The popularity heuristic: Using search query data for forecasting

Andreas Graefe
Institute for Technology Assessment and Systems Analysis
Forschungszentrum Karlsruhe, Germany
graefe@kit.edu

J. Scott Armstrong
The Wharton School
University of Pennsylvania, Philadelphia, PA
armstrong@wharton.upenn.edu

August 13, 2009

Abstract. We used search query data to measure aggregated name recognition of tennis players. Then, applying the recognition heuristic, we generated forecasts to predict the outcomes of the 1,016 matches of the eight Grand Slam tennis tournaments in 2007 and 2008 and compared them to forecasts based on ATP rankings and betting odds. The search query forecasts correctly predicted 70% of the matches and were only slightly inferior to the ATP rankings (72% correct predictions); betting odds were most accurate (77 to 79%). In addition, consensus forecasts of the three approaches improved accuracy. In relying on the recognition heuristic and providing different information, search query data has the potential to improve forecasting.

1. The recognition heuristic

Goldstein and Gigerenzer (1999) gave American students from the University of Chicago pairs of the 22 largest American cities (such as Los Angeles versus San Diego) and the 22 largest German cities (such as Munich versus Berlin) and asked them which of the two cities has the larger population. The students recognized all of their own cities and scored a median of 71% correct inferences. Surprisingly, for the German cities, of which only X% were recognized, the median score was slightly higher: 73%.

Apparently, the students' lack of knowledge about German cities was beneficial in making correct inferences for this two-alternative choice task. In making their judgments, the students relied on the *recognition heuristic* (Goldstein & Gigerenzer 1999): if one of two alternatives is recognized and the other one is not, then infer that the recognized alternative has the higher value.

When is the recognition heuristic useful?

Gigerenzer (2008) summarized findings from 11 studies that analyzed the usefulness of the recognition heuristic as well as the degree to which people rely on it when making decisions. These studies replicated the city task, extended it to rivers, mountains, and islands, or used other two-alternative choice tasks like judging which of two NHL players has the higher career point total. The studies showed that people intuitively made the right choice in relying on the recognition heuristic in situations where it is most useful; that is, if there is a strong correlation between recognition and criterion, which can be measured as the *recognition validity* α. For example, in the city population task, α is the proportion of times a recognized city has a higher population than an unrecognized city.

Table 1 shows an overview of the findings from two studies conducted by Goldstein and Gigerenzer (1999) and Pohl (2006). For example, for pairs of the 20 largest Swiss cities, Pohl conducted a pilot study in which he asked a group of German students which of the cities they recognize. Based on the results, he calculated the recognition validity. These studies found that, if the recognition heuristic can be applied (i.e. if one of the two alternatives is *not* recognized), it would have led to more than 80% correct inferences when inferring population. Then, Pohl asked one group of participants which of the two cities is larger. When making their assessments, students were not informed about recognition validities. Nonetheless, in 89% of all cases, students intuitively relied on the recognition heuristic when assessing population. This was similar to results from other studies that analyzed the city population task. A second group was asked which of the two cities is farther away from the Swiss city Interlaken. Not surprisingly, recognition was not helpful to solve this task (α=0.51) and only 54% of the students made inferences that were consistent with the recognition heuristic.

Table 1: Validity of human recognition and search query date to solve two-alternative choice tasks

Task	Performed by	Reference	Human recognition		Popularity validity
			Recognition validity	Inference consistent with recognition heuristic	
Population of 22 American cities	American students	Goldstein and Gigerenzer 1999	-	-	.78
Population of 22 German cities			.80	90%	.77
Population of 20 Swiss cities	German students	Pohl 2006, Experiment 1	.86	89%	.77
Population of 11 Belgian cities	German students	Pohl 2006, Experiment 3.	.89	88%	.75
Population of 11 German cities			-	-	.87
Population of 11 Italian cities			.82	89%	.65

The collective recognition heuristic

While the recognition heuristic appears to work well to solve two-alternative choice tasks, its application is limited: an individual decision-maker cannot apply the heuristic if he recognizes both of the alternatives. A way to solve this problem is the *collective recognition heuristic* (Goldstein & Gigerenzer 2009). That is, one asks a group of people whether they have heard of each of the alternatives or not. Then, one ranks the alternatives according to collective recognition. Finally, for each pair of alternatives, the alternative with the higher rank is predicted to win.

The collective recognition heuristic in forecasting

Goldstein and Gigerenzer (2009) summarized three studies that directly tested the performance of the collective recognition heuristic for predicting sports events. Thereof, two similar studies compared the recognition heuristic to ATP rankings and an expert seeding to forecast the 127 matches of the Wimbledon tennis tournaments in 2003 and 2005 (Serwe & Frings 2006, Scheibehenne & Bröder 2007). In both studies, the recognition heuristic performed equally well than the two benchmarks. The third study, conducted by Pachur and Biele (2007), analyzed the performance of the recognition heuristic to predict the 24 first-round matches of the 2004 European Soccer Championship. Forecasts based on laypeople's recognition were correct 65% of the time but were less accurate than three benchmarks: expert forecasts, the FIFA world ranking of soccer teams, and the performance of the teams in the qualifying round. Furthermore, Goldstein and Gigerenzer (2009) related the results of two further studies on predicting the results of soccer games to the recognition heuristic. The authors of these studies reported that laypeople – in primarily predicting that the team which was more familiar to them would win the match – provided more accurate forecasts than experts.

In certain situations, collective recognition appears to be able to provide as accurate predictions as more knowledge-intense approaches. However, obtaining information on name recognition from a group of people may be costly – and is certainly not fast.

2. The potentials of search query data for forecasting

To our knowledge, not much work has been done on using search query data for forecasting thus far. However, two similar studies have shown the potential of using such information to monitor and predict influenza outbreaks in the United States. In analyzing the relationship between search queries for influenza from the Yahoo! search engine and actual influenza outbreaks, Polgreen et al. (2008) developed linear models that predicted influenza outbreaks 1-3 weeks before they occurred. Furthermore, they were able to predict an increase in mortality attributable to pneumonia and influenza up to 5 weeks in advance. However, they did not compare their model to a benchmark approach. Similarly, Ginsberg et al. (2008) developed a model, which relates Google search queries to influenza-related physician visits. They reported that the search query data was consistently able to predict influenza-related physician visits 1-2 weeks ahead of the surveillance reports of the U.S. Centers for Disease Control and Prevention. These two studies demonstrate the potential of search query data to improve forecasting in the highly relevant field of health care. The forecasts may enhance the possibilities of health care providers to react to seasonal epidemics. For example, it might allow for preparing for a need of additional pharmaceuticals or vaccine in an area that expects an influenza outbreak.

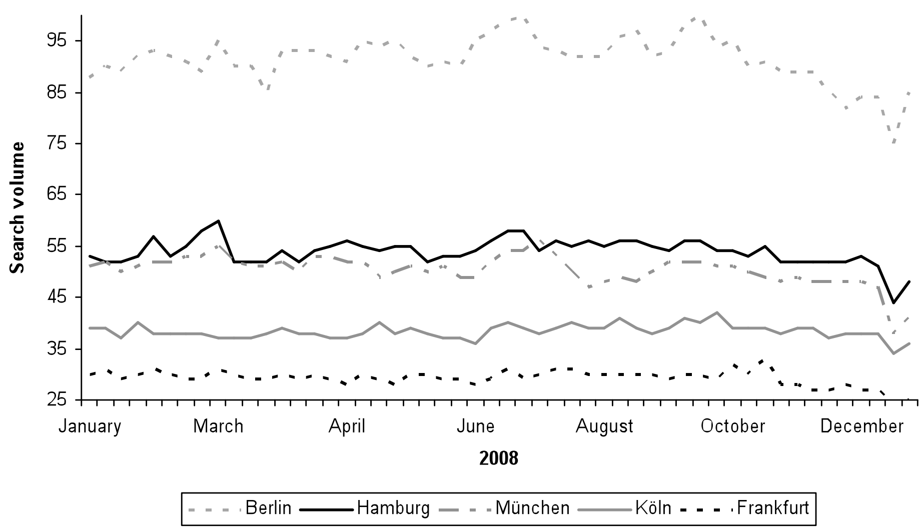
With a similar approach, Askitas and Zimmermann (2009) demonstrated the usefulness of Google data for generating short-term forecasts of the monthly unemployment rate in Germany. Although the analysis was limited to a small data set from January 2004 to April 2009, their simple statistical model indicated strong correlations between keyword searches and unemployment rates. The authors concluded that the continuously available search query data is useful in complex and changing situations where it is impossible to incorporate traditional flows of information. In particular, compared to traditional approaches, the method seems to have advantages in anticipating turning points or structural breaks.

Using search query data to measure popularity

In looking for information on certain keywords, Internet search engine users reveal information. Search engines like Google and Yahoo! save such information as search query data, which provides information on how many searches have been done for certain keywords, relative to the total number of searches in a certain period of time. Thus, search query data should be useful to measure the relative popularity of certain keywords. The more users search for a certain keyword, the higher should be the keyword's popularity.

Figure 1 shows the relative search volumes for the names of the five largest German cities from *Google Insights for Search*[1]. This service allows for comparing the frequency of searches for certain keywords. Thereby, it is possible to restrict the results to certain regions, categories, and time frames. The resulting data has been scaled and normalized by Google. We limited our search to data from 2008 and only for queries submitted within Germany (using German city names). Search volume was highest for Berlin, followed by Hamburg, Munich, Cologne, and Frankfurt. Thus, the search volumes exactly reflect the actual ranking of these five cities in terms of population.

Figure 1: Search query volumes for the five largest German cities



Note that the information incorporated in search query data differs from recognition. Studies on name recognition show people a pair of alternatives and ask them which of the two they recognize. By comparison, search engine users who query certain keywords have heard of these keywords before and actively look for further information. Thus, search query data does not measure recognition of keywords but rather their popularity.

[1] http://www.google.com/insights/search

First tests of popularity validity

To test for the ability of search query data to measure popularity, we replicated the studies reported in Table 1. That is, we searched *Google Insights for Search* for the search volumes of each city name. In the case of Goldstein and Gigerenzer (1999), who conducted the study with American students, we limited the search to queries submitted within the United States. In the case of Pohl (2006), we limited our search to queries submitted within Germany. Then, we used a heuristic similar to the recognition heuristic: *infer that the city with the higher search volume has the higher value*. In the following, we refer to this heuristic as the *popularity heuristic*. Accordingly, the *popularity validity* for the population task is the proportion of times a city with the higher search volume has a higher population.

The results are shown in the last column of Table 1. Although the popularity validities are not as high as the recognition validities, they would have led to correct inferences clearly above chance level. Note also that the search query validities were highest when using query data from within the United Stated for American cities (.78) and from within Germany for German cities (.87).

3. Using the popularity heuristic for forecasting two-alternative outcome tasks

These initial findings suggest that, for solving a two-alternative choice task, search query data can lead to similar performance than the recognition heuristic. We tested the usefulness of the popularity heuristic for (1) forecasting the outcomes of tennis matches and (2) predicting election outcomes.

3.1 Forecasting tennis matches

The task

Predicting the outcomes of tennis matches has been popular to test the performance of the recognition heuristic (Serwe & Frings 2006, Scheibehenne & Bröder 2007). We tested our approach to forecast the four men's Grand Slam tennis tournaments (the Australian Open, the French Open, Wimbledon, and the U.S. Open) in both 2007 and 2008. Each tournament starts with 128 players who, in total, play 127 matches (in each match, one player is eliminated, except for the final winner who never loses a match). Thus, over the eight tournaments, we analyzed forecasts for 1,016 matches.

Popularity heuristic forecasts

We obtained data for the six months prior to the start month of each tournament. For example, as the U.S. Open tournament usually starts in late August, we obtained data from February to July of the same year. Using the players' full names as keywords, we obtained the relative frequency with which search engine users looked for information on each of the two players playing each other. For example, in the event of Roger Federer playing Rafael Nadal, we entered *Roger Federer* and *Rafael Nadal* in the respective search fields. If players' names contained umlauts or hyphens, we used variations of the names (e.g. 'Jonas Björkman OR Jonas Bjorkman OR Jonas Bjoerkman" or "Paul Henri Mathieu OR Paul-Henri Mathieu").

Then, using the popularity heuristic, we derived the *popularity prediction* (PP) for each match: *the player with the higher search volume was predicted as the winner*. If the search volumes for both players were equal or too low to be displayed, the time horizon for obtaining the data was extended by one month until the search query data discriminated between the two players.

Benchmark forecasts

We compared the PPs to two benchmark forecasts that incorporate extensive information about the players: the ATP (Association of Tennis Professionals) rankings and betting odds. For both Wimbledon tournaments, we used the expert seeding as an additional benchmark.

ATP ranking

Each match was predicted based on the ATP Entry ranking, which is commonly referred to as the 'world ranking'. At any point, it reflects each player's performance at major tournaments during the immediate past 52 weeks. The player ranked first on this list is referred to as the Number 1 player in the world. In analyzing the performance of rankings for predicting the outcomes of basketball, tennis, and NFL games, Boulier and Stekler (1999, 2003) found that rankings based on past performance are good predictors.

The rule for predicting the matches from the ATP rankings was similar to the rule used for the PP: predict that the player with the higher ranking will win the match.

Betting odds

Betting odds from five online bookmakers (bet365.com, expekt.com, ladbrokes.com, pinnaclesports.com and unibet.com) were used as additional benchmark forecasts. The player with the better odds was predicted to win the match.

Betting odds are known for high accuracy in the domain of sports forecasting. For example, in analyzing 1,212 NFL games, Boulier and Stekler (2003) found betting odds to be more accurate than tipsters and ranking based forecasts. Similar results were reported by Forrest et al. (2005) for predicting nearly 10,000 English soccer games: betting odds were superior to tipsters' as well as to statistical models.

Wimbledon expert seeding

Among the 128 players that compete in each of the Grand Slam tournaments, 32 are seeded. For the Australian Open, the French Open, and the U.S. Open, the seeding is based on the ATP ranking derived from past performance. In contrast, an expert committee that evaluates players' performance on the unique grass courts determines the Wimbledon seeding, which has often differed from the ATP ranking. Thus, we used the expert seeding as an additional benchmark for the two Wimbledon tournaments.

Again, the player with the higher ranking in the seeding was predicted to win the game. If none of the two players is seeded, then guess. In our sample, the seeding forecasts had to guess 40 times in 2007 and 44 times in 2008.

Results

Table 2 shows the results for the popularity heuristic, ATP rankings, and betting odds for each tournament as well as the predictions based on the seeding for the two Wimbledon tournaments. In total, predictions based on the ATP rankings were available for all 1,016 matches. For four matches, no search query data was available, which reduced the number of PPs to 1,012.[2] Also, for certain matches, no betting odds were available from some of the online bookmakers. Thus, the number of observations for the betting odds varied from 1,002 to 1,013.

Performance of the popularity heuristic

The popularity heuristic performed well and clearly outperformed chance. Overall, the PPs correctly predicted 707 out of 1,012 matches (or 70%). The highest score was obtained for the Australian Open 2007 (76% correct predictions), whereas for both French Open tournaments, the performance of the popularity heuristic was worst.

Comparison to benchmark forecasts

The ATP ranking was slightly more accurate than the popularity heuristic and correctly predicted 735 out of 1016 matches (or 72%). Except for the two French Open tournaments, differences were small, ranging between 2 and 4% of correct predictions. In the case of Wimbledon 2008, the PPs outperformed the ATP ranking by providing accurate predictions for 74% of the matches (vs.

---

[2] *Google Insights for Search* only shows results for search terms that receive a significant amount of traffic. In our sample, even by extending the search period back to January 2004, we were unable to obtain information for four matches. This reduces our sample of PPs to 1,012.

66% for ATP). Not surprisingly, betting odds were most accurate and provided correct predictions for 77 to 79% of all matches. Only for the Australian Open 2007, both PPs and ATP outperformed betting odds. The relative performance of the Wimbledon experts' seeding and the popularity heuristic was mixed. While in 2007, the seeding based predictions were more accurate, they were less accurate in 2008.

Table 2: Percentage of correct predictions of the popularity heuristic, ATP rankings, Wimbledon seeding and betting odds

		Popularity heuristic	ATP	Wimbledon Seeding	Betting odds
2007	Australian Open	.76	.79	-	.72-.74
	French Open	.66	.72	-	.76-.78
	Wimbledon	.69	.73	.73	.84-.86
	U.S. Open	.68	.70	-	.76-.80
2008	Australian Open	.72	.75	-	.77-.79
	French Open	.62	.69	-	.76-.79
	Wimbledon	.74	.66	.63	.75-.78
	U.S. Open	.74	.76	-	.75
Total		.70 (n=1,012)	.72 (n=1,016)	-	.77-.79 (n=1,002-1,013)

Majority and consensus forecasts

We calculated consensus forecasts for pairs of PP, ATP, and betting odds[3] as well as the combination of all three methods. That is, we analyzed only the cases in which all components agreed on the outcome of a match. The results are shown in Table 3. All combinations of consensus forecasts led to a higher percentage of correct predictions compared to the percentage of correct predictions when relying only on one method. With 82% of correct predictions, forecasting accuracy was highest for the consensus forecasts of all three methods.

Table 3: Consensus forecasts of the popularity heuristic (PP), ATP, and betting odds

	PP = ATP	PP = ODDS	ATP = ODDS	PP = ATP = ODDS
Number of consensus forecasts	796	794	855	716
Correct forecasts	612	641	685	588
Percentage of correct forecasts	.77	.81	.80	.82

Discussion

Our results showed that the popularity of tennis players, measured as aggregated search query information, could be helpful to predict the winner of tennis matches. In correctly predicting 70%

[3] We calculated the mean odds of the five online bookmakers.

of 1,012 Grand Slam tennis matches, the popularity heuristic did clearly better than chance and was nearly as accurate as performance based ATP rankings (72% of correct predictions).

These results closely resemble the findings of two similar studies by Serve and Frings (2006) and Scheibehenne and Bröder (2007), who obtained tennis player name recognition from amateur tennis players and laypeople to predict the Wimbledon tournaments in 2003 and 2005, respectively. For the 2003 tournament, Serve and Frings (2006) reported an average prediction accuracy of 72% for the aggregated amateur recognition and 66% for the aggregated recognition of laypeople. For the 2005 tournament, Scheibehenne and Bröder (2007) found an average prediction accuracy of 68% for the aggregated amateur recognition and 67% for the aggregated recognition of laypeople. Overall participants the average prediction accuracy was 70%, which matched our results. In both studies, the performance of the respective ATP ranking was worse than in our sample (66% and 69% of correct predictions).

Among the competitors, betting odds performed best. Again, this conforms to the findings of Serve and Frings (2006) and Scheibehenne and Bröder (2007), who both found betting odds to be superior to recognition based forecasts and rankings. The reason for this superiority might be that betting odds change dynamically throughout the tournament by shifting in response to the weight of betters' money. In placing wagers, bettors continuously reveal information that was not accessible prior to the tournament. Also, informed betters likely possess information that is not incorporated in the simple heuristics based on recognition or ATP rankings. For example, they might have information on player performance, which often varies depending on the type of surface a particular match is played on: the French Open are played on clay; Wimbledon is played on grass, and the U.S. and Australian Open are played on hard courts. Or, betters might know about the record of two specific players playing each other, whether a player might suffer from injuries, or whether a player had the opportunity to rest longer than his opponent. In aggregating such information, betting odds are a highly sophisticated forecasting method and can hardly be outperformed by other approaches. Thus, it may appear unfair to compare them to simple heuristics like recognition or ATP ranking based forecasts.

Nonetheless, our results suggest that popularity – and ATP – based forecasts provide information that is not yet incorporated in betting odds and, thus, have the potential to improve forecasting. Over all matches, the percentage of correct predictions of the five online bookmakers ranged between .77 and .79. While this performance clearly outperformed the popularity heuristic and ATP, consensus forecasts of betting odds and other approaches were able to increase the percentage of correct predictions. For example, the consensus forecasts of PP and ODDS led to 81% of correct predictions. The consensus forecasts for ATP and ODDS generated 80% of correct predictions. Finally, with 82% of correct predictions, the best performance was achieved by relying on the consensus forecasts of all three approaches. It appears as if consensus forecasts of methods can be used to add certainty to the forecast.

3.2 Forecasting election winners

The task

We tested our approach to forecast the winner of the 2006 U.S. Congressional and Senate Election. We only analyzed races in which one Republican candidate ran against one Democratic candidate. Our sample included 366 Congressional races and 33 Senate races.

Popularity heuristic forecasts

We obtained search query data for the last week prior to Election Day (i.e. October 29 – November 4, 2006). Similar to above, the candidates' full names, as reported by the 2006 election statistics[4], were used as keywords. While we did not include initials, we also searched for nicknames. For example, in the case of the candidate "Howard P. 'Buck' McKeon", we searched for "Howard McKeon OR Buck McKeon". Search query data was obtained only from the state in which the election has been held. If the search volumes for both candidates were equal, we extended our search to the whole United States.

Again, we used the popularity heuristic to determine the winner of each race: *the candidate with the higher search volume was predicted as the winner*.

Benchmark forecasts

Voters often use other simple heuristics to make their decisions. We use two common heuristics as benchmarks: familiarity (incumbent advantage) and endorsements (money).

---

[4] http://clerk.house.gov/member_info/electionInfo/2006/2006Stat.htm

Incumbent advantage

The advantage of the incumbent in being reelected is well known, particularly in low recognition elections like House elections. The reason is similar to popularity and goes back to the voter's familiarity with candidates. Voters simplify their choice by focusing only on candidates they are already familiar with. Thus, the incumbent advantage heuristic predicted the incumbent to win the election. Of course, it can only be applied if an incumbent is in the race.

Money

Another good predictor for the success of a candidate is the amount of money raised. The more money a candidate has available, the more he can spend on campaigning and, thus, the more visible he will become to the electorate. The money heuristic predicted the candidate who raised more money during the campaign to win the election. The amount of money raised from each candidate was derived from opensecrets.org.

Results

Table 4 shows the results for the popularity heuristic as well as the money-based and incumbent-advantage forecasts. In total, we obtained PPs for 366 House races and the 33 Senate races.

Performance of the popularity heuristic

The popularity heuristic correctly predicted the winner for 90% of the House races and for 24 out of 33% (73%) of the Senate races. For both election types, the PPs were more accurate if there was an incumbent running and less accurate in open-seat elections.

Table 2: Percentage of correct predictions of the recognition heuristic and its competitors for the 2006 House and Senate Elections

House	**No. of races**	**PP**	**Money**	**Incumbent**
Incumbent in the race	334	.91	.92	.94
Open-seat	32	.75	.94	-
Total	366	.90	.92	-
Senate				
Incumbent in the race	29	.79	.76	.79
Open-seat	4	.25	.50	-
Total	33	.73	.73	-

Comparison to benchmark forecasts

Overall, the money-based forecasts (MF) were slightly more accurate than the PPs, although for the Senate races with an incumbent running, the PPs led to better results. Both times, the incumbent-advantage heuristic (IAF) performed well, yielding 94% correct predictions for the House races and 79% for the Senate races.

Majority and Consensus forecasts

We calculated majority and consensus forecasts of PP, MF, and IAF. The results are shown in Table 5. Majority forecasts were available for each race. If there was an incumbent in the race, the majority forecasts drew on all three forecasts and thus were unambiguous. In case of open-seat races, only on two cues (PP and MF) were available for the majority forecasts. Thus, if each PP and MF indicated a different candidate to win, the majority forecast predicted a tie. This happened for 6 House races and 3 Senate races. The majority forecasts performed about as well as the best individual forecast. For the House races, the majority forecasts overall yielded 92% correct predictions and thus performed similar to the MF. For the Senate races, the majority forecasts yielded 74% correct predictions, which slightly outperformed both PP and MF.

Table 3: Majority and consensus forecasts of PP, MF, and IAF

	Majority forecasts		Consensus forecasts	
House	**No. of races**	**Percentage correct**	**No. of races**	**Percentage correct**
Incumbent in the race	334	0.93	294	0.98
Open-seat	32	0.84	30	0.92
Total	366	0.92	324	0.97
Senate				
Incumbent in the race	29	0.79	24	0.83
Open-seat	4	0.38	1	0
Total	33	0.74	25	0.8

Consensus forecasts (i.e. cases in which each of the three[5] forecasts predicted the same winner) were available for 324 House races and 25 Senate races. The consensus forecasts performed well and correctly predicted the winner in 97% of the House races for which such a forecast was available. For Senate races, the consensus forecasts correctly predicted the winner in 80% of the races.

Discussion

The results showed that the popularity of political candidates, measured as aggregated search query information, could be helpful to predict election winners. In correctly predicting 90% of the House races and 73% of the Senate races in our sample, the popularity heuristic did clearly better than chance and performed similar to two benchmark heuristics. Furthermore, similar to the results from the tennis example, the popularity heuristic can add certainty through combining with other forecasts. The consensus forecasts of three heuristics (popularity, incumbent advantage, and money) correctly predicted 97% of the House races and 80% of the Senate races.

[5] Two forecasts for open-seat races: PP and MF.

Also, the popularity heuristic – as well as the benchmark heuristics – was clearly more accurate for predicting the winners in House races than in Senate races. House elections are generally low recognition elections where most voters do not have extensive information and motivation to evaluate the candidates. Thus, many voters might only recognize a candidate's name but might not be able to associate a face with that name. Then, they cannot evaluate the candidate by any other means than recognition, which might be why the popularity heuristic works particularly well in these cases. By comparison, Senate elections are high profile and more voters can be expected to have additional knowledge about the candidates, for example, they might recognize the candidates' faces. In that case, they can use other means to evaluate the candidates, which might explain the inferior performance of the popularity heuristic for Senate elections.

One other heuristic for voters to chose their preferred candidate is facial competence: select the candidate who looks more competent. Todorov et al. (2005) presented 31 subjects with pictures of candidates running in U.S. House and Senate elections. Based on one-second exposures, the subjects rated each candidate's competence (subjects who recognized a candidate were excluded). For the three Senate elections from 2000 to 2004, the most competent-looking candidates won 71% of the 95 races. For the two House elections in 2002 and 2004, the most competent-looking candidate won 67% of the 600 races in their sample. Although the comparison might be flawed due to a different and smaller sample in our case, the popularity heuristic performed about equally well for the Senate races but clearly outperformed facial competence for the House races. Again, the reason might be the information available to voters. For Senate races, more voters might recognize the candidate's faces and thus can apply the facial competence heuristic. For House races, many voters might not know how the candidates look like. They can only make their decision based on name recognition or popularity. The popularity heuristic appears to work well if no other information is available.

4. Potentials and limitations of search query data

In relying on search query data, the popularity heuristic benefits from search engine users' lack of information. In actively looking for information on certain keywords, search engine users reveal information that can be used to produce accurate forecasts. The method works well if there is a strong correlation between keyword popularity and criterion. In our case studies, the popularity heuristic yielded almost as accurate forecasts as approaches involving more information. Examples are the ATP ranking, which incorporates the performance of tennis players for the last 12 months, or the money based forecasts, which incorporate endorsements for – and thus pre-evaluation of – candidates from political elites and social groups. However, not surprisingly, in predicting the outcome of tennis matches, the popularity heuristic was clearly less accurate than forecasts derived from betting odds.

The popularity heuristic is very easy to use. This makes it valuable if decisions have to be made fast, at low costs, and with limited information. And, of course, if no other forecasting methods are available. In harnessing collective information from Internet users, it can also be used to make inferences for multiple-choice outcomes, for example, to decide about which movie to go to, which newspaper to read, or where to travel. While the popularity heuristic only makes binary choices between two outcomes, search query data has also been used to create quantitative forecasts, for example, to predict influenza outbreaks or unemployment rates. Thereby, the method seems to have advantages over traditional approaches in situations involving high uncertainties and structural breaks.

However, there are some open questions about using search query data for forecasting. First, the data might include information that is unrelated to the forecasting task and thus biases the results. For example, 'Donald Young' is not only a common American name but also an Alaskan Congressman in the U.S. House of Representatives, a retired Major League Baseball player as well as a professional tennis player. Accordingly, the keyword 'Donald Young' obtained high search volumes and the tennis player Donald Young was often mistakenly predicted to win his matches.[6] (The Congressman Donald Young was not included in our sample since he did not have a rival in the 2006 election.) Similar flaws happened for forecasting the election winners. For example, the Californian Democrat 'Charlie Brown' by far accumulated the highest search volume but closely lost the election against John Doolittle. Thus, caution is required when using search query data for forecasting, as sometimes judgment might be necessary when interpreting the results. Over time, one might learn about a keywords' validity in certain contexts and might even find ways to solve these problems computationally. Furthermore, users have the possibility to restrict *Google Insights for Search* to searches in predefined categories (like sports or politics), which may limit the incorporation of irrelevant data.

Second, it is unclear to what extent the approach is vulnerable to manipulation. Since search engines like Google only collect information about the search behavior of Internet users, such data might easily be manipulated. People who have an interest in manipulating the forecast, could program robots, which perform a large number of certain search queries and thus distort the data. For example, imagine health care providers would base their strategy for dealing with seasonal epidemics completely on forecasts derived from search query data. Then, pharmaceutical companies could try to manipulate search query data in order to increase demand for certain medication. Or, since search engine providers are the owners of the data, the companies

[6] Donald Young's highest ranking in the ATP Entry Ranking was 73.

themselves could only publish data, which represents their interests. It is a general problem that we have only limited information about how search query data is generated.

Such questions need to be resolved before search query data should be used for dealing with important problems like influenza outbreaks or economic indicators. Further research is necessary to learn more about the potentials and limitations of using search query data for forecasting.

5. Conclusion

We used search query information to measure popularity of tennis players and political candidates. Then, using a simple heuristic, we generated forecasts to predict the outcomes of the 1,016 matches of the eight Grand Slam tennis tournaments in 2007 and 2008 as well as of 366 House and 33 Senate races in the 2006 U.S. elections. The forecasts of our popularity heuristic performed well and correctly predicted 70% of the tennis matches, which was only slightly inferior to forecasts based on the ATP ranking (72% of correct predictions). Betting odds were most accurate. In predicting election winners, the method correctly predicted 90% of the House races and 73% of the Senate races. In both applications, a combination of the popularity heuristic with benchmark approaches improved forecast accuracy.

Search query data provides information that can be valuable for forecasting, especially for making decisions between multiple-choice outcomes. It is easy to use and allows for making fast decisions under limited information. However, further research is necessary before using search query data for solving highly important problems.

Acknowledgments

Thanks to Joseph Buchdahl for providing free access to tennis results and betting odds at www.tennis-data.co.uk. Thanks to Ishika Das for collecting data.

References

Askitas, N. & Zimmermann, K. F.: Google econometrics and unemployment forecasting, *Applied Economics Quarterly*, 55, 2009, 107–120.

Boulier, B. L. & Stekler, H. O. (1999). Are sports seedings good predictors? An evaluation, *International Journal of Forecasting*, 15, 83-91.

Boulier, B. L. & Stekler, H. O. (2003). Predicting the outcomes of National Football League games, *International Journal of Forecasting*, 19, 257-270.

Forrest, D., Goddard, J. & Simmons, R. (2005). Odds-setters as forecasters: the case of the football betting market, *International Journal of Forecasting*, 21, 552-564.

Gigerenzer, G. (2008). Fast and frugal heuristics. In: G. Gigerenzer (Ed.), *Rationality for mortals*. Oxford University Press, pp. 20-45.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S. & Brilliant, L. (2008). Detecting influenza epidemics using search engine query data, *Nature*, 457, 1012-1015.

Goldstein, D. G. & Gigerenzer, G. (1999). The recognition heuristic: How ignorance makes us smart. In: G. Gigerenzer & Todd, P. M. (Eds.), *Simple heuristics that make us smart*. Oxford University Press, pp. 37-58.

Goldstein, D. G. & Gigerenzer, G. (2009). Fast and frugal forecasting, *International Journal of Forecasting* (forthcoming), doi:10.1016/j.ijforecast.2009.05.010.

Pachur, T. & Biele, G. (2007). Forecasting from ignorance: The use and usefulness of recognition in lay predictions of sports events, *Acta Psychologica*, 125, 99-116.

Pohl, R. F. (2006). Empirical tests of the recognition heuristic, *Journal of Behavioral Decision Making*, 19, 251-271.

Polgreen, Philip M., Chen, Y., Pennock, David M. & Nelson, Forrest D. (2008). Using Internet searches for influenza surveillance, *Clinical Infectious Diseases*, 47, 1443-1448.

Scheibehenne, B. & Bröder, A. (2007). Predicting Wimbledon 2005 tennis results by mere player name recognition, *International Journal of Forecasting*, 23, 415-426.

Serwe, S. & Frings, C. (2006). Who will win Wimbledon? The recognition heuristic in predicting sports events, *Journal of Behavioral Decision Making*, 19, 321-332.

Todorov, A., Mandisodza, A. N., Goren, A. & C. C. Hall (2005). Inferences of competence from faces predict election outcomes, *Science*, 308, 1623-1626.