Annotated Bibliography For Senior Thesis

Tom Silver
Harvard University
tsilver@college.harvard.edu

Stuart Shieber
Harvard University
shieber@seas.harvard.edu

Abstract

TODO

Date: September 15, 2015

Table of Contents

1 The Turing Test **3**
1.1 Philosophy of the Turing Test 3
1.2 Practical Implementations of the Turing Test 3
1.3 Proposed Alternatives to the Turing Test 3

2 General Test Design **3**
2.1 Inducement Prizes 3
2.2 Rating Systems 3

3 Training Judges **3**
3.1 Learning to Rank 4

1 The Turing Test

Overview: [22]

1.1 Philosophy of the Turing Test

Beginning with Turing's original paper [24], this collection of papers examines the philosophical underpinning of the Test. These papers attempt to rigorously define intelligence and determine whether the Turing Test reflects the proposed definitions.

Related citations: [24], [23], [20], [16], [17], [10], [6]

1.2 Practical Implementations of the Turing Test

These papers characterize attempted practical implementations of the Turing Test, such as the Loebner Prize. The collection also includes papers that formalize the Turing Test mathematically and explore its properties as a practical test.

Related citations: [26], [21], [14], [1]

1.3 Proposed Alternatives to the Turing Test

In this collection, authors consider alternatives to the Turing Test, i.e. other tests that aim to evaluate the intelligence of a machine. There have been a number of recent seminars and informal discussions on the subject; these will also be included here.

Related citations: [18], [15], [13], [12], [5]

2 General Test Design

The centerpiece of this thesis is a new test for intelligence. The design of this test needs to be informed by general test design principles, which apply regardless of the quality that is under consideration.

2.1 Inducement Prizes

These papers survey existing inducement prizes, including their triumphs and pitfalls. They consider the ultimate goals of such prizes and attempt to elucidate general principles for successful inducement.

Related citations: [19]

2.2 Rating Systems

The proposed test in this thesis does not output a binary value for intelligence, but instead assigns a rating that reflects the relative intelligence of the evaluated subject relative to other subjects in the system. One closely related rating system is the Elo rating system used in chess. These papers consider the Elo rating system and other related systems. The statistical foundation of Elo is of primary interest.

Related citations: [11], [9], [8], [2], [7]

3 Training Judges

An interesting machine learning problem that has immediately emerged from the proposed competitive rating system is that of training judges. A machine in our system needs an algorithm for

estimating the intelligence ratings of its competitors. Papers in this section suggest various avenues for approaching this problem.

3.1 Learning to Rank

One subfield of machine learning that has recently received a lot of attention is the problem of learning rank functions. While the problem is not exactly that of training a judge, there may be parallels between the two problems and perhaps inspiration to take from the recent progress in the former.

Related citations: [25], [4], [3]

References

[1] Phillip G Bradford and Michael Wollowski. A formalization of the turing test. *ACM SIGART Bulletin*, 6(4):3–10, 1995.

[2] Ralph Allan Bradley and Milton E Terry. Rank analysis of incomplete block designs the method of paired comparisons. *Biometrika*, 39(3-4):324–345, 1952.

[3] Zhe Cao, Tao Qin, Tie-Yan Liu, Ming-Feng Tsai, and Hang Li. Learning to rank: from pairwise approach to listwise approach. In *Proceedings of the 24th international conference on Machine learning*, pages 129–136. ACM, 2007.

[4] Olivier Chapelle and Yi Chang. Yahoo! learning to rank challenge overview. In *Yahoo! Learning to Rank Challenge*, pages 1–24, 2011.

[5] Paul R Cohen. If not turing's test, then what? *AI magazine*, 26(4):61, 2005.

[6] Jamie Cullen. Imitation versus communication: Testing for human-like intelligence. *Minds and Machines*, 19(2):237–254, 2009.

[7] Arpad E Elo. *The rating of chessplayers, past and present*. Arco Pub., 1978.

[8] Mark E Glickman and Albyn C Jones. Rating the chess rating system. *CHANCE-BERLIN THEN NEW YORK-*, 12:21–28, 1999.

[9] Trevor Hastie, Robert Tibshirani, et al. Classification by pairwise coupling. *The annals of statistics*, 26(2):451–471, 1998.

[10] Patrick Hayes and Kenneth Ford. Turing test considered harmful. In *IJCAI (1)*, pages 972–977, 1995.

[11] James P Keener. The perron-frobenius theorem and the ranking of football teams. *SIAM review*, 35(1):80–93, 1993.

[12] Hector J Levesque. Is it enough to get the behaviour right. *Proc. of IJCAI-09, Pasadena, CA*, 2009.

[13] Hector J Levesque, Ernest Davis, and Leora Morgenstern. The winograd schema challenge. In *KR*, 2012.

[14] Paweł Łupkowski and Andrzej Wiśniewski. Turing interrogative games. *Minds and Machines*, 21(3):435–448, 2011.

[15] G Marcus. What comes after the turing test? *New Yorker*, 2014.

[16] James H Moor. An analysis of the turing test. *Philosophical Studies*, 30(4):249–257, 1976.

[17] James H Moor. The status and future of the turing test. *Minds and Machines*, 11(1):77–93, 2001.

[18] Mark O Riedl. The lovelace 2.0 test of artificial creativity and intelligence. *arXiv preprint arXiv:1410.6142*, 2014.

[19] Alex Schroeder. The application and administration of inducement prizes in technology. *Independence Institute Research Paper*, 2004.

[20] John R Searle. Minds, brains, and programs. *Behavioral and brain sciences*, 3(03):417–424, 1980.

[21] Stuart M Shieber. Lessons from a restricted turing test. *arXiv preprint cmp-lg/9404002*, 1994.

[22] Stuart M Shieber. *The Turing test: verbal behavior as the hallmark of intelligence*. Mit Press, 2004.

[23] Stuart M Shieber. The turing test as interactive proof. *Noûs*, 41(4):686–713, 2007.

[24] Alan M Turing. Computing machinery and intelligence. *Mind*, pages 433–460, 1950.

[25] Hamed Valizadegan, Rong Jin, Ruofei Zhang, and Jianchang Mao. Learning to rank by optimizing ndcg measure. In *Advances in neural information processing systems*, pages 1883–1891, 2009.

[26] Joseph Weizenbaum. Elizaa computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45, 1966.