Artificial Intelligence: Individual Project

Modern Matchmaking

Agathe Benichou and Michael Teddick

Problem statement

- Successful matchmakers are being swamped with very eligible candidates within their communities looking for marriage and are having trouble keeping up with the demand.

- They are looking for a revolutionary recommender system that is easily downloadable onto their personal computers which they can use to ease their work while still creating prosperous matches for their communities.

- The matchmakers still want to retain their conventional and personal methods of matchmaking, such as meeting with their clients.


Parameter Highlights

Input and Knowledge

Explicit Input:

- An extensive form given to the client which inquires about personal information, family background, personality, lifestyle, views, professional ambitions and dating preferences.
- Meeting with the matchmaker for them to gain insight on their character.

Implicit Input:

- Utilizing this explicit input, the system will crawl the web to try and find additional information about the client that they did not disclose in their form and that the matchmaker did not pick up on.

Matchmaker Dating Philosophy:

- Each matchmaker differs in how they make matches and what they constitute as most important.

Output

- The system ultimately outputs a rank of eligible candidates for a client
  - Higher rank means higher matchability.

- All the eligible candidates above a threshold of 0.5 will given to the matchmaker.

- If they agree, the matchmaker will inform the client of the contact information and a basic profile summary of their highest match, as well as any advice on next steps.

Assumptions and Constraints

Assumptions:

- A matchmaker will man the system.
- The matchmaker can reject clients.
- The client accepts the matchmakers dating philosophy.
- Clients will provide strict requirements and dating preferences.

Constraints:

- Clients will only receive one match at a time.
- After having received a match, the client must go on at least one date.
- If a match is unsuccessful, the client must meet with the matchmaker before receiving another match.

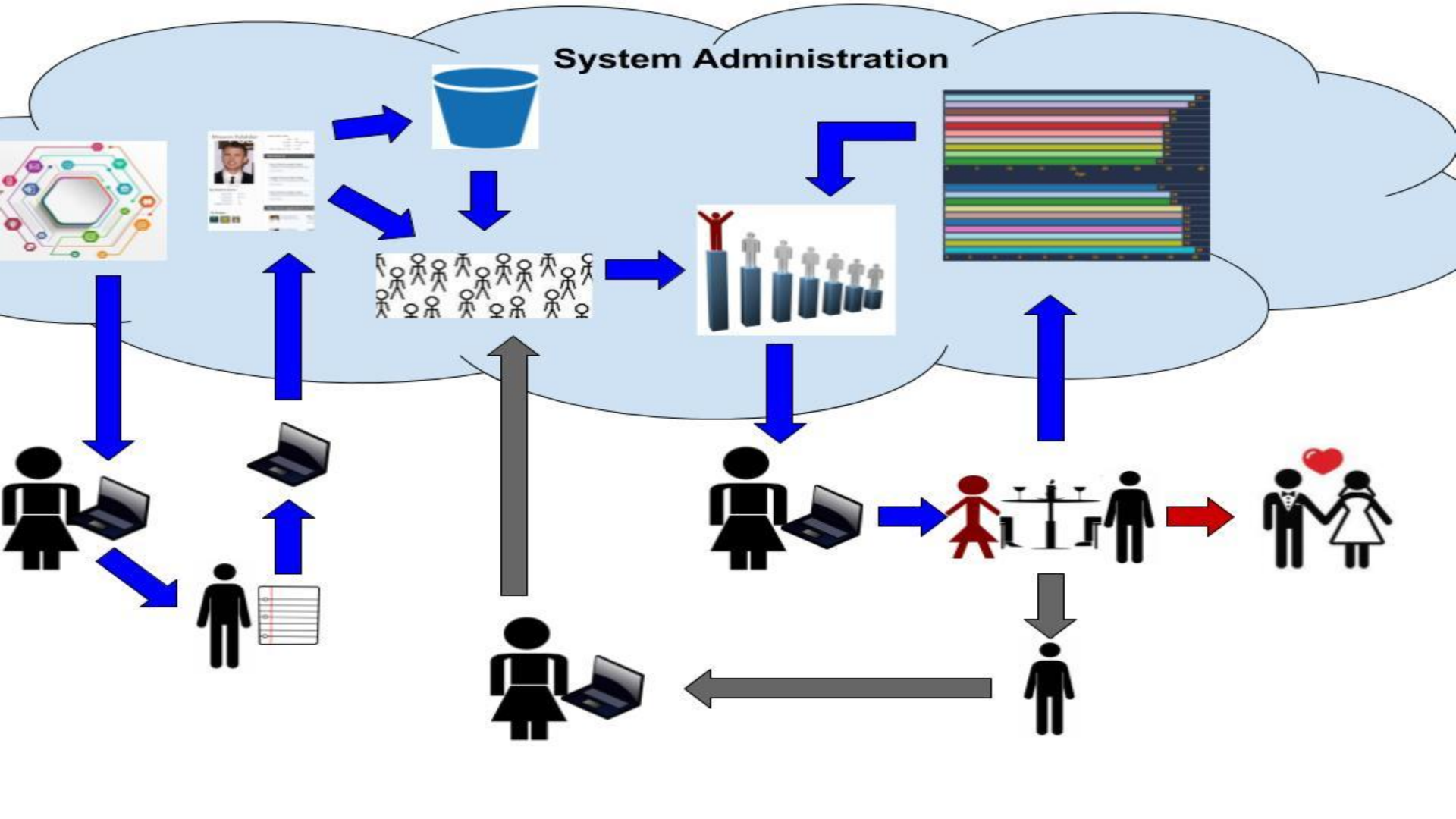
Evaluation Criteria

- Success in the matchmaking process is defined as marriage.
- To measure success, a scale was created which assigns values to each type of possible outcome from a match:

Marriage: +3%, Engagement: +2%, >5 dates: +1%, 3-5 dates: 0%, 2 dates: -1%, 1 date: -2%

- If clients stop dating at any point in this process, they meet with a matchmaker and fill out a feedback form which helps pinpoint what went wrong or what they did not like.


Solution Overview


System Administration


Preprocessing Steps

Data Collection

Explicit Data Collection:

- Occurs through the extensive, mandatory multiple-choice form that a client fills out.

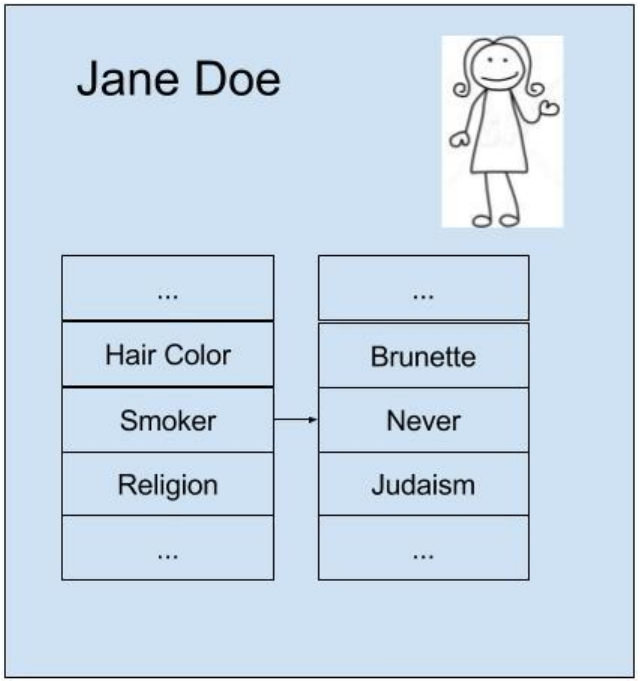
Implicit Data Collection:

- Occurs using three web scraping techniques: Human Copy-Paste, Regular Expressions, Semantic Annotations
- Given a webpage that the system determines should be scanned for information:
    - Human copy-paste extracts all the text from that page to be analyzed.
    - Regexs find keywords within the text to help identify a sentence containing important information. These sentences containing relevant keywords are set aside.
    - Semantic annotation extract the important information from that sentence that relates to defined keywords.

Data Representation and Organization

- Once data collection has been completed, this data will be stored in a profile.
  - Profile composed of answers recorded from the categorial questions posed on the form, their strict requirements and preferences for a match.

- Each matchmaker contains their own bucket of storage within the systems private cloud platform.
  - This bucket contains directory of clients, their match preferences, and history of matches.

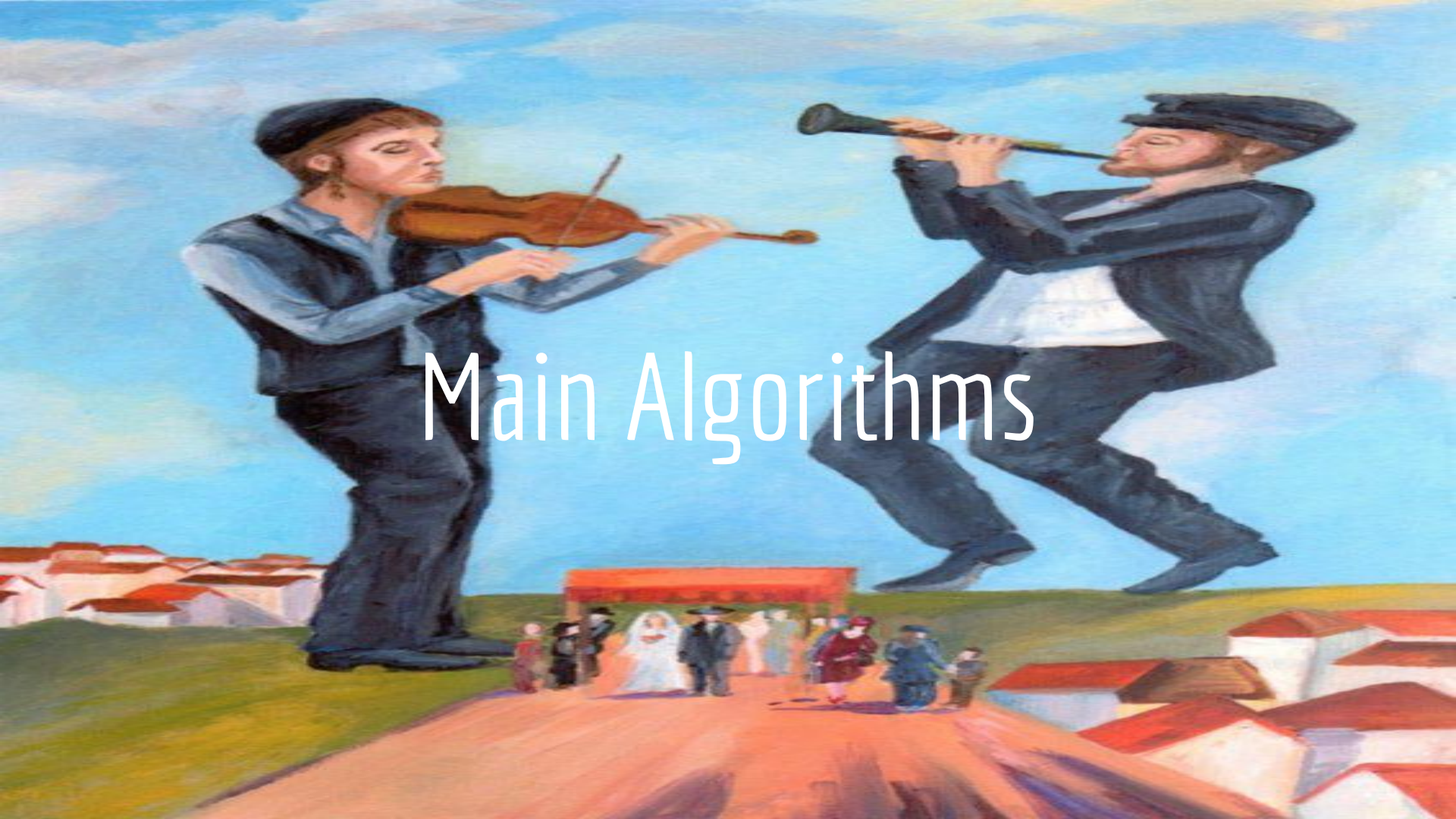


Preprocessing

Strict Requirements vs Dating Preferences:

- Strict Requirements (Hard Constraints) are necessities that a client's match must have, such that the client will not consider anyone who does is outside of this.
  - ‘Man’, ‘Between 21-25’, ‘Jewish’, ...
- Dating Preferences (Soft Constraints) are the desired preferences that a client would like their match to have but are not necessary.
  - ‘Brunette’, ‘Tall’, ‘Intelligent’, ...

- The system utilizes strict requirements to query the matchmakers directory to obtain a pool of eligible candidates. From these candidates, the system ranks them by utilizing dating preferences.


Main Algorithms

Matchmaking Algorithm

- The matching algorithm that this system uses is a derivative of a combination of algorithms, namely, Description Logics, Logic Programming, and Fuzzy Theory.

ID	Name	Hair	Religion	Age	Income	Ethnicity	Weight
24	Jeff Love	Blonde	Catholic	28	92,500	White	210
321	Levi Berg	Brown	Jewish	32	153,000	White	185

ID	Name	Hair	Religion	Age	Income	Ethnicity	Weight
33	Jen Li	Blonde	None	23	33,400	Asian	105
121	Cat Smith	Brown	Catholic	27	103,000	White	115

Matchmaking Algorithm

This example will use a Knowledge Base K = {F, O, P}.

Women(121, Cat Smith, Brown, Catholic, 27, 103,000, White, 115)

Catholic is a subset of Christianity.

Matchmaking Algorithm

- The Knowledge Base (K) contains the rule component (P) which is the component where ‘rules’ or preferences of the client will be stored.

Hard Constraints:

- *Must be Jewish or Catholic.*

Soft Constraints:

- *(1) I would like him to make over 100k if he is over 30.*
- *(2) I would prefer him under 200 lbs.*
- *(3) His hair should be Blonde.*

$Wa(x) = Jewish(x)$

$Wa(x) = Catholic(x)$

$W(x) = Wa(x)$

$W1(x, a) = Income(x), Age(a)$

$W2(x) = Weight(x)$

$W3(x) = Hair(x)$

$$Woman(x, a, m) = W1(x, a), W2(x), W3(x), m = .5 * s1 + .4 * s2 + .1 * s3$$

Matchmaking Algorithm

- This same process will be completed for the male resulting in two functions.
- The male and female function will then be combined into one all inclusive function.

$$\textit{Hard Constraints: } M(x)$$

$$\textit{Soft Constraints: } Man(x, a, m)$$

$$Match(x, a, m) = W(x), Woman(x, a, m1), M(x), Man(x, a, m2), m = m1 * m2$$

$$ans_2(K_{match}, Match)$$

Matchmaking Algorithm

- Results in probabilities that correlate to each individual that the system tries to match Cat Smith with, creating a table:
  - The ID of the individual with the probability that this will be a successful match.

solve the **Top-k retrieval** problem:

$$ans_k(\mathcal{P}, Match) = \mathtt{Top}_k\{\langle x, \mathbf{y}, u\rangle \mid \langle \mathbf{y}, u\rangle \in \mathtt{Top}_1\{\langle x, \mathbf{y}', u'\rangle \mid \mathcal{P} \models Match(x, \mathbf{y}', u')\}\}.$$

where $Match$ is the conjunctive query

$$Match(x, \mathbf{y}, u) \leftarrow \beta(x, \mathbf{y}_\beta), Buyer(x, \overline{\mathbf{y}_\beta}, u_\beta), \sigma(x, \mathbf{y}_\sigma), Seller(x, \overline{\mathbf{y}_\sigma}, u_\sigma), u = u_\beta * u_\sigma$$

and for each variable in the array $\mathbf{y}$, the same variable occurs in $\mathbf{y}_\beta, \overline{\mathbf{y}_\beta}, \overline{\mathbf{y}_\sigma}$ or $\mathbf{y}_\sigma$.

ID	Probability
24	0.3010
321	0.1885


Learning

Learning

- How are the matches made improved? By learning!
- The main method of learning employed in this system is using word vectors to build rich profiles for each characteristic of a person.

RICH

Rich	Blonde	Athletic	Brunette	...
.5	.5	.5	.5	...

BLONDE

Rich	Blonde	Athletic	Brunette	...
.5	.5	.5	.5	...

...

...	...	...	...	...
...	...	...	...	...

Learning

Updating the word vectors based on feedback. For this example consider the match got married (+3%). These probabilities are utilized as soft constraints.

For an increase: CurProb + (ModProb * (1 - CurProb))

For a decrease: CurProb * (1 - ModProb)

BROWN

Rich	Blonde	Athletic	Brunette	…
.5	.5	.5	.5	…

BROWN

Rich	Blonde	Athletic	Brunette	…
.5	.515	.5	.5	…

This was calculated from this formula: ***CurProb + (ModProb * (1 - CurProb)) = (.5)+ ((.03) * (1 - (.5)))***


Questions?

References

[2] Vargiu, Eloisa; Urru, Mirko, et al. *“Exploiting web scraping in a collaborative filtering-based approach to web advertising.*” Sciedu, Artificial Intelligence Research, 5 December 2012.

[3] Amandi, Analia; Schiaffino, Silvia, et al. *“Intelligent User Profiling.”* SpringerLink, Springer US, International Federation for Information Processing, 2010.

[4] Torlak E., Jackson D. (2007) *“Kodkod: A Relational Model Finder.”* In: Grumberg O., Huth M. (eds) Tools and Algorithms for the Construction and Analysis of Systems. TACAS 2013. Lecture Notes in Computer Science, vol 4424. Springer, Berlin, Heidelberg

[5] Pazzani, Michael; Billsus, Daniel. *“Learning and Revising User Profiles: The Identification of Interesting Web Sites.”* Kluwer Academic Publishers Machine Learning (2009) 27: 313.

[6] Munzert, Simon; Rubba, Christian; Meibner, Peter; Nyhis, Dominic. *“Automated Data Collection with R: A Practical Guide to Web Scraping and Data Mining”*. John Wiley & Sons Ltd., 2015

[7] Malik, Sanjay Kumar; Rizvi, Sam. *“Information Extraction Using Web Usage Mining, Web Scraping and Semantic Annotation.”* IEEE, CICN: International Conference On Computational Intelligence and Communication Networks, 2011.