Thoughts About (good) Research

Sergej Sizov
What is Our Mission?

Do good research!

Find truth!

Be relevant!

Achieve impact!
The Cycle in Which we Live…

Research → Publications → Dissemination → People & Money → Research

Achievements of research: a latent variable – not directly measurable … but what about indicators?!
Objectives and Achievements: Outline

1. Publication objectives
   - where and how to submit?
   - where not to submit?
   - how to measure success?

2. Dissemination objectives

3. People objectives

4. Money objectives
Part 1:
Publication Objectives
Peer review: general idea

Most scientists regarded the new streamlined peer-review process as ‘quite an improvement.’
Peer review: general idea (2)

given: set of reviewers $V = \{v_1, \ldots, v_k\}$, confidence grades $res(v_i, d)$ for submission $d$

collective result (restrictivity by thresholds $t_1$ and $t_2$, tuning by weights $w(v_i)$):

\[
\text{decision}(d) = \begin{cases} 
+1 & \text{if } \sum_i res_i(d) \cdot w(v_i) > t_1 \\
-1 & \text{if } \sum_i res_i(d) \cdot w(v_i) < t_2 \\
0 & \text{otherwise}
\end{cases}
\]

Special cases:
- “Unanimous Decision”
- “Voting”
- “Weighted Average” (e.g., weighted by some quality estimator)
General idea: accurate restrictive decisions

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>R</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>AA</td>
<td>AR</td>
<td>A0</td>
</tr>
<tr>
<td>R</td>
<td>RA</td>
<td>RR</td>
<td>R0</td>
</tr>
<tr>
<td>Junk</td>
<td>JA</td>
<td>JR</td>
<td>J0</td>
</tr>
</tbody>
</table>

committee decision

submissions

accepted papers

reduction

\[
\text{reduction} = \frac{|A0| + |B0| + |J0|}{|U|}
\]

error

\[
\text{error} = \frac{|AR| + |RA| + |JA| + |JR|}{|AA| + |RR| + |AR| + |RA| + |JA| + |JR|}
\]

junkred

\[
\text{junkred} = \frac{|J0|}{|J0| + |JA| + |JR|}
\]

loss

\[
\text{loss} = \frac{|A0| + |R0|}{|AA| + |RR| + |AR| + |RA| + |A0| + |R0|}
\]

tradeoff!
Peer review: a simple model

**Given:** set of reviewers $V = \{ v_1, ..., v_L \}$, binary decision (accept/reject)

**Wanted:** Approximations for *loss* and *error* for „unanimous decision“

$$X_i = \begin{cases} 
1 & \text{if } v_i \text{ assigns paper correctly} \\
0 & \text{otherwise}
\end{cases}$$

Probability of correct collaborative decision:

$$P(X_1 = 1, X_2 = 1|\bar{J}) = P(X_1 = 1|\bar{J}) \cdot P(X_2 = 1|\bar{J}) + \text{cov}(X_1, X_2|\bar{J})$$

$$P(X_1 = 1, ..., X_k = 1|\bar{J}) =$$

$$P(X_1 = 1|\bar{J}) \cdot \prod_{i=1}^{k-1} \frac{P(X_i = 1|\bar{J})P(X_{i+1} = 1|\bar{J}) + \text{cov}(X_i, X_{i+1}|\bar{J})}{P(X_i = 1|\bar{J})}$$

.. analogously we obtain $P(X_1 = 0, ..., X_k = 0|\bar{J})$

$$\text{junkred}(\text{Meta}) = 1 - P(X_1 = ... = X_L|J)$$

$$= 1 - [P(X_1 = 1, ..., X_L = 1|J) + P(X_1 = 0, ..., X_L = 0|J)]$$

$$\text{loss}(\text{Meta}) = 1 - P(X_1 = ... = X_L|\bar{J})$$

$$= 1 - [P(X_1 = 1, ..., X_L = 1|\bar{J}) + P(X_1 = 0, ..., X_L = 0|\bar{J})]$$

$$\text{error}(\text{Meta}) = \frac{P(X_1 = 0, ..., X_L = 0|\bar{J}) \cdot P(\bar{J}) + P(X_1 = ... = X_L|J) \cdot P(J)}{1 - \text{junkred} \cdot P(J) - \text{loss} \cdot P(\bar{J})}$$
Peer review: a simple model (2)

Example:
- Probability $p < 0.5$ to misassign doc from pos/neg for all k methods
- 50% of docs are Junk
- junkDoc is assigned to pos/neg with prob. 0.5
- $c < p(p-1)$ (no perfect correlation)

\[
\begin{align*}
\text{junkred} & = 1 - \left( \frac{c + \frac{1}{4}}{\frac{1}{2}} \right)^{L-1} \\
\text{loss} & = 1 - \left( 1 - p \right) \cdot \left( \frac{c + (1-p)^2}{1-p} \right)^{L-1} + p \cdot \left( \frac{c + p^2}{p} \right)^{L-1} \\
\text{error} & = \frac{1}{2} \cdot \left( \frac{p \cdot \left( \frac{c + p^2}{p} \right)^{L-1} + \left( \frac{c + \frac{1}{4}}{\frac{1}{2}} \right)^{L-1}}{\left( 1 - p \right) \cdot \left( \frac{c + (1-p)^2}{1-p} \right)^{L-1} + p \cdot \left( \frac{c + p^2}{p} \right)^{L-1} + \left( \frac{c + \frac{1}{4}}{\frac{1}{2}} \right)^{L-1}} \right)
\end{align*}
\]
Peer reviewing is not perfect!

Progress changes rules and ways of thinking!

Famous rejected papers:
  - B-trees
  - The first paper about the Web (Berners-Lee et al)
  - The first paper (Hendler et al) and the second paper (Fensel et al) about Semantic Web

See also:
S.Santini:
  We Are Sorry to Inform You..
  How much damage could be caused by a peer reviewer having a bad day?
  IEEE Computer, Dec 2005, pp. 126-128
Where to submit?

Events with

- peer review
- high visibility and impact in the community
- high restrictivity (low acceptance rate, 5-15 %)
- good organizers and reviewers

Sources of recommendations

- your supervisor and colleagues
- Microsoft Libra
- Australian Ranking of ICT Conferences
Differences in publication culture

Computer Science
- Peer-reviewed conferences
- Top conferences have 5-15% acceptance rate
- Specialized and small conferences (attendance of 500+)
- Often value conferences > journals

Pure Sciences (eg, Math, Physics)
- Pre-print at Arxiv.org
- Rigorous reviews for journals
- Huge flagship conference (ICM 98 attracted ~4000)

Social Sciences
- Often value journals > conferences
- Conferences are mostly for gathering or short abstract
- based screening
- Rigorous reviews for journals
Where not to submit?

Bogus conferences

- Known conferences and journals of (very) dubious reputation
- Nagib-Callaos-Conferences, Khalid-Soliman-Conferences
- Blacklists are impossible to keep up (threats by organizers, e.g. fakeconferences.org)

Indicators: curious OC and PC, fake venues, missing or questionable reviewing process, paper presentation not required.

see also:
SCIgen - An Automatic CS Paper Generator
Bogus conferences

To name just a few..

- IMCSE: International Multiconference in Computer Science and Computer Engineering
- WMSCI or SCI: World Multiconference on Systemics, Cybernetics and Informatics
- ICCCT: International Conference on Computing, Communications and Control Technologies
- PISTA: Conference on Politics and Information Systems: Technologies and Applications
- SSCCI: Symposium of Santa Caterina on Challenges in the Internet and Interdisciplinary Research
- CITSA: International Conference on Cybernetics and Information Technologies, Systems and Applications
- ISAS: International Conference on Information Systems Analysis and Synthesis
- CISCI: Conferencia Iberoamericana en Sistemas, Cibernética e Informática
- SIECI: Simposium Iberoamericano de Educación, Cibernética e Informática
- WCAC: World Congress in Applied Computing
- Any IPSI International Conference or journal
- Any GESTS international conference or journal
- KCPR: International Conference on Knowledge Communication and Peer Reviewing (!!)
- International e-Conference on Computer Science
- …
Publications: success indicators

**visibility**: we publish on high-quality conferences (A+, A, B)
- check DBLP profile against CORE

**impact**: our work is frequently cited by others
- check Google scholar, Citeseer, ..
- evaluate with Publish or Perish

**reproducibility**: method details, tuning parameters, evaluation datasets, libraries, etc. are documented and available to public

**reliability**: our models are correct, evaluation is consistent
Visibility: check the DBLP Profile!

2007


- Lyublana Antova, Christoph Koch, Dan Olteanu: MayBMS: Managing Incomplete Information with Probabilistic World-Set Decompositions. ICDE 2007: 1479-1480

- Michael Schmidt, Stefanie Scherzinger, Christoph Koch: Combined Static and Dynamic Analysis for Effective Buffer Minimization in Streaming XQuery Evaluation. ICDE 2007: 236-245

- Lyublana Antova, Christoph Koch, Dan Olteanu: 1018 Worlds and Beyond: Efficient Representation and Processing of Incomplete Information. ICDE 2007: 606-615


- Christoph Koch: XPath Leashed. PLAN-X 2007: 0-1

2006

- Walker M. White, Alan J. Demers, Christoph Koch, Johannes Gehrke, Rajmohan Rajagopalan: Scaling games to epic proportion. SIGMOD Conference 2007: 31-42


- Lyublana Antova, Christoph Koch, Dan Olteanu: Query language support for incomplete information in the MayBMS system. VLDB 2006: 1422-1425


Impact: try Publish or Perish!

Citation metrics: example

- **h-Index**: $N_p$ papers have at least $h$ citations each, and the other $(N_p-h)$ papers have no more than $h$ citations each.


- **g-Index**: the (unique) largest number such that the top $g$ articles received (together) at least $g^2$ citations (gives more weight to highly cited articles)

Reproducibility

For better transparency and reproducibility:

- Provide software (evaluation methods and libraries used) to public
- Make also evaluation datasets available (e.g. online)
- Contribute to challenges and evaluation initiatives (e.g. TREC, CLEF, Semantic Web Challenge, Billion Triple Challenge etc.)
- Provide advanced versions of your contributions with full proofs, explanations, and complete experiments (technical reports, journal papers, online material)
Reliability

Try to avoid tricks and hacks (if possible.. you know..)

Adopt established evaluation methods and best practices of the target community
- systematic experiments with large-scale reference data sets are very welcome in many communities
- example: UCI Machine Learning Repository

Statistical significance is often important..

Ensure testbed transparency and reproducibility (as discussed before)

Clearly identify the novelty, originality, and expected benefits of your work wrt existing stuff! Seven out of ten PhD status applications fail at this point
Part 2:
Dissemination Objectives
Dissemination Objectives

Ensure visibility of you and your work!

Be present at relevant events
(.. ideally with accepted contributions :-)

Consider to submit an advanced journal version after conference acceptance

Make interesting software prototypes available to public and keep them up to date
  • successful examples: SVM light, Snowball, Mallet..

Contribute to evaluation initiatives and challenges
Part 3:
People Objectives
People objectives / Networking

Organisation of workshops

- what are the key ones?
  → New themes!

- Reviewing
  → Reviewers are also cited!

- Summer schools
  - learning
  - teaching

Communities

- Thematically focused
  - E.g. EKAW, BTW..

- National Communities
  - GI focused groups

- Evaluation activities

- Standardization activities

- EU / DFG projects

- Special interest groups

- Open source prototypes and libraries, public datasets, etc.

Supervision of students

(lectures, projects, diploma thesis) is also important
Presentation Skills
and
Scientific Writing

Sergej Sizov
Scientific writing: major points

Important points:

Subject
Purpose
- to exchange the scientific knowledge
- to ask and answer specific questions

Audience
- scientists and those interested in the subject
- a publisher or an editor
The principle of the diamond

1. Introduction
2. Related Work
3. Your Message
   - Methods
   - Evaluation: methodology, results
   - Discussion
4. Conclusion and Future Work
5. References
The other side: reviewing

Be always polite, provide constructive criticism and positive recommendations!

Key aspects:
- Novelty / originality, step beyond state of the art
- Technical quality and depth
- Relevance / appropriateness for target audience
The other side: reviewing

Be always polite, provide constructive criticism and positive recommendations!

Paper analysis in detail

- Structure and clarity, quality of text / charts
- Study of related work
- Appropriate formalization
- Run-through examples
- Evaluation: goals, methodology, testbed, result presentation, interpretation
- Reproducibility of results
The other side: reviewing

Be always polite, provide constructive criticism and positive recommendations!

Possible recommendations

- Accept the contribution, possibly with minor changes
- Major revision, reject and resubmit later
- Reject, do not submit further, think first what you are doing
- Submit to other conference (which ones)
- Submit as a workshop paper (which workshop)
- Submit as short paper or poster

Overall recommendation should be aligned with discussion and scores you give to the contribution.
Presentation skills: before we start:

What is a successful presentation for you?

When have you seen a really good presentation? Why do you think that presentation was good?

Can you establish other criteria from presentations you have seen?

A present
Getting started

You want to present your work to an audience in X

Define your audience
  • expert; non-expert; mixed

Define your time
  • fixed time limit: seems long, but usually too short

Define your environment
  • accommodate - in a strange room - to the equipment (beamer, microphone, board)
  • have back-ups (power supply, memory stick, CD, handouts, board, …)

Define your design
  • logo, name of the institute, colour, layout, structure, …
Accept that you are probably going to be nervous

Find your own solution(s):
- Something to drink
- Deep breathing
- Go for a walk
  ⇒ fresh air
- ...

The only effective remedy: Accept it. Have strategies!
Strategies when nervous (1)

Know your slides:
- try to present your talk to friends etc. before presenting it at a conference
- don’t finish preparation one minute before your talk starts

Think about your equipment:
- notebook: hotkey for external monitor, beamer resolution, beamer, disable screen saver, remote control, presenter/mouse, laser pointer, power supply+adapter, light (where is the dimmer? off/on?)
- overhead projector (where is the switch?), pointer, board, pen ...

The most effective strategy: Know the first minute of your talk – your introductory material / your first slide(s)- off by heart
Strategies when nervous (2)

Show your *title / first* slide

- eye contact to (or neck region of) audience, greet, title, your name, where you work, smile

⇒ audience turns towards your slide and doesn’t focus on you

Show your *second* slide: overview of your presentation

- give informative overview

- don’t: Overview
  Introduction
  Details
  Conclusion
  References

Continue with your presentation/slides
Strategies when nervous (3)

**Voice**: Don’t worry if it is cracking or squeaking
- Try to speak clear and slowly, loud enough

**Blushing**: Ignore
- the audience is watching your slides!

**If reading**: Format your notes for easy path finding

Try to **look at the audience** as much as possible
- You are the expert, observe them

**Give presentations** as often as possible

Be aware of what people will do when they are **nervous**
- Find out what you do -> work on it
don’t excuse yourself!

... using notes
... graph not nice
... spelling errors
... etc.
The major fault with many presentations is:

The structure of the material doesn’t harmonize with the visual aids or the way you deliver it.

Common mistakes:
- Unstructured facts
- No obvious framework
- The audience becomes disoriented
Be very selective. You can’t include everything:
Structure material as a diamond of detail:

1. Title slide
2. Introductory overview slide
   brief summary: place your work in context, give the big picture; why don’t put results in here? **Tell them what you are going to do**
3. Place work in context
   good summary of methods, results and conclusion
4. The detail
   then/here give detail; **tell them**
5. Concluding overview slide
   of what your work means: Your conclusions and further directions. **Tell them what you’ve told them**
Slides for the basic structure

Title slide
Overview slide
Detail
Conclusion slide
Presenting: Pausing and Interruptions

If you lose your place or have to pause, say nothing:
- Control your body language, find your place again

Interruption beyond your control, say

“I’ll repeat that…”

or continue without hesitation

A pause for the speaker seems longer than for the audience

Pause because of thinking: look at someone in audience; not at ceiling or floor

Interruption = question(s)
Presenting: Finishing the Presentation

Be professional:
- Put up your concluding slide and BRIEFLY describe your conclusions
- Then just slightly nod your head and say "Thank you"
  ⇒ job of the chair to ask for questions
- Don’t fluster

Common mistakes:
- Nervous -> body language
  "Well, that’s all I’ve got to say, really"
  "That’s it – so – um – thank you."
  "I’d like to thank you for your kind attention in listening to this talk. " (too artificial)
Demonstrating: Finishing in a hurry (1)

- Don’t fluster
- Smoothly finish the sentence you are saying
- Say something like
  “I’m sorry I don’t have time to give you the details here. But I’d be pleased to talk to anybody afterwards.”
- Put up your conclusion slide
- Say
  “And so, in conclusion, …”
- If time is very short
  put up Conclusions slide for audience to read
The other side

speaker

chairman

audience
The other side: to be the chairman (1)

The chairman is in charge of controls

- objectiveness
- efficiency
- time
  - of speaker
  - of whole event
The other side: to be the chair(wo)man (2)

It is the job of the chairman:

- To stand in front at the beginning and at the questions‘ session. Be competent / confident (no hands in pockets)
- To introduce the speaker, the topic / title (+ where s/he works, etc.)
- That the talk and questions session go smoothly
  - To ask questions him/herself (important if there are no questions)
  - To admit to the floor (one questions after the other)
- To clarify incomprehensible (even inaudible) questions
- To summarize (if necessary)
- To conclude the talk and thank the speaker and audience / continue with next speaker
The other side: ask questions

Always ask questions and give comments:
- To clarify what you did not understand
- To recommend something
- To add new / unknown / important material
- To give another viewpoint
- …

If you don’t have questions at all, ask anyway about:
- Why is this work / research important
- How about costs / real time / etc.
- Future work / direction (if not stated before)
- …
The other side: feedback (1)

Feedback is important

- because it is like the applause for the actor on stage
- because you learn from your own mistakes
- always give feedback
- (unfortunately: no feedback culture)

‘Who plays up to me is my enemy,
who blames me is my teacher.’
**The other side: feedback (2)**

**How** to give feedback:

Start with the positive / good things (on the talk, etc.)
Move on to the things the presenter could have done better
  - Don’t say: This and that was very bad / idiotic / stupid
  - But: You could improve here…; it wasn’t too bad, but you could do better if…

Put your criticism in a positive way! Always praise, but make (necessary) improvements clear!
Thank-you slide