Socio-spatial Properties of Online Location-based Social Networks

1. the main idea behind the paper (1-2 sentences)

The main idea behind this paper is to introduce a method to identify sentiments of a sentence in a document using conditional random fields to incorporate sentence structure and context information. The paper also proposes and evaluates two different active learning strategies for labelling sentiment data.

1. two questions that arose as you were reading the papers

* The authors say, “All of the state-of-the-art algorithms perform well on individual sentences without considering any context information”, which algorithms are they talking about?
* The authors also say, “we also employ active learning to help collect more labeled data”, what exactly do they mean by “labelled data”?

1. two strengths of the paper

* This paper proposes two different strategies to select data with high uncertainty for human beings to label, and the experimental results on customer reviews show faster convergence when compared to baselines.
* The approach used by the paper showed an increase in accuracy by 5-15% on Amazon Customer reviews compared to existing supervised learning and rule based methods

1. two weaknesses or areas for future work

* The authors say that their approach worked better for the Facebook comments because they had emoticons which conveyed sentiments better, the existing algorithm can be tweaked to obtain better results for the other datasets as well.
* The authors only use two strategies, ranking on the average probability and then in an ascending order, they can investigate further strategies such as weighting the most important sentences, ignoring some irrelevant text etc.

Towards Multimodal Sentiment Analysis: Harvesting Opinions from the Web

1. the main idea behind the paper (1-2 sentences)

This paper addresses the task of multimodal sentiment analysis, and conducts proof-of-concept experiments that demonstrate that a joint model that integrates visual, audio, and textual features can be effectively used to identify sentiment in Web videos. It also makes three important contributions, it introduces the concept tri-modal sentiment analysis, it identifies a subset of audio-visual features for sentiment analysis and lastly it introduces a new dataset consisting of real online data which will be useful for future research.

1. two questions that arose as you were reading the papers

* The dataset consists of 47 videos, I feel that this is an extremely tiny dataset and there needs to be at least a few hundred to claim that a model performs accurately.
* Analyzing certain visual cues to better understand the user’s sentiment can be done using some Computer Vision techniques, the authors have only used Smile Duration and Look away duration, why are they only using these two feature? I’m sure there are more features which can be used.

1. two strengths of the paper

* This paper introduces tri-modal sentiment analysis by integrating visual, audio and linguistics features and using it to determine polarity.
* The paper performs qualitative and statistical analysis identifying five multimodal features which are helpful to identify different sentiments, words, pitch etc.
* The authors create a new dataset which consists of video opinions, collected from YouTube and are annotated to help further research

1. two weaknesses or areas for future work

* The authors of this paper performed analysis on a relatively smaller dataset when compared to standard datasets, so for future work, they could probably analyze a larger dataset
* Also, they focus on a very narrow portion of the available type of data on the internet (Video Opinions), the authors can explore other domains such as Reviews, Political Debates etc.

You Are Where You Tweet: A Content-Based Approach to Geo-locating Twitter Users

1. the main idea behind the paper (1-2 sentences)

The authors propose and evaluate a probabilistic framework for estimating a Twitter user’s location entirely based on their tweets, even in the absence of any other geospatial cues. The paper goes on to identify words in tweets with a strong geo-local scope. The models generated by the authors places the twitter users within 100 miles of their location with an accuracy of 51%.

1. two questions that arose as you were reading the papers

* The authors say that “In particular, we extract a set of active users with 1000+ tweets who have listed their location in the form of latitude/longitude coordinates. “. Why did they only choose authors with over 1000 tweets?
* The authors say that “Instead of using stemming, we use the Jaccard Coefficient to check whether a newly encountered word is a variation of a previously encountered word “. Could they have used any other measures?

1. two strengths of the paper

* Only 24% of the tweets have actual geo-locations, therefore, finding a unique approach which can predict the location without using any geo-spatial data would be helpful for classifying the other 76% of tweets
* The authors also introduce a lattice based neighborhood smoothing model for refining the estimated results. This would allow personalized location based services, which would be helpful for users.

1. two weaknesses or areas for future work

* Since this is a data driven approach, more data can be added to probably improve the accuracy further.
* The authors can also try a different approach where they combine the purely content based approach with a social network based inference.

Socio-Spatial Properties of Online Location-Based Social Networks

1. the main idea behind the paper (1-2 sentences)

The paper presents a comprehensive study of the spatial properties of the social networks arising among users of three main popular online location-based services. They observe robust universal features across all of them, along with some heterogenous features. They also provide evidence that mechanisms akin to gravity models may influence how these social connections are created over space. This work is the first large scale study to investigate socio-spatial properties of online location based social networks

1. two questions that arose as you were reading the papers

* The probability of two individuals have a connection depending on the distance ‘d’ is still debatable and I’m not quite sure we can generalize it in the manner done by the previous work
* The authors could have chosen any three spatial social networks, why did they restrict themselves to these three and not anything else?
* In Network randomization, why did the authors choose those two particular models? They could have chosen anything else, why these two?

1. two strengths of the paper

* The methodology used by the authors is based on two randomized null models and highlights how observed properties deviate from what would be expected by chance with purely social or geographic mechanisms.
* They also deduce that LBSNs present universal spatial features across them, regardless of the service, its number of users or the adopted sampling method
* They observe and discuss heterogeneity in user socio-spatial behavior and observe that users exhibit friendship connections across a wide range of geographic distances and show similar variability in the social triads they belong to

1. two weaknesses or areas for future work

* One topic which can be investigated for future work can be understanding how heterogeneity arises in correlation with the temporal evolution of the social network
* If the generative mechanisms behind the various properties of how these models work can be understood, better models can be developed to perform more accurate predictions