

Emergence of Terminological Conventions as an Author-Searcher Coordination Game

David Bodoff
University of Haifa
Graduate School of Management
Haifa, Israel
+972 (0)4-8249579
dbodoff@gsb.haifa.ac.il

Sheizaf Rafaeli
University of Haifa
Graduate School of Management
Haifa, Israel
+972 (0)4-8249578
sheizaf@rafaeli.net

ABSTRACT

All information exchange on the Internet – whether through full text, controlled vocabularies, ontologies, or other mechanisms – ultimately requires that an information provider and seeker use the same word or symbol. In this paper, we investigate what happens when both searchers and authors are dynamically choosing terms to match the other side. With each side trying to anticipate the other, does a terminological convention ever emerge, or do searchers and providers continue to miss potential partners through mis-match of terms? We use a game-theoretic setup to frame questions, and learning theory to make predictions about whether and which term will emerge as a convention.

Categories and Subject Descriptors

H.3.3 [Information Search and Retrieval]: Query formulation;
H.3.1 [Content Analysis and Indexing]: Indexing methods

General Terms

Experimentation, Human Factors, Standardization

Keywords

Game theory, Learning, Manual indexing

1. INTRODUCTION

On the Internet, producers and seekers of information come together through one primary mechanism: their use of the same words. This applies in the widespread case of search engines, which index the author's words and process the searcher's query. But it also applies where the vocabulary is controlled rather than free-text. And it applies to ontologies where categories are labeled using words, which are then browsed or searched by the person or agent seeking information. It also applies to both human and robotic providers and seekers of information. In all cases, the producer and seeker of information come together through the use of identical symbols or "words".

Our research is about how providers and seekers of information on a given topic learn through experience to use identical words to represent it. Our contribution is in our simultaneously modeling the two complementary roles of seeker and provider, and how both sides choose words to try to match the other. The literature on "relevance feedback" is about just one of these, i.e. how information *searchers* learn from experience which words to use. But Internet technologies have made the opposite role of content authoring and indexing increasingly dynamic and strategic. Content authors consciously choose words to include in their

Webpages, which they think will be used by searchers. Similarly, keyword advertisers consciously choose to advertise on keywords that they think will be used by searchers. Thus, searchers and providers are trying to anticipate the word usage of one other.

Early in the development of the field of information retrieval, query term modification and document term modification were viewed as complementary; the challenge was how to do both simultaneously, in a probabilistically coherent way [1]. In this work, we address the behavioral version of this question. Here the question is not what our algorithms should do to account for both perspectives, but what human players actually do, when they play the roles of dynamic authors and searchers. Both sides receive feedback, and both have the opportunity to modify their choice of words to better match the other. But if both sides modify their choice to suit the other, then nothing will have been gained! The goal of our research is to characterize the simultaneous process through which authors and seekers learn to modify their choice of terms to match one another. An example of a related practical question is: Should keyword advertisers constantly change their choice of terms to match the term that is used by more interested users, or will this moving target only make it more difficult for a terminological convention to emerge?

2. THEORY AND METHODS

We frame the above questions in game-theoretic terms, and apply theories of learning to predict the outcome. The basic game is as follows: There are two populations, representing searchers and authors/advertisers, respectively. There is a single topic, but it can be represented by two different keywords. In each round of the game, each player from among the searchers is randomly paired with a random, anonymous partner from among the authors/advertisers. They are both asked to choose one of the two terms to represent the information being offered (in the case of the author) or sought (in the case of the searcher). If they choose the same term, they both "win"; otherwise, they both lose. Following van Huyck [2] we call this the "convention game", since neither side particularly cares which word is used, only that it matches the partner. After receiving feedback about whether their partner for that round had chosen the same term, the game proceeds to the next round, in which each player is paired with a random new partner. Using standard game-theoretic notation, the game is as shown in Table 1. Each cell shows the payoffs for the two players under a given combination of the choices they make.

Table 1 Basic Convention Game

		Term chosen by indexer	
		"A"	"B"
Term chosen by searcher	"A"	1, 1	0, 0
	"B"	0, 0	1, 1

The question is, will the whole “system” of players eventually converge on a single term? Or will it remain a perpetual guessing game, with players sometimes missing and sometimes matching their partner? What are the factors that influence whether -- and which term -- will emerge as the convention to represent a given topic? Although we have framed the issue in terms of a game, pure game theory makes no predictions about such a case, in which there are two identical Nash equilibriums. But theories of evolutionary learning or individual learning do. A series of papers by van Huyck et al. [2, 3] studies a generic game of this form. Using theories of evolutionary learning, they find that players converge on a choice, and that as between convergence to “A” versus “B”, this depends on the fraction of players who make each choice in the first round. This is the basis for our similar predictions in our case in which the task is to choose between two terms to represent a given text.

We also extend the basic game to consider that there are documents and searchers about more than one topic which use similar words. In this game, the populations of searchers and providers are each subdivided into two types. Each type is choosing terms for a different topic. In this game, a seeker and provider wish to match terms, but only if they are of the same type, i.e. dealing with the same topic. If they are dealing with two different topics, then they emphatically do not wish to match one another’s choice of terms. Whereas the basic game considered only recall (Type II error), this game also considers precision (type I error). This game is depicted in figure 2. We refer to it as the “separation game”, since equilibrium requires that all players of one type (i.e. topic) use one term (either one), and all players of the other type use the other term. This game has a more complex state space than the basic convention game, so it is not known whether players will “figure out” how to converge. We also propose a new theory for predicting which (if any) term will “go with” each topic in equilibrium. This constraint-based approach models that for some topics, one term is significantly more appropriate than any other, while for other topics, the choice is more arbitrary. The constraint-satisfaction approach predicts that the topics for which one term is far better will anchor the solution, with the other topics being represented by the leftover terms.

Table 2 Separation Game

Searcher		Indexer			
		Assigned to Text 1		Assigned to Text 2	
		Chooses term “A”	Chooses term “B”	Chooses term “A”	Chooses term “B”
Assigned to Text 1	Chooses term “A”	1, 1	0, 0	-1, -1	0, 0
	Chooses term “B”	0, 0	1, 1	0, 0	-1, -1
	Assigned to Text 2	Chooses term “A”	-1, -1	0, 0	1, 1
	Chooses term “B”	0, 0	-1, -1	0, 0	1, 1

3. EXPERIMENTS AND RESULTS

We ran experiments of the basic and extended game types. In each session, 16 subjects played five 30-round computerized games of one of the 2 game types. Typical results of the basic convention game are shown in figure 3. Each line represents a 30-round game based on one text. The y-axis represents the percentage of all players who were choosing the labeled keyword over the alternative shown in parentheses. Convergence to equilibrium is clear, as is the influence of the starting point on determining the final outcome. We found that given the initial

choices, other factors such as how well each term actually fits the topic, didn’t influence the outcome.

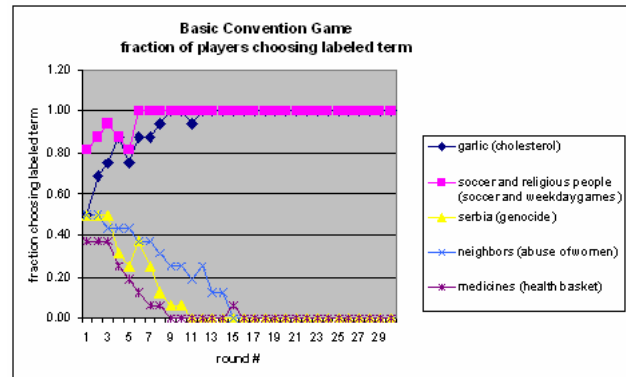


Figure 3 Results of 5-game session of basic convention game

Regarding the separation game, convergence was even more rapid even though this game is formally more complex. Players from each topic congregated around one of the two keywords and thereby separated. Regarding which word was adopted for which topic, this was determined by the predicted pattern. For example, figure 4 shows a case where the term “garlic” was favored for each of two topics, as can be seen in players’ first-round choices. But players working with the topic of garlic and cholesterol quickly switched to the term “cholesterol”, because players with the topic of garlic and health couldn’t be expected to be the ones to switch and use the inappropriate (to them) term “cholesterol”.

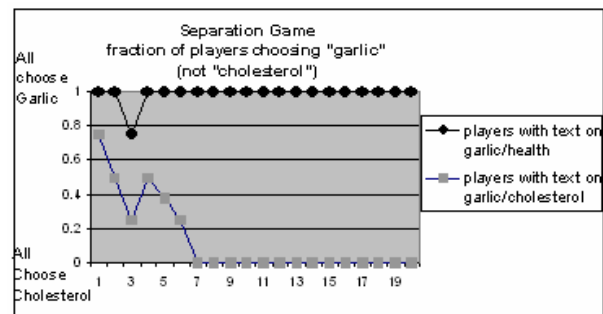


Figure 4 Results of one separation game

4. CONCLUSION

Our research provides a framework for predicting the evolution over time of which words will be used by searchers and authors/advertisers to represent different topics. A key idea is that this mapping of topic to keyword is expected to evolve, and according to predictable patterns.

5. REFERENCES

- Robertson, S.E., M.E. Maron, and W.S. Cooper, Probability of Relevance: A Unification of Two Competing Models For Document Retrieval. Information Technology -- Research and Development, 1982. 1: p. 1-21.
- van Huyck, J.B. R.C. Battalio, and F.W. Rankin, On the Origin of Convention: Evidence from Coordination Games. The Economic Journal, 1997. 107(442): p. 576-596.
- van Huyck, J.B., R.C. Battalio, and R.O. Beil, Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure. The American Economic Review, 1990. 80(1): p. 234-248