum_{one\ top-most\ nonrelevant} D_i}_{}$$

ONE Top-most Non-relevant

• Ide Regular

$$Q_{new} = Q_{old} + \sum_{all\ relevant} D_i - \underbrace{\sum_{all\ nonrelevant} D_i}_{}$$

ALL Non-relevant

Compare to Rocchio:

- Without Normalization
- Without Weighing terms

Implementation in WordSeer

Query Collection Personal Ads Search Reset

147 results

Search term

This is Tanzir, 28 years old man from central London.

Hellow, my name is Jason i am 29 years old and what i am looking for is a marriage minded family oriented girl woman between 18 and 45 years old for a long term relationship what to lead to marriage i currently i live in europe in romania but i have relatives in PA and friends in OR and TX and i am looking for a girl or a woman who is willing and agree to try a long distance relationship to see where lord would lead us i am christian and i am attending a baptist church in my area, well this is all what i had to say for the moment who is interested in to get to know me get to know eachoder please email me.

Sentences

Hey I'm on here looking to get to know some and see how it goes I'm 25 years old no kids single and looking I want to date someone and see where it leads if you are looking for a one night stand or a bbw Please Do n't email me that is Not what im looking for I like to go out to eat movies bowling shopping hanging with friends bars playing video games writing poems reading going to the zoo and the aquarium I like to try new things I'm tall small bbw so if u not into bbw I'm not the one for you please send me a pic and info about you if not I will not get back to you

This is Tanzir, 28 years old man from central London.

I am a 20 year old girl looking for someone who will take care of me and show me all the nice places that London has to offer.

Hi I'm a 47 year old, tall, slim black woman in search of a man who is ready to stop being single.

PlusSize GreekLady (36 years old) Looking For A Single GreekMan 36-46 years old From queens That speaksGreek For a Serious Relationship Leading to Marriage Include Your Age Location & A Photo With Your Reply

Hardworking, friendly, affectionate, 41 year old Latin bbw, hoping to find a friend... a companion to go out with occasionally, have fun... get to know each other and see where it may lead.

Non-Relevant **neutral** **Relevant**

Not relevant Neutral Relevant

SELECT

Refine results

Relevant sentences

- Call me old fashioned, but that's not cool.
- old and very single.
- Are you an old fashioned man who still believes in romance?
- I am open to men that are older than me but not too old to the point he can be my father:) I am looking for a man that know what he wants does n't play games and not a man that looking for another one night stand or another girlfriend or mistress.
- I'm looking for a relationship that's new and exciting, and even when it gets old, it feels new everyday.

Non-relevant sentences

- This is Tanzir, 28 years old man from central London.
- Hellow, my name is Jason i am 29 years old and what i am looking for is a marriage minded family oriented girl woman between 18 and 45 years old for a long term relationship what to lead to marriage i currently i live in europe in romania but i have relatives in PA and friends in OR and TX and i am looking for a girl or a woman who is willing and agree to try a long distance relationship to see where lord would lead us i am christian and i am attending a baptist church in my area, well this is all what i had to say for the moment who is interested in to get to know me get to know eachoder please email me.

Relevant

Non-Relevant

Rocchio

```
//vector adjustment: Rocchio
function calculate_sentence_adjustment($query, $vector_query, $relevant, $irrelevant){
    global $ALPHA_plus; // relevant sentences weight
    global $ALPHA_minus; // irrelevant sentences weight
    $already_relevant = (array) $query['relevant'];
    $already_irrelevant = (array) $query['irrelevant'];

    $new_relevant = array_subtract($already_relevant, $relevant);
    $no_longer_relevant = array_subtract($relevant, $already_relevant);

    $new_irrelevant = array_subtract($already_irrelevant, $irrelevant);
    $no_longer_irrelevant = array_subtract($irrelevant, $already_irrelevant);

    $relevant_vect = convert_sentence_IDs_to_sparse_vector($new_relevant);
    $relevant_vect->normalize();

    $no_longer_relevant_vect = convert_sentence_IDs_to_sparse_vector($no_longer_relevant);
    $no_longer_relevant_vect->normalize();
    Normalization
    $irrelevant_vect = convert_sentence_IDs_to_sparse_vector($new_irrelevant);
    $irrelevant_vect->normalize();

    $no_longer_irrelevant_vect = convert_sentence_IDs_to_sparse_vector($no_longer_irrelevant);
    $no_longer_irrelevant_vect->normalize();

    $positive_adjustment = $relevant_vect->scalarMultiply($ALPHA_plus);
    Weigh terms for relevant and non-relevant sentences
    $negative_adjustment = $irrelevant_vect->scalarMultiply(-1*$ALPHA_minus);
    $no_longer_negative_adjustment = $no_longer_irrelevant_vect->scalarMultiply($ALPHA_minus);

    $adjustment = $positive_adjustment->vectorAdd($negative_adjustment);
    $adjustment = $adjustment->vectorAdd($no_longer_positive_adjustment);
    $adjustment = $adjustment->vectorAdd($no_longer_negative_adjustment);
    return $adjustment;
}
```

Ide dec-chi

```
// vector adjustment: Ide dec-hi
function calculate_sentence_adjustment_ide_dec($query, $vector_query, $relevant, $irrelevant){
    GLOBAL $ALPHA_w_plus;
    $already_relevant = (array) $query['relevant'];
    $already_irrelevant = (array) $query['irrelevant'];

    $new_relevant = array_subtract($already_relevant, $relevant);
    $no_longer_relevant = array_subtract($relevant, $already_relevant);

    // get the descending scores
    $descend_irrelevant = retrieve_sentences_from_vector_query($irrelevant);
    //get the id of the least relevant (the last one)
    $top_irrelevant = array_shift(array_values($descend_irrelevant));

    $new_irrelevant = array_subtract($already_irrelevant, $top_irrelevant);
    $no_longer_irrelevant = array_subtract($top_irrelevant, $already_irrelevant);

    $relevant_vect = convert_sentence_IDs_to_sparse_vector($new_relevant);
    $no_longer_relevant_vect = convert_sentence_IDs_to_sparse_vector($no_longer_relevant);
    $irrelevant_vect = convert_sentence_IDs_to_sparse_vector($new_irrelevant);
    $no_longer_irrelevant_vect = convert_sentence_IDs_to_sparse_vector($no_longer_irrelevant);

    $positive_adjustment = $relevant_vect->scalarMultiply($ALPHA_w_plus);
    $negative_adjustment = $irrelevant_vect->scalarMultiply(-1*$ALPHA_w_plus);
    $no_longer_negative_adjustment = $no_longer_irrelevant_vect->scalarMultiply($ALPHA_w_plus);

    $adjustment = $positive_adjustment->vectorAdd($negative_adjustment);
    $adjustment = $adjustment->vectorAdd($no_longer_positive_adjustment);
    $adjustment = $adjustment->vectorAdd($no_longer_negative_adjustment);
    return $adjustment;
}
```

Get Top-most Irrelevant sentence

Without Normalization

No weighing terms for relevant and non-relevant sentences

Ide Regular

```
// vector adjustment: Ide regular
function calculate_sentence_adjustment_ide_regular($query, $vector_query, $relevant, $irrelevant){
    GLOBAL $ALPHA_w_plus;
    $already_relevant = (array) $query['relevant'];
    $already_irrelevant = (array) $query['irrelevant'];

    $new_relevant = array_subtract($already_relevant, $relevant);
    $no_longer_relevant = array_subtract($relevant, $already_relevant);

    $new_irrelevant = array_subtract($already_irrelevant, $irrelevant);
    $no_longer_irrelevant = array_subtract($irrelevant, $already_irrelevant);

    $relevant_vect = convert_sentence_IDs_to_sparse_vector($new_relevant);
    // $relevant_vect->normalize();
    $no_longer_relevant_vect = convert_sentence_IDs_to_sparse_vector($no_longer_relevant);
    // $no_longer_relevant_vect->normalize();
    $irrelevant_vect = convert_sentence_IDs_to_sparse_vector($new_irrelevant);
    // $irrelevant_vect->normalize();
    $no_longer_irrelevant_vect = convert_sentence_IDs_to_sparse_vector($no_longer_irrelevant);
    // $no_longer_irrelevant_vect->normalize();
```

Without Normalization

```
$positive_adjustment = $relevant_vect->scalarMultiply($ALPHA_w_plus);

$no_longer_positive_adjustment = $no_longer_relevant_vect->scalarMultiply(-1*$ALPHA_w_plus);
$negative_adjustment = $irrelevant_vect->scalarMultiply(-1*$ALPHA_w_plus);

$no_longer_negative_adjustment = $no_longer_irrelevant_vect->scalarMultiply($ALPHA_w_plus);

$adjustment = $positive_adjustment->vectorAdd($negative_adjustment);
$adjustment = $adjustment->vectorAdd($no_longer_positive_adjustment);
$adjustment = $adjustment->vectorAdd($no_longer_negative_adjustment);
return $adjustment;
```

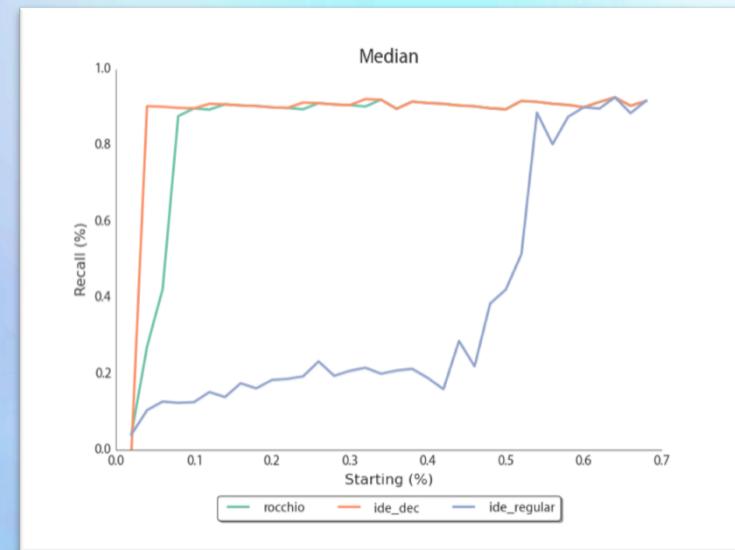
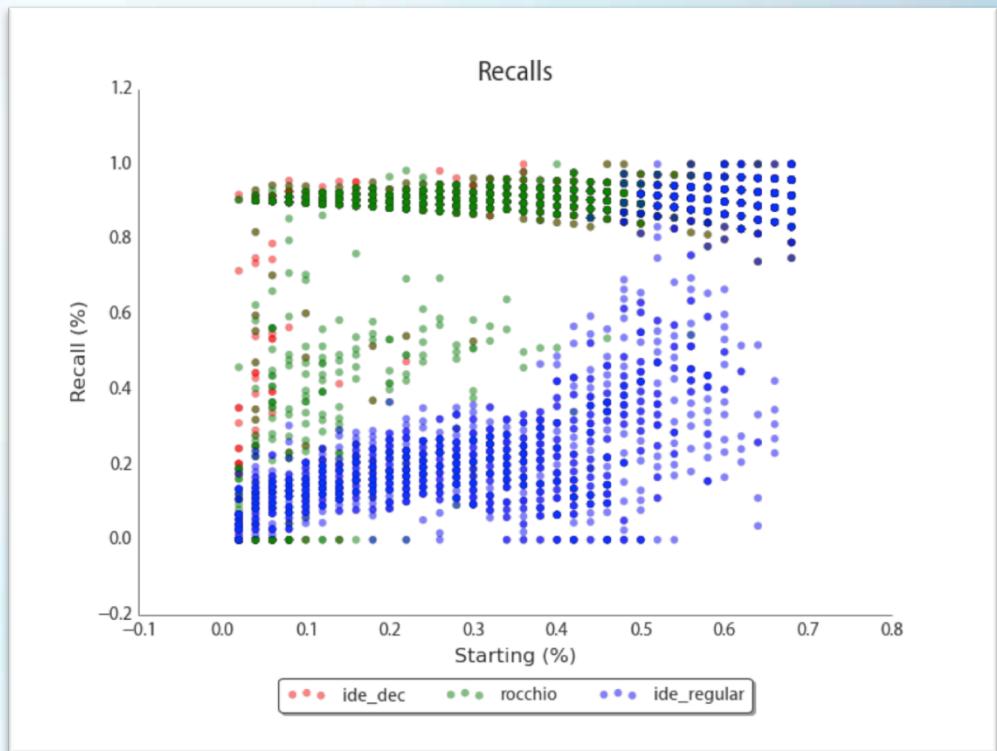
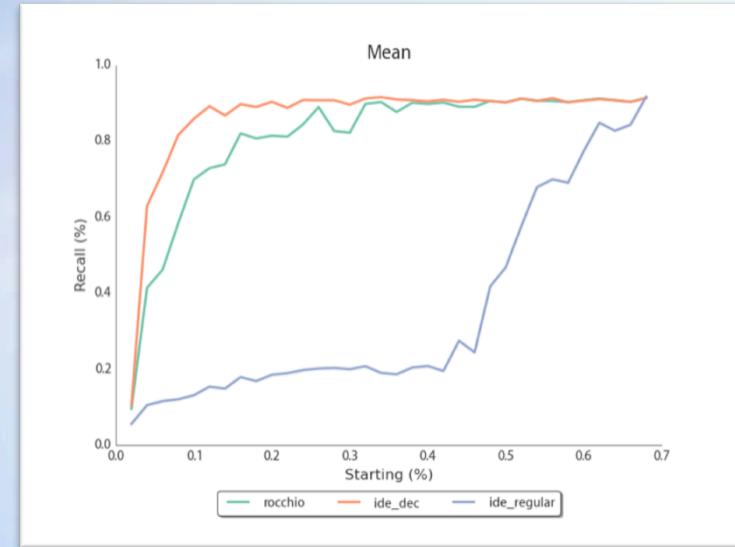
}

No weighing terms for relevant and non-relevant sentences

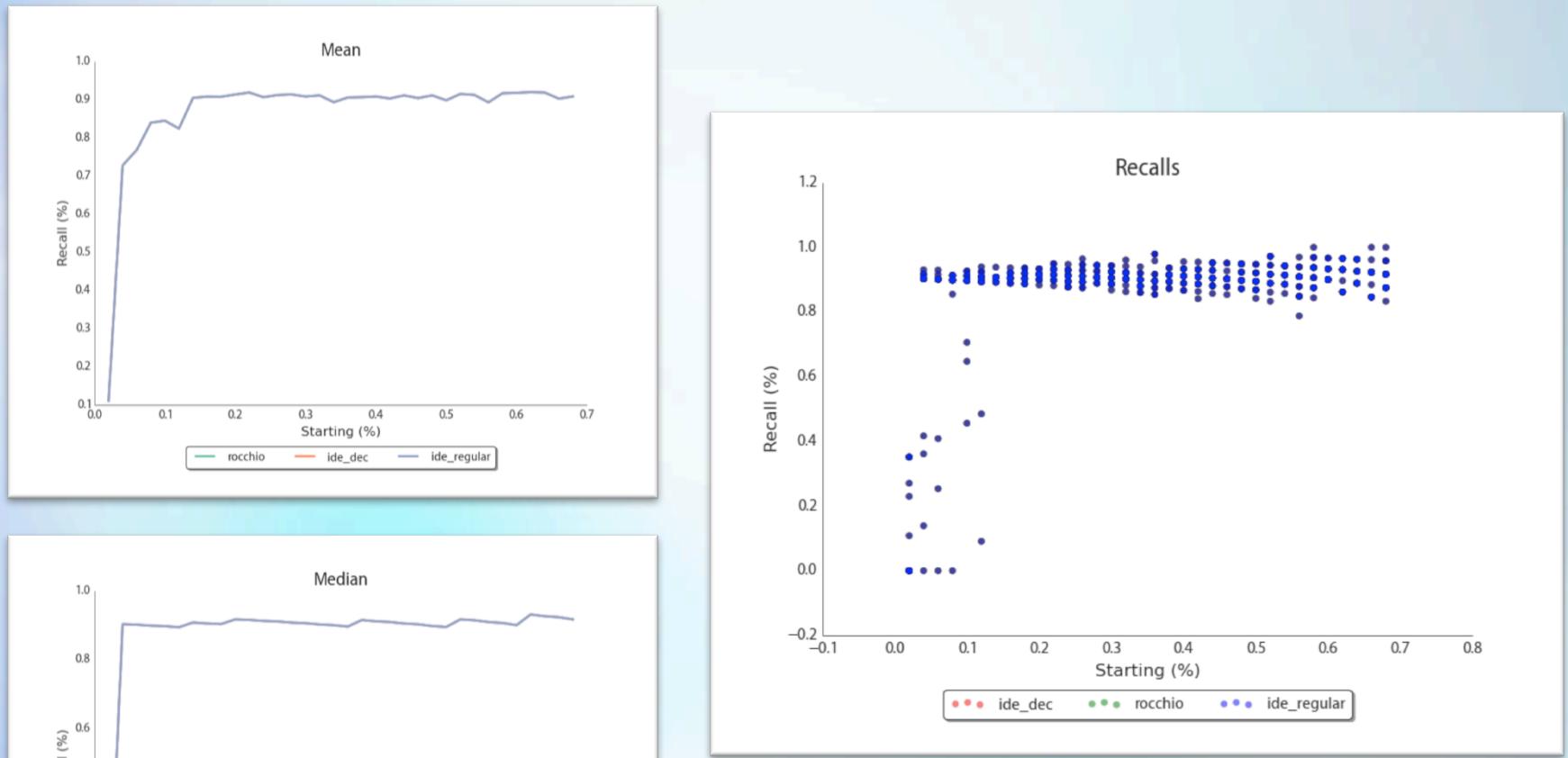
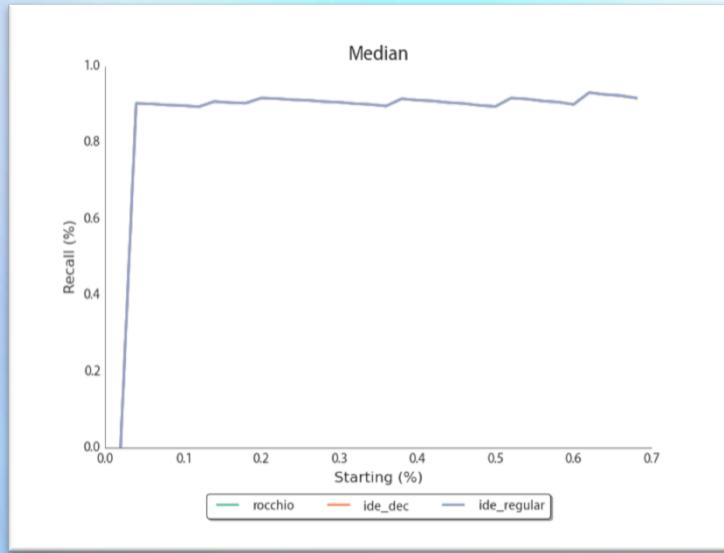
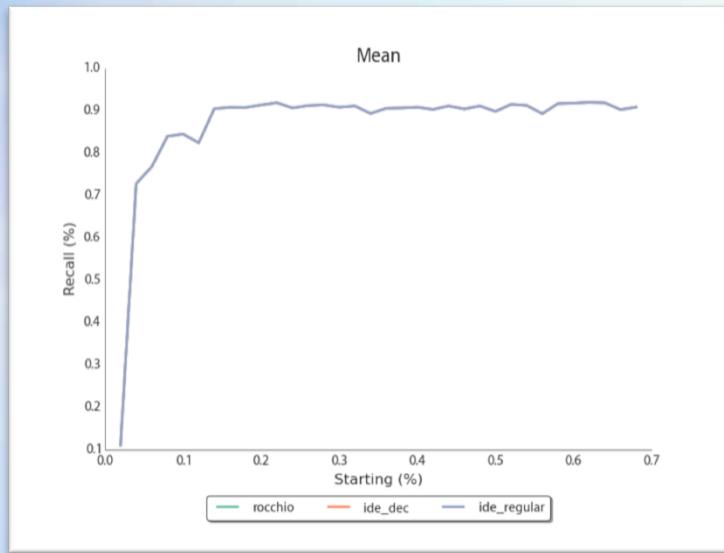
Testing Approach

- Get the sentence IDs from manually selected relevant and irrelevant sentences base on a selected keyword.
- Feed these sentence IDs into the relevance feedback algorithm
- Run

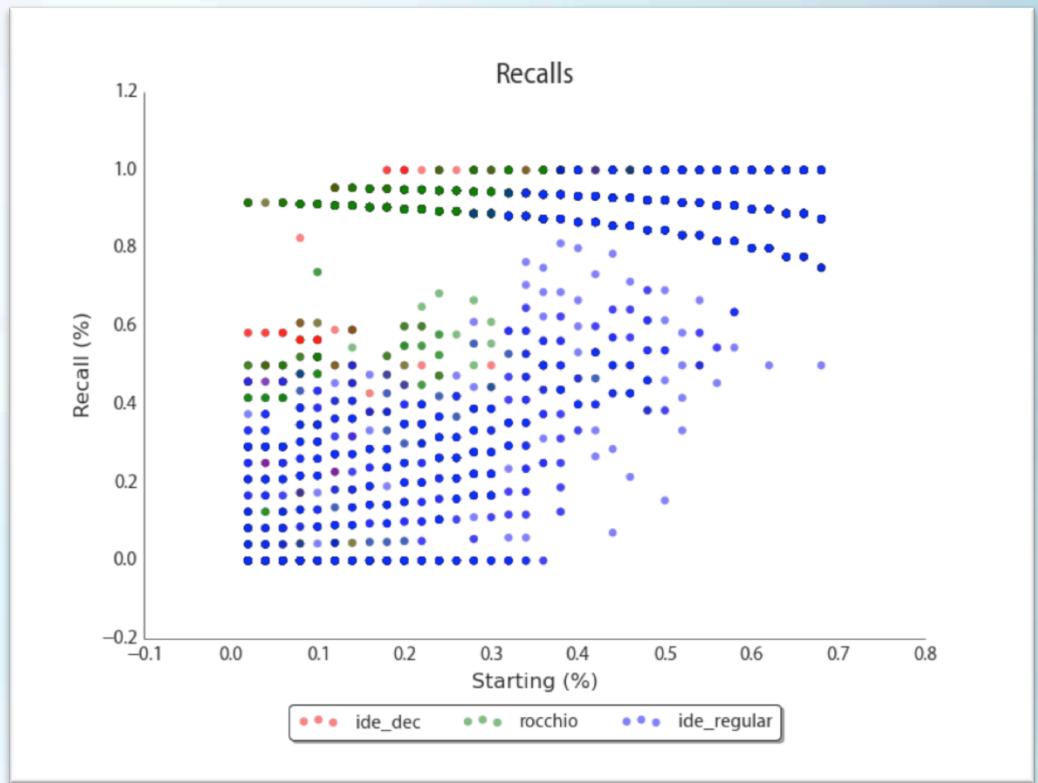
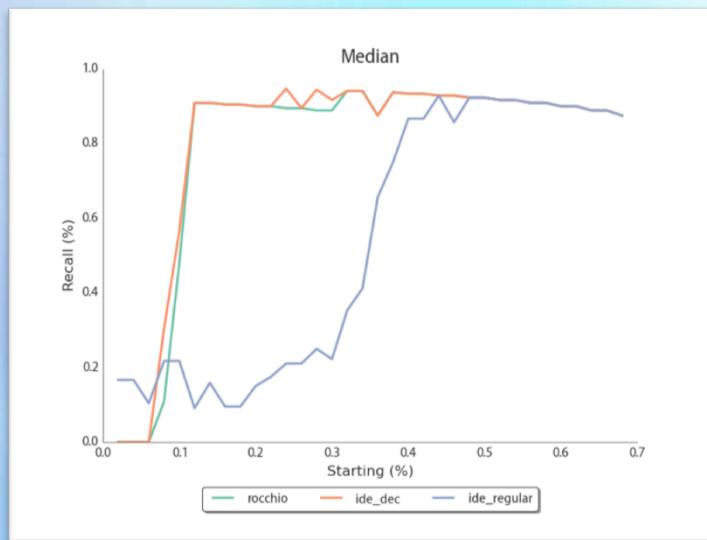
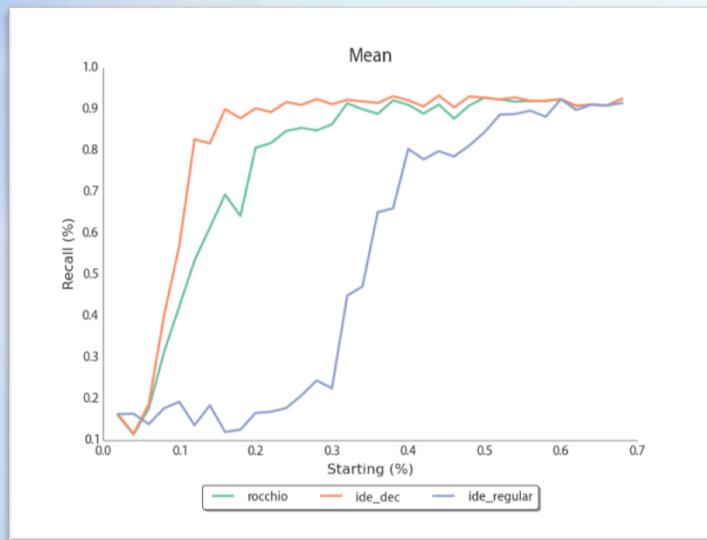
Black



Black – No Irrelevant



Curvy



References