Brennan, Jimmy, and Zach notes

11/12/2019

- Yes there are issues in using ML, but we can still use it to help solve societal problems
- Look at Amulya's research to solve problems within low resource communities
    - Access to public health
    - Access to healthy food
    - Access to rehab facilities
    - Access to good governance
- How can we use AI to solve the above issues?

Example 1: HIV among homeless Youth

- Interventions among homeless youth
    - Try to raise awareness about HIV prevention
    - Urge them to adopt safer behaviors
    - Encourage them to spread the message among social circles
- Homeless shelters: Resource Constraints
    - Cannot intervene with every homeless youth
    - Rely on word-of-mouth effects to inform about HIV
- Standard influence maximization
    - Inputs:
        - Social network G
        - Influence Model I
        - Number of Nodes :
    - Outputs:
        - Optimal set of K nodes to maximize expected influence spread
- State of the art in influence maximization
    - Algorithms with theoretical guarentees
        - Greedy
        - Celf
        - TIM
    - Efficient heuristics
        - IRIE
        - Sketch based heuristics
    - But do these algorithms suffice
- uncertainties: what do they mean?
    - In social network structure, there are inconsistencies. For example, homeless people are not the most reliable so we assume that they are not 100% correct
    - Uncertainties about state of influence of nodes - uncertainty about which nodes influence each other
    - Critical challenges in the real world: multi-stage selection

    - Problem - dynamic influence maximization under certainty

Pilot Study

- 3 pilot studies done with 173 homeless youth in LA
    - 2 different homeless shelters
    - Raise awareness
    - Head to head comparison of algorithms
- HEALER and DOSIM in the field,
    - Test healer and dosim
    - Verify need for using AI algorithms
- 1st pilot study
    - HEALER - 62 youth
- 2nd pilot study
    - DOSIM - 56 youth
- 3rd pilot study - around
    - 

Possible Issues with pilot studies

- Different case study groups
    - Healer (62 youth)
    - DOSIM (56 youth)
    - Degree centrality (55 youth)
    - Groups from similar locations
    - Small size

Standard influence maximization

- Input
    - Social network G
    - Influence model I
    - Number of nodes to choose K
- Output
    - Optimal set of K nodes to maximize expected influence spread

**Maximizing the spread of influence through a social network**

Problem Setting
- Given
  - A limited budget B for initial advertising (give way free sample of product)
  - Estimates for influence between individuals
- Goal
  - Trigger a large cascade of influence (further adoptions of a product)
- Field of maximization is governed by S-curve of adoption
- Question:
  - Which set of individuals should B target at?
- Application besides product marketing
  - Spread an innovation
  - Detect stories in blogs

What we need
- Form models of influence in social networks
- Obtain data about particular network
- Devise algorithm to maximize spread of influence

Outline
- Models of influence
  - Linear threshold
  - Independent

General operational view
- A social network is represented as a directed graph, with each person (customer) as a node
- Nodes start either active or inactive
- An active node may trigger activation of neighboring nodes
- Monotonicity assumption: active nodes never deactivate

Linear Threshold Model
- A node v, has random threshold
  - Threshold between 0,1
- A node v is influence by each neighbor w according to a weight b such that
- A noed v becomes active when at least fractions of its neighbors are active

Independent Cascade
- When node v becomes active

Critical challenge in the real world
- Strength of friendship can be in ranges
- Emerging trends in ML - blackbox, morality, ethics (all negative)

- AI for social good
    - Despite issues with ML, AI has many positive impacts
    - How can AI machine learning solve societal problems
- How can we use AI to solve low resource communities problem?
    - No access to public health facilities
    - Healthy food
    - Access to rehab facilities
    - Access to good governance (farming communities)
- Many of the problems of low resource communities can be solved by increasing social networks (<u>*not necessarily online*</u>)
- Current problems: Typically NGO's who help these communities don't have enough money, time and volunteers
- How can AI be used to raise awareness of HIV prevention to homeless youth?
    - Homeless use 10x more likely to get affected by HIV

- Traditional method: NGO’s conduct “interventions” peer leader is given information and taught how to spread this information to their peers.
    - Urge them to adopt safer behaviours
    - Encourage to spread message
    - Diffusion of information
    - Resource constraints, cannot intervened on every homeless youth themselves.
    - Rely on word of mouth
    - GOAL: maximize number of youth who is informed about HIV
- Friend network, which people should we target to maximize spread of information (similar to viral marketing)
- Standard Influence maximization
    - Inputs:
        - Social network (G)
        - Influence model (I): description of how influence spreads between people
        - Number of Nodes to choose (K): depends on number of people in homeless shelter
    - Output
        - Optimal set of K nodes to maximize expected influence spread

    - INFLUENCE maximization algorithms are not great with real problems
- Critical challenges in the real world: uncertainties
    - Uncertainty in social network structure
        - Homeless don't use social media (hard to gather data)
        - Uncertain about existence of
    - Uncertainty about state of influence of nodes
        - Unable to sample every node
        - Unknown which nodes are influenced and which are not
    - Multi-stage selection
        - Does not work for influence maximization
- given a  soical network G = (V,E)
    - Picking M subsets of nodes (M interventions organized by shelter) 3
    - Size of each subset is K (Max capacity of shelter) 4 leaders
- 2 algorithms to solve this problem, HEALER and DOSIM
    - 3 studies with 173 homeless youth in LA
    - 2 shelters (my friends place and Safe place for youth)
    - Raise awareness about HIV

    - Goal: verify need for using AI algorithms, verify usability
- HEALER/DOSIM/ Degree Centrality - select peer leaders
    - Network application: interacts with homeless youth
    - Algorithm: determines which people should be recommended for the intervention
    - Shelter official collects information and refines algorithm (are people really friends or not)
    - Looks at social network data and selects a set of nodes for first intervention
    - What percentage of non peer leaders were informed about HIV after the interventions 70% healer and dosim
    - What percentage of informed, non peer leader, untested people made a proactive move afterwards and got an HIV test
    - Improvement of 150% over degree centrality
- Degree Centrality (selects most popular people)
    - Creates too many redundant edges
    - Influencing people who are already influenced
    - Cross community edges are poor

    - Ignores cross community connections (only need to influence 1 person in a small group)
- What is wrong with the study
    - 3 different populations
    - Sample size too small
    - Randomness
    - Only ran the algorithm once
    - Can’t run twice
    - Cant determine statistical significance
    - Different messages spread differently

11/14/19

1. Network application (connects with youth) Algorithm (Feedback)
2. # of non leaders informed: Healer and Dosim 70% DC 27%
3. Maximizing the spread of influence through social network
    a. How to mitigate/maximize spread of influence
4. Influence: can cover anything (awareness/product/propaganda)

5. Scenario Given: a limited budget B for advertising, estimate **influence between individuals?**
    a. Goal: trigger a large cascade of influence
    b. Question : which set of individuals should B target
    c. **What you need:**
        i. Form models of influence in social networks
        ii. Obtain data about particular network (estimate inter personal influence)
        iii. Devise algorithm to maximize spread of influence)
6. Viral marketing: goal - ensure further adoptions of a product
7. S- curve of adoption: the spread of innovations curve
    a. Y: number of adopters
    b. X: time S - top part: sold too much
8. Applications besides product marketing
    a. Spread innovation
    b. Detect stories in blogs
9. Models of influence
    a. Two basic classes of diffusion models: threshold and cascade
    b. Operational View

i. A social network is represented as a directed graph and each node is a person/customer
ii. Nodes start either active or inactive (influenced or not)
iii. An active node may trigger activation of neighboring nodes
iv. Monotonicity assumption: active nodes never deactivate
v. Directed vs undirected graph (directional edges)
vi. Bidirectional normal friendship

10. Linear model threshold
    a. A node v has random threshold (0.7, will not become influenced until 70% of people in social network become influenced)
    b. Every relationship as different influential threshold number on a person (person around them .3 .5 .2 etc..)
11. Independent Cascade
    a. Probability that if you get influenced the other person will get influenced
    b. Person you rarely see, low percentage -- good friend, high
    c. Strength of friendships

12. How to solve influence maximization problem
    a. Set of K-nodes (initial adopters)
    b. Greedy algorithm
        i. Repeat procedure for K timesteps
        ii. For each timestep: add a single node to the result set
        iii. Results - 4 nodes
    c. Also works with any algorithm that has diminishing returns property(return on investment) math name - submodularity
    d. Influence maximization problem with diminishing returns property means always use the greedy algorithm
13. Real world algorithms: just give ranges of nodes from each other
    a. Rank people based off of the discussion of their friends, how good of friends is the other person
    b. Use ranges (be robust)
    c. Assume there is
14. 3 ways games can be solved
    a. Nash Equilibrium - what we are using
    b. Dominance
    c. Something else