Evaluations in Computer Science
Negative Results

Maria-Esther Vidal
Agenda

1. Scientific Method
2. Negative Results
3. NOiSE Program
High Level Approach for Evolving Knowledge [Basil97]
Scientific Method

According to Oxford English Dictionary: "a method or procedure that has characterized Natural Sciences since the 17th century, consisting in systematic observation, measurement, and experiment, and the formulation, testing, and modification of hypotheses".
Every baby knows the scientific method!

1. Make an observation.
2. Form a hypothesis.
3. Perform the experiment.
4. Analyze the data.
5. Report your findings.
6. Invite others to reproduce the results.
Every baby knows the scientific method!

1. Make an observation.
2. Form a hypothesis.
3. Perform the experiment.
4. Analyze the data.
5. Report your findings.
6. Invite others to reproduce the results.
Every baby knows the scientific method!

1. Make an observation.
2. Form a hypothesis.
3. Perform the experiment.
4. Analyze the data.
5. Report your findings.
6. Invite others to reproduce the results.
Every baby knows the **scientific method!**

1. Make an observation.
2. Form a hypothesis.
3. Perform the experiment.
4. Analyze the data.
5. Report your findings.
6. Invite others to reproduce the results.
Every baby knows the scientific method!

1. Make an observation.
2. Form a hypothesis.
3. Perform the experiment.
4. Analyze the data.
5. Report your findings.
6. Invite others to reproduce the results.
Every baby knows the scientific method!

1. Make an observation.
2. Form a hypothesis.
3. Perform the experiment.
4. Analyze the data.
5. Report your findings.
6. Invite others to reproduce the results.
Every baby knows the **scientific method!**

1. Make an observation.
2. Form a hypothesis.
3. Perform the experiment.
4. Analyze the data.
5. Report your findings.
6. Invite others to reproduce the results.
Scientific Research
Scientific Research

Observation

Part of the World

The World
Scientific Research

System or Formal Model of Ideas

Part of the World

The World
Scientific Research

Evaluation of the Theory

Validation of the System/Model of Ideas

Theory

System or Formal Model of Ideas

Observation

Part of the World

The World

NOiSE2015
Theory

According to Oxford English Dictionary:
Theory

According to Oxford English Dictionary:

- 
- 
- 
- 
-
Theory

According to Oxford English Dictionary:

“A scheme or system of ideas or statements held as an explanation or account of a group of facts or phenomena;”
Theory

According to Oxford English Dictionary:

- “A scheme or system of ideas or statements held as an explanation or account of a group of facts or phenomena;
- a hypothesis that has been confirmed or established by observation or experiment, and is propounded or accepted as accounting for the known facts;
Theory

According to Oxford English Dictionary:

▪ “A scheme or system of ideas or statements held as an explanation or account of a group of facts or phenomena;
▪ a hypothesis that has been confirmed or established by observation or experiment, and is propounded or accepted as accounting for the known facts;
▪ a statement of what are held to be the general laws, principles, or causes of something known or observed”.
Validation of a Theory

Formal Analysis + Empirical Evaluation = Validation of the Theory
Negative Result in Computer Science [Pechelt97]

- Something that **should have worked** does not.
- Negative Results are usually **camouflaged** as positive results
  - non-evaluating, mis-evaluating, or redefining the problem
- **Spurious** insights
  - Distorted evaluations
  - Positive results sometimes occur by chance
Evaluating a Research Project
Formal evaluation: state and prove theorems in some logic formal system.

**Extreme Value Theorem**

A continuous function on a closed interval always has an absolute maximum and minimum. They are located either at the end points or at critical values.

**Strategy:**
Find the critical values and then evaluate the original function at each critical value and at the end points. The highest value is the absolute max. The lowest value is the absolute min.

**Note:** Watch out for closed intervals and don’t forget the end points! If the problem says “find the maximum value,” it is probably referring to the absolute max.

“Where” is an x-value. “What” is a y-value.
Formal evaluation

- 
- 
- 
- 
- 
- 
- 

Formal evaluation

Formalization allows us to demonstrate:

•
•
•
•
•
•
•
Formal evaluation

Formalization allows us to demonstrate:
- Complexity of the problem.
Formal evaluation

Formalization allows us to demonstrate:
- Complexity of the problem.
- Optimality Conditions for the problem.
Formal evaluation

Formalization allows us to demonstrate:

- Complexity of the problem.
- Optimality Conditions for the problem.
- Properties of the proposed solution:
  - Soundness,
  - Completeness,
  - Optimality.
Goal of Formal Evaluations
Goal of Formal Evaluations

- Formal proofs:
Goal of Formal Evaluations

- Formal proofs:
  - Soundness and Completeness of the algorithms.
Goal of Formal Evaluations

- Formal proofs:
- Soundness and Completeness of the algorithms.
- Time and Space Complexity of the solution
  - Best, worst, and average cases.
Contra-examples to “theorems” or “properties”
Intractability – High Complexity
Undecidability
Exponential and Superpolynomial lower bounds
Empirical Evaluations
Anatomy of a Research Project*

1. Define the Research Questions
2. Formulate a hypothesis (or several)
3. Design an evaluation to test the hypothesis
4. Run the study
5. Analyze the results
6. Report and Publish

*https://www.merlin.uzh.ch/contributionDocument/download/6915
Anatomy of a Research Project*

Define the Research Questions

Formulate a hypothesis (or several)

Design an evaluation to test the hypothesis

Run the study

Analyze the results

Report and Publish

Evaluation of the Theory

*https://www.merlin.uzh.ch/contributionDocument/download/6915
Research Questions

Focus the study

Relate the theory and real-world scenarios

Determine where and what kind of research will be conducted

Allow for identifying specific objectives of the study
Experimental Set Up

Benchmarks
- Datasets
- Use Cases
- Tasks
- Size of the sample
- Description of the sample
- Type of Environment

Metrics or Observed Variables
- Efficiency
- Effectiveness
- Precision
- Recall
- AUC
- AUCPR
- Correlations

Statistical Methods
- T-student
- Wilcoxon Signed Rank Test
- Median or Average
- Chi-squared test

NOiSE2015
Experiment Configuration

Study Protocol

Number of times the experiment is replicated

Parameters that impact on the results

NOiSE2015
Experimental Study – Procedure Common in Natural Sciences

HYPOTHESES

EXPERIMENTS

NEGATIVE RESULTS

POSITIVE RESULTS
Negative or Inconclusive Results in Experimental Evaluations

- “Low/Bad” values of the observed variables.
- Observed Results are not statistically significant (high values of p-values)
  - Results provide little or no evidence that the null hypothesis is false.
  - No evidence that the null hypothesis is true.
- Low values or negative values of correlation
- High Values of Standard Deviation
Why can an Experimental Result be Negative?

- Large **number of parameters** impact on the outcome of an experiment
  - Unknown **relevant parameters** may be required
  - **Irrelevant parameters** are considered (multi factorial analysis)
- Heuristics and models based upon simple relationships or observations
  - No formal proofs can be provided
- The **experiment was not replicated** a large number of times (**sample** is not large enough)
Why should Negative Results be published?

- Avoid repeating the same experiments.
- Encouraging the definition of new hypotheses.
- Identifying new Challenges
  - Advancing the research areas!
14:10-15:30 Glorious failure session (4 x 10 min + 10 min discussion) (Chair: Miel Vander Sande)

- Heitmann, B., & Hayes, C. The Role of Negative Results for Choosing an Evaluation Approach - A Recommender Systems Case Study.
- Huelss, J., & Paulheim, H. What SPARQL Query Logs Tell and do not Tell about Semantic Relatedness in LOD.

15:30-16:00 Coffee Break
16:00-16:55 Confessions session
   (Chair Jacco van Ossenbruggen)

Interview with a witness (15 min)

- Kjetil Kjernsmo, How can scientific methods provide guidance for Semantic Web Research and Development?

Confession papers (2 x 2 min + Group discussion)
   (Chair Anastasia Dimou)

- Mutharaju, R., & Kapanipathi, P. Are We Really Standing on the Shoulders of Giants?
- Chapman, K., & Chapman, C. Breaking the Paper Barrier and Publishing At the Speed of Thought.

16:55-17:35 Breakout Discussion: Submission guidelines for reporting negative results

17:35-17:45 Report back

17:45 Closing
References


http://www.merlin.uzh.ch/publication/show/9417