Reviews For Paper

Paper ID	504
Title	Forecasting a Moving Target: Ensemble Models for ILI Case Count Predictions

Masked Reviewer ID: Assigned_Reviewer_1

Review:

Question	
Brief summary of the paper's contributions.	The paper addresses the problem of presenting a detailed prospective analysis of the temporal trends of flu activity using data sources for 15 Latin American countries. The paper proposes to integrate social and physical indicators using a novel matrix factorization based regression approach; investigates the efficacy of combining diverse sources at the data fusions level and model level and discussing relative merits; proposed different ways of handling uncertainties into prediction models and evaluating these methods in a real domain.
Overall rating	Weak Reject: I vote for rejection, but will not argue if it is accepted.
Brief justification for overall rating.	The paper addresses an important question in an interesting application domain: namely forecasting flu trends over a range of Latin American countries. The paper covers a wide range of issues. The matrix factorization approach seems technically sound and the results seem reasonable; however the presentation of the ideas and description of the results seemed somewhat disjointed and not at a level of detail to be convincing. In other words, the clarity and organization of the paper could be significantly improved and that would make this is a stronger paper. If accepted, the authors need to significantly improve the clarity of the paper.
	The paper has a good set of ideas but it seems to be hastily written. There are a plethora of typos and the core ideas seem to be glossed over. For instance,

Detailed comments. For more information, go to http://www-users.cs.umn.edu/~banerjee/sdm14/guidelines.html.	how does this approach compare to time series approaches (HMM, SVM/LD, ECTS, EDSC etc.) frequently used for prediction? What is the matrix factorization method more appropriate for this problem compared to other approaches? Table 1 is key to this paper and the discussion of the results int his paper are sorely lacking - this seriously degrades the clarity of this paper. The results in Tables 2-3 on combining data sources and effectively forecasting a moving target are encouraging. The novelty of the results in Table 4 is unclear. I think the idea of forecasting methods for a moving target has huge potential in areas ranging from epidemiology to network security. I strongly encourage the authors to provide enough detail in their paper so that the approach and results would be clear so that other researchers would be motivated to harness a similar approach.

Masked Reviewer ID: Assigned_Reviewer_4
Review:

Question	
Brief summary of the paper's contributions.	This paper presented a detailed prospective analysis of the generation of robust quantitative predictions of temporal trends of flu activity using several surrogate data sources for 15 Latin American countries.
Overall rating	Strong Accept: I will fight for acceptance.
Brief justification for overall rating.	This paper is targeting a very interesting problem, with significant amount of the data collection, analysis, and results.
Detailed comments. For more information, go to http://www-users.cs.umn.edu/~banerjee/sdm14/guidelines.html.	This paper presented a detailed prospective analysis of the generation of robust quantitative predictions of temporal trends of flu activity using several surrogate data sources for 15 Latin American countries. This is a very interesting problem and the paper in general is well-organized and well written. Among the various contributions of this

paper, this reviewer is particularly impressed by the approach that it integrate both social indicators and physical indicators to improve the performance. This seems to be a very promising approach for this difficult problem.

Masked Reviewer ID: Assigned_Reviewer_5
Review:

Question	
Brief summary of the paper's contributions.	This paper presents an approach that integrates both social indicators and physical indicators in predictions of temporal trends of flu activity and flu case counts. Its experiment shows a better performance against the baseline methods, and when combing different sources of data, model-level fusion performs better than data-level's.
Overall rating	Accept: I will argue for acceptance.
Brief justification for overall rating.	It is a very good idea to combine physical indicators and social indicator to forecast ILI incidence, because the former is a main cause of ILI case and the latter is a sign of its happening. Also, it provides comprehensive analyses in their experiment, which shows a superiority of its methods.
Detailed comments. For more information, go to http://www-users.cs.umn.edu/~banerjee/sdm14/guidelines.html.	Nevertheless, there are some issues (and suggestions) as given below: First, though 15 countries are involved in the dataset, yet it seems that they are independent from each other, and the prediction is made for each country based on its own data, and then what the difference does the number of countries make? I mean there may be some relations among the countries are supposed to be taken in consideration, e.g., similar physical indicators or some influence even infection may take place among them. Second, the approach is presented towards creating the keyword dictionary in different languages by tracking in multiple media, where the keywords are treated equally. However, I think there must be some strategies in choosing the

keywords because it will directly affect the final results. My suggestion is to incorporate ideas on choosing keywords and assigning their weights, which can distinguish their importance in prediction.

Third, it will be better if the baseline methods have included the methods using search engine queries as mentioned in the present work, and it is expected to be compared with some pioneering work and see the performance exactly.